

# Perbandingan Metode Naive Bayes Classifier dan Support Vector Machine Pada Analisis Sentimen Wisata Biru Berdasarkan Ulasan Twitter, Instagram, dan Google Maps Review

Selvi Rahmadila, Debby Alita\*

Fakultas Teknik dan Ilmu Komputer, Program Studi Informatika, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: <sup>1</sup>selvi\_rahmadila@teknokrat.ac.id, <sup>2,\*</sup>debbyalita@teknokrat.ac.id

Email Penulis Korespondensi: debbyalita@teknokrat.ac.id

Submitted: 26/11/2025; Accepted: 30/12/2025; Published: 31/12/2025

**Abstrak**—Wisata biru di Provinsi Lampung telah diidentifikasi sebagai potensi unggulan yang meliputi kawasan pantai, pulau, dan wilayah bahari yang mempunyai daya tarik tinggi bagi wisatawan. Respons publik terhadap destinasi tersebut dapat direpresentasikan melalui ulasan daring yang dipublikasikan pada berbagai platform media sosial dan layanan ulasan. Dalam penelitian ini, dilaksanakan analisis serta perbandingan kinerja algoritma klasifikasi sentimen Naive Bayes Classifier dan Support Vector Machine terhadap ulasan wisata biru di Lampung. Sebanyak 3.950 data ulasan dikumpulkan dari Twitter atau X, Instagram, dan Google Maps Review. Data yang diperoleh selanjutnya diproses melalui tahap pembersihan teks untuk menghilangkan unsur yang tidak relevan, kemudian dilaksanakan pelabelan tema dan sentimen dengan pendekatan semi supervised learning. Representasi fitur teks dibentuk memakai metode Term Frequency Inverse Document Frequency. Hasil pelabelan memperlihatkan bahwa sebaran sentimen bersifat tidak seimbang dan didominasi oleh sentimen positif. Pengujian model dilaksanakan dengan skema pembagian data latih dan data uji sebesar 80 banding 20. Berdasarkan hasil evaluasi, tingkat akurasi sebesar 91,90 persen dihasilkan oleh algoritma Support Vector Machine, sedangkan Naive Bayes Classifier mencapai akurasi sebesar 90,38 persen. Temuan tersebut mengindikasikan bahwa Support Vector Machine mempunyai kemampuan yang lebih baik dalam menangani data teks berdimensi tinggi serta kondisi distribusi sentimen yang tidak seimbang. Hasil penelitian ini diharapkan dapat dimanfaatkan sebagai rujukan dalam pemilihan metode analisis sentimen yang tepat untuk mendukung pengelolaan dan pengembangan wisata biru berbasis data.

**Kata Kunci:** Analisis Sentimen; Naive Bayes Classifier; Support Vector Machine; TF-IDF; Wisata Biru

**Abstract**—Blue tourism in Lampung Province has been recognized as a leading regional asset encompassing coastal areas, islands, and marine zones with strong appeal to visitors. Public responses toward these destinations can be captured through online reviews distributed across multiple digital platforms. In this study, the performance of sentiment classification algorithms, namely Naive Bayes Classifier and Support Vector Machine, was examined and compared using reviews related to blue tourism in Lampung. A total of 3,950 review records were collected from Twitter or X, Instagram, and Google Maps Review. The collected data were subjected to a series of preprocessing stages, including text cleaning to remove irrelevant elements, followed by theme and sentiment labeling using a semi supervised learning approach. Feature representation was generated through the Term Frequency Inverse Document Frequency method to transform textual data into numerical form. The labeling results revealed an imbalanced sentiment distribution with a strong dominance of positive sentiment. Model evaluation was conducted using an 80 to 20 split between training and testing datasets. The evaluation results indicated that the Support Vector Machine achieved an accuracy of 91.90 percent, while the Naive Bayes Classifier reached an accuracy of 90.38 percent. These findings suggest that the Support Vector Machine demonstrates superior capability in handling high dimensional textual data and imbalanced sentiment distributions. The outcomes of this study are expected to provide empirical guidance in selecting appropriate sentiment analysis algorithms to support data driven management and development of blue tourism destinations.

**Keywords:** Sentiment Analysis; Naive Bayes Classifier; Support Vector Machine; TF-IDF; Blue Tourism

## 1. PENDAHULUAN

Pariwisata merupakan salah satu sektor strategis yang mempunyai kontribusi besar terhadap pertumbuhan ekonomi, pembangunan daerah, dan peningkatan kesejahteraan masyarakat di Indonesia [1]. Sebagai negara kepulauan, Indonesia mempunyai berbagai jenis destinasi wisata, termasuk wisata bahari yang menjadi daya tarik penting di banyak provinsi. Lampung, sebagai pintu gerbang Pulau Sumatera, termasuk wilayah yang mempunyai potensi wisata pesisir dan bahari yang sangat besar. Salah satu ruang lingkup pengembangannya adalah konsep wisata biru, yang meliputi pantai, pulau, serta kawasan bahari lain yang menawarkan keindahan visual, keunikan ekosistem, serta pengalaman rekreatif berbasis laut. Seiring meningkatnya kunjungan wisatawan [2], pemahaman terhadap persepsi, opini, dan kepuasan pengunjung menjadi penting sebagai bahan evaluasi dan perumusan strategi pengembangan. Dalam beberapa tahun terakhir, berbagai penelitian memperlihatkan bahwa pemanfaatan data digital khususnya dari media sosial menjadi salah satu pendekatan yang efektif untuk membaca suara publik secara *real time* [3].

Perkembangan teknologi informasi telah menghasilkan volume data ulasan wisata yang sangat besar dan beragam. Media sosial seperti Twitter/X, Instagram, maupun Google Maps Review menghasilkan opini langsung dari wisatawan, baik berupa komentar positif, keluhan, kritik, maupun saran [4]. Data ini menjadi sumber informasi berharga yang dapat dianalisis memakai pendekatan analisis sentimen, yaitu teknik untuk mengidentifikasi dan mengklasifikasikan pendapat publik ke dalam kategori sentimen positif, negatif, atau netral [5]. Penggunaan analisis sentimen dalam bidang pariwisata telah berkembang pesat karena terbukti mampu memberikan gambaran objektif mengenai persepsi wisatawan terhadap destinasi [6].

Analisis sentimen merupakan bagian dari ranah *Natural Language Processing* (NLP) yang mengolah teks

dalam jumlah besar menjadi informasi terstruktur. Dalam penelitian sebelumnya, berbagai metode klasifikasi telah digunakan untuk analisis sentimen ulasan wisata, seperti Support Vector Machine (SVM), Naive Bayes Classifier (NBC), Random Forest, dan jaringan saraf. Namun, SVM dan NBC menjadi dua metode yang paling banyak digunakan karena efektivitasnya pada data teks berdimensi tinggi [7].

Support Vector Machine dipahami sebagai salah satu algoritma klasifikasi yang dirancang untuk menangani data dengan tingkat kompleksitas yang tinggi. Pada mekanisme kerjanya, pemisahan antar kelas dilaksanakan dengan membentuk sebuah hyperplane yang dipilih sedemikian rupa agar jarak pemisah atau margin antar kelas menjadi paling optimal. Pendekatan ini memungkinkan perbedaan karakteristik antar kelas data direpresentasikan secara lebih jelas, terutama pada data berdimensi besar. Dalam penerapannya, kemampuan SVM diperluas melalui penggunaan teknik kernel, sehingga pola data yang bersifat non-linear tetap dapat dipetakan ke dalam ruang berdimensi lebih tinggi. Dengan cara ini, struktur data yang semula sulit dipisahkan dapat diperlakukan sebagai data yang lebih mudah diklasifikasikan. Kondisi tersebut menyebabkan tingkat akurasi yang dihasilkan pada tugas klasifikasi teks sering kali berada pada nilai yang relatif tinggi.

Berbagai penelitian terdahulu telah melaporkan bahwa performa SVM memperlihatkan hasil yang unggul dalam konteks analisis sentimen. Keunggulan ini ditemukan pada beragam jenis data teks, termasuk ulasan di sektor pariwisata maupun ulasan produk. Temuan tersebut mengindikasikan bahwa SVM mempunyai kemampuan yang baik dalam menangkap variasi bahasa, konteks, dan pola ekspresi sentimen yang kompleks, sehingga sering dijadikan sebagai metode andalan dalam penelitian klasifikasi sentimen berbasis teks [8]. Sementara itu, Naive Bayes Classifier (NBC) merupakan metode probabilistik yang populer karena kecepatan, kesederhanaan, serta performa yang stabil. NBC memakai *Teorema Bayes* dengan asumsi independensi antar fitur. Meskipun asumsi ini tidak selalu realistis pada data bahasa alami, NBC sering memberikan hasil yang cukup kompetitif dan mudah digunakan untuk dataset besar [9]. Pada penelitian analisis ulasan wisata, NBC juga banyak diterapkan karena kemampuannya yang baik dalam memproses teks pendek dari media sosial [10].

Penelitian ini dilaksanakan untuk menganalisis sentimen masyarakat terhadap wisata biru di Provinsi Lampung, dengan mencakup ulasan dari pantai, pulau, serta kawasan bahari yang menjadi daya tarik wisata. Data dikumpulkan dari tiga platform utama: Twitter/X, Instagram, dan Google Maps Review. Pemilihan ketiga platform tersebut mengikuti studi-studi sebelumnya yang memperlihatkan bahwa kombinasi data media sosial memberikan gambaran persepsi publik yang lebih komprehensif [11]. Data yang diperoleh kemudian melalui tahap pembersihan teks, normalisasi, serta pelabelan berdasarkan tema estetika wisata biru. Ekstraksi fitur dilaksanakan memakai metode TF-IDF, yang merupakan teknik umum dalam representasi teks karena mampu menekankan kata penting yang berbeda antar dokumen [12].

Dalam penelitian ini, dua algoritma klasifikasi utama diterapkan, yaitu Support Vector Machine dan Naive Bayes Classifier. Pemanfaatan Support Vector Machine dilaksanakan karena kemampuannya dalam memproses dan memodelkan data berdimensi tinggi secara efektif. Sementara itu, Naive Bayes Classifier digunakan sebagai metode perbandingan karena proses komputasinya yang relatif cepat serta efisiensi yang tinggi dalam pengolahan data teks [13]. Kedua algoritma diterapkan untuk mengelompokkan sentimen ulasan menjadi tiga kategori utama: positif, negatif, dan netral.

Penelitian sebelumnya oleh Ningsih *dkk.* menjelaskan perbandingan algoritma Naive Bayes dan Support Vector Machine dalam analisis sentimen data Twitter. Hasil penelitian memperlihatkan bahwa algoritma Support Vector Machine menghasilkan nilai akurasi yang lebih tinggi dibandingkan Naive Bayes, yaitu sebesar 70,83%, sedangkan Naive Bayes hanya mencapai akurasi 63,02%. Hasil tersebut memperlihatkan bahwa SVM lebih efektif dalam menangani data teks berdimensi tinggi dibandingkan Naive Bayes [14].

Penelitian lain oleh Mukharyahya *dkk.* melakukan perbandingan kinerja algoritma Naive Bayes dan Support Vector Machine pada data dengan distribusi kelas tidak seimbang. Hasil pengujian memperlihatkan bahwa Support Vector Machine memperoleh akurasi sebesar 81%, lebih tinggi dibandingkan Naive Bayes yang hanya mencapai 76%. Penelitian ini menegaskan bahwa SVM mempunyai performa yang lebih stabil dibandingkan Naive Bayes, terutama pada data dengan ketidakseimbangan kelas [15].

Pada penelitian lanjutan yang dilaksanakan oleh Kurniawan dan rekan, perbandingan kinerja dilaksanakan terhadap dua algoritma klasifikasi, yaitu Naive Bayes dan Support Vector Machine, dalam konteks analisis sentimen pada data Twitter yang berkaitan dengan ulasan marketplace. Dalam studi tersebut, data teks media sosial dianalisis untuk menilai kemampuan masing-masing algoritma dalam mengelompokkan sentimen yang terkandung di dalamnya. Proses evaluasi dilaksanakan pada beberapa dataset yang berbeda guna memperoleh gambaran kinerja yang lebih menyeluruh. Berdasarkan hasil pengujian yang dilaporkan, nilai akurasi yang dihasilkan oleh Support Vector Machine secara konsisten tercatat lebih tinggi dibandingkan dengan Naive Bayes pada seluruh dataset yang digunakan. Temuan ini memperlihatkan bahwa pola bahasa dan karakteristik teks pada media sosial dapat ditangkap dengan lebih baik oleh pendekatan Support Vector Machine. Keunggulan tersebut mencerminkan kemampuan SVM dalam membangun batas pemisah yang lebih optimal pada data berdimensi tinggi, sehingga variasi ekspresi sentimen pada teks media sosial dapat dikenali secara lebih akurat. Dengan demikian, hasil penelitian ini mengindikasikan bahwa Support Vector Machine mempunyai kinerja yang lebih andal dibandingkan Naive Bayes dalam tugas klasifikasi sentimen berbasis teks media sosial [10].

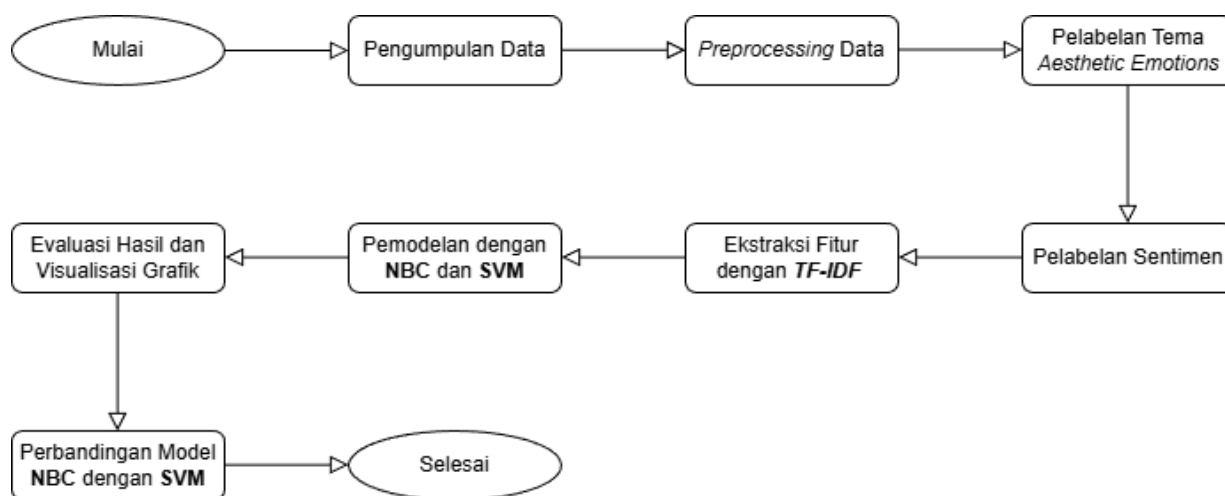
Kondisi tersebut menjadikan analisis sentimen sebagai pendekatan yang semakin relevan untuk memahami dinamika persepsi publik yang berubah dengan cepat. Pemilihan algoritma SVM dan NBC sebagai metode

pembandingan juga dinilai tepat karena keduanya mempunyai karakteristik performa yang berbeda dalam mengolah data teks, terutama ulasan wisata yang cenderung singkat, beragam, dan tidak terstruktur. Dengan demikian, penelitian ini diharapkan mampu mengisi kesenjangan dalam literatur sekaligus memberikan kontribusi bagi penyusunan strategi pengelolaan wisata biru yang lebih adaptif dan berorientasi pada pengalaman wisatawan.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Metode yang digunakan dalam penelitian untuk menganalisis sentimen berbasis *aesthetic emotions* memakai pendekatan *machine learning*. Metodologi mencakup keseluruhan alur kerja mulai dari pengumpulan data, pelabelan tema dan sentimen, *preprocessing*, hingga ekstraksi fitur dan pemodelan. Setiap tahapan dirancang untuk memastikan bahwa data yang digunakan mempunyai kualitas yang baik dan relevan sehingga model yang dibangun mampu memberikan hasil klasifikasi yang optimal [16]. Selain itu, pada bab ini juga dipaparkan proses evaluasi kinerja model serta perbandingan antara algoritma yang digunakan, yaitu Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM).



Gambar 1. Tahapan Penelitian

### 2.2 Pengumpulan Data

Data ulasan mengenai Wisata Biru Lampung dikumpulkan dari tiga platform utama, yaitu Instagram, Google Maps, dan Twitter/X. Proses pengumpulan dilaksanakan secara otomatis memakai Apify Console, sebuah platform berbasis cloud yang menyediakan *Actor* atau modul *scraping* siap pakai untuk mengekstraksi data dari berbagai sumber secara efisien.

#### a. Instagram (*Hashtag Scraper*)

1. Pengumpulan data dilaksanakan memakai *Actor* Instagram *Hashtag Scraper*.
2. Parameter yang digunakan berupa hashtag yang relevan dengan tema penelitian, seperti #wisatabirulampung, #pantailampung, #lampungwisata, #pulaulampung dan hashtag lain yang berkaitan dengan wisata biru.
3. Data yang diperoleh meliputi teks *caption*, jumlah *likes*, jumlah komentar, tanggal *posting*, dan *username*. Informasi ini berfungsi untuk mengidentifikasi opini publik dalam bentuk ulasan singkat maupun narasi panjang.

#### b. Google Maps (*Google Maps Reviews Scraper*)

1. Untuk platform ini digunakan *Actor* Google Maps *Reviews Scraper*.
2. Data diambil dari lokasi yang relevan dengan wisata biru Lampung, seperti nama pantai, pulau, atau tempat wisata bahari lainnya.
3. Data yang dikumpulkan berupa teks ulasan pengguna, rating bintang, tanggal ulasan, dan nama profil pengulas (jika tersedia). Data ini penting karena bersifat eksplisit dalam memperlihatkan kepuasan atau ketidakpuasan pengunjung.

#### c. Twitter/X (*Tweet Harvest/Twitter Scraper*)

1. Data dari Twitter dikumpulkan memakai *Actor* Twitter *Scraper* atau *Tweet Harvest*.
2. Parameter pencarian meliputi kata kunci seperti “Wisata Biru Lampung”, “Pantai Lampung”, “Sunset Lampung”, serta istilah lain yang relevan dengan tema estetika laut (ombak, pasir, senja, pulau).
3. Data yang diperoleh mencakup teks *tweet*, tanggal unggahan, jumlah *retweet*, jumlah *like*, dan *username*. Informasi ini memberikan gambaran *real-time* mengenai opini publik dalam bentuk spontan.

Seluruh data hasil *scraping* dari ketiga platform kemudian digabung menjadi satu file dengan format *.csv* agar



dapat diproses lebih lanjut pada tahap *preprocessing*, analisis sentimen, dan visualisasi. Jumlah total data yang dihasilkan sebanyak 3950.

**Tabel 1.** Jumlah dataset

Platfrom	Jumlah data
instagram	273
Google maps	2736
Twitter/X	941

**Tabel 2.** *Sample Dataset*

Nama	Teks	Platfrom
Arjuna	Viewnya bagus, bersih dan indah bnget pantai dan batu karangnya cocok buat	googlemaps
Poenya	fotber dwngan ombak yg menawan	
kontak Jo	Bagus pantainya	googlemaps
Mulyadi	Sangat menarik,rekom silahkan berkunjung,tidak akan mengecewakan,perjalanan dari Jambi ke Lampung ,fokus pantai marina,semoga ke depan semangkin di tingkatkan	googlemaps

### 2.3 Preprocessing Data

Tahap prapengolahan data ditempatkan sebagai langkah awal dalam rangkaian pengolahan data. Pada tahap ini, data mentah diperlakukan dan disesuaikan agar berada dalam bentuk yang lebih terstruktur serta layak untuk dianalisis lebih lanjut. Melalui proses ini, informasi yang semula belum seragam diubah sehingga dapat dipahami dan diolah oleh sistem komputasi secara efektif. *Preprocessing* dilaksanakan untuk memastikan bahwa data yang digunakan telah memenuhi kebutuhan proses data *mining*. Berbagai penyesuaian diterapkan agar noise dapat diminimalkan, ketidakkonsistenan dapat dikurangi, dan struktur data menjadi lebih jelas. Dengan demikian, hasil akhir dari tahap ini adalah data yang telah siap digunakan pada tahapan analisis berikutnya, sehingga proses penambangan data dapat berjalan dengan lebih optimal dan menghasilkan keluaran yang lebih akurat [17]. Tahap ini mempunyai peran penting karena menentukan kualitas hasil analisis yang akan diperoleh. Pada penelitian ini, proses *preprocessing* difokuskan pada pembersihan data dari elemen-elemen yang tidak relevan atau *noise*, sehingga data siap untuk dianalisis lebih lanjut [18].

#### a. Case Folding

Semua teks diubah menjadi huruf kecil (*lowercase*) memakai fungsi `.lower()` pada Python. Langkah ini bertujuan menyamakan bentuk kata sehingga “Pantai” dan “pantai” diperlakukan sama [19].

**Tabel 3.** Case folding

text	case folding
Pantai Baru di daerah pesawaran.	Pantai baru di daerah pesawaran.
Pantai dengan pasir putih.	pantai dengan pasir putih.
HTM 100rb untuk dewasa, anak <sup>2</sup> diatas 10tahun dihitung dewasa, dibawah 10tahun bayarnya 80rb.	htm 100rb untuk dewasa, anak <sup>2</sup> diatas 10tahun dihitung dewasa, dibawah 10tahun bayarnya 80rb.
Dengan HTM segitu sudah dapat welcome drink, bisa pilih sendiri, ada air mineral, fr**t tea, teh s***o.	dengan htm segitu sudah dapat welcome drink, bisa pilih sendiri, ada air mineral, fr**t tea, teh s***o.
Kamar mandi yg bersih dengan dilengkapi shower dan air yg lancar.	kamar mandi yg bersih dengan dilengkapi shower dan air yg lancar.
Pemandangan yg bagus sekali, angin sepoi sepoi.	pemandangan yg bagus sekali, angin sepoi sepoi.
Saran: klo bisa akses jalan masuknya diperbagus lagi, ya walupun pas kami kesana 25 Juni 2025 jalan masuknya sudah padat tadi masih belum rata, klo bisa diaspal juga biar mulus.	saran: klo bisa akses jalan masuknya diperbagus lagi, ya walupun pas kami kesana 25 juni 2025 jalan masuknya sudah padat tadi masih belum rata, klo bisa diaspal juga biar mulus.

#### b. Cleaning Text

Teks dibersihkan dengan menghapus URL, mention, tanda # (*hashtag*), angka, dan tanda baca memakai *regular expression* (`re.sub`) serta `2139ustaka string.punctuation`. Karakter kosong di awal dan akhir teks dihapus memakai `.strip()`.

**Tabel 4.** Cleaning Text

case folding	cleaned
pantai baru di daerah pesawaran.	pantai baru di daerah pesawaran
pantai dengan pasir putih.	pantai dengan pasir putih
htm 100rb untuk dewasa, anak <sup>2</sup> diatas 10tahun dihitung dewasa, dibawah 10tahun bayarnya 80rb.	htm rb untuk dewasa anak <sup>2</sup> diatas tahun dihitung dewasa dibawah tahun bayarnya rb



case folding	cleaned
dengan htm segitu sudah dapat welcome drink, bisa pilih sendiri, ada air mineral, fr**t tea, teh s***o. kamar mandi yg bersih dengan dilengkapi shower dan air yg lancar. pemandangan yg bagus sekali, angin sepoi sepoi.	dengan htm segitu sudah dapat welcome drink bisa pilih sendiri ada air mineral frt tea teh so kamar mandi yg bersih dengan dilengkapi shower dan air yg lancar pemandangan yg bagus sekali angin sepoi sepoi
saran: klo bisa akses jalan masuknya diperbagus lagi, ya walupun pas kami kesana 25 juni 2025 jalan masuknya sudah padat tadi masih belum rata, klo bisa diaspal juga biar mulus.	saran klo bisa akses jalan masuknya diperbagus lagi ya walupun pas kami kesana juni jalan masuknya sudah padat tadi masih belum rata klo bisa diaspal juga biar mulus

c. *Tokenizing*

Teks dipecah menjadi potongan kata (tokens) memakai *word\_tokenize()* dari pustaka NLTK. Tokenisasi memudahkan proses penghapusan kata tidak penting (*stopwords*) dan *stemming*

**Tabel 5.** Tokenizing text

cleaned	tokens
pantai baru di daerah pesawaran pantai dengan pasir putih htm rb untuk dewasa anak <sup>2</sup> diatas tahun dihitung dewasa dibawah tahun bayarnya rb dengan htm segitu sudah dapat welcome drink bisa pilih sendiri ada air mineral frt tea teh so kamar mandi yg bersih dengan dilengkapi shower dan air yg lancar pemandangan yg bagus sekali angin sepoi sepoi	['pantai', 'baru', 'di', 'daerah', 'pesawaran', 'pantai', 'dengan', 'pasir', 'putih', 'htm', 'rb', 'untuk', 'dewasa', 'anak <sup>2</sup> ', 'diatas', 'tahun', 'dihitung', 'dewasa', 'dibawah', 'tahun', 'bayarnya', 'rb', 'dengan', 'htm', 'segitu', 'sudah', 'dapat', 'welcome', 'drink', 'bisa', 'pilih', 'sendiri', 'ada', 'air', 'mineral', 'frt', 'tea', 'teh', 'so', 'kamar', 'mandi', 'yg', 'bersih', 'dengan', 'dilengkapi', 'shower', 'dan', 'air', 'yg', 'lancar', 'pemandangan', 'yg', 'bagus', 'sekali', 'angin', 'sepoi', 'sepoi', 'saran', 'klo', 'bisa', 'akses', 'jalan', 'masuknya', 'diperbagus', 'lagi', 'ya', 'walupun', 'pas', 'kami', 'kesana', 'juni', 'jalan', 'masuknya', 'sudah', 'padat', 'tadi', 'masih', 'belum', 'rata', 'klo', 'bisa', 'diaspal', 'juga', 'biar', 'mulus']
saran klo bisa akses jalan masuknya diperbagus lagi ya walupun pas kami kesana juni jalan masuknya sudah padat tadi masih belum rata klo bisa diaspal juga biar mulus	

d. *Stopword Removal*

Menghapus kata-kata tidak umum yang tidak mempunyai makna penting dalam analisis seperti “yang”, “dan”, “di”, dan “ke”. Daftar *Stopwords* memakai bahas Indonesia dari NLTK yang telah disesuaikan dengan kebutuhan penelitian ini [20].

**Tabel 6.** Stopword removal

tokens	tokens no stopwords
['pantai', 'baru', 'di', 'daerah', 'pesawaran', 'pantai', 'dengan', 'pasir', 'putih', 'htm', 'rb', 'untuk', 'dewasa', 'anak <sup>2</sup> ', 'diatas', 'tahun', 'dihitung', 'dewasa', 'dibawah', 'tahun', 'bayarnya', 'rb', 'dengan', 'htm', 'segitu', 'sudah', 'dapat', 'welcome', 'drink', 'bisa', 'pilih', 'sendiri', 'ada', 'air', 'mineral', 'frt', 'tea', 'teh', 'so', 'kamar', 'mandi', 'yg', 'bersih', 'dengan', 'dilengkapi', 'shower', 'dan', 'air', 'yg', 'lancar', 'pemandangan', 'yg', 'bagus', 'sekali', 'angin', 'sepoi', 'sepoi', 'saran', 'klo', 'bisa', 'akses', 'jalan', 'masuknya', 'diperbagus', 'lagi', 'ya', 'walupun', 'pas', 'kami', 'kesana', 'juni', 'jalan', 'masuknya', 'sudah', 'padat', 'tadi', 'masih', 'belum', 'rata', 'klo', 'bisa', 'diaspal', 'juga', 'biar', 'mulus']	['pantai', 'daerah', 'pesawaran', 'pantai', 'pasir', 'putih', 'htm', 'rb', 'dewasa', 'anak <sup>2</sup> ', 'diatas', 'dihitung', 'dewasa', 'dibawah', 'bayarnya', 'rb', 'htm', 'segitu', 'welcome', 'drink', 'pilih', 'air', 'mineral', 'frt', 'tea', 'teh', 'so', 'kamar', 'mandi', 'yg', 'bersih', 'dilengkapi', 'shower', 'air', 'yg', 'lancar', 'pemandangan', 'yg', 'bagus', 'angin', 'sepoi', 'sepoi', 'saran', 'klo', 'akses', 'jalan', 'masuknya', 'diperbagus', 'ya', 'walupun', 'pas', 'kesana', 'juni', 'jalan', 'masuknya', 'padat', 'klo', 'diaspal', 'biar', 'mulus']

e. *Stemming*

Mengubah kata menjadi bentuk dasarnya memakai algoritma “Nazief-Adriani yang diimplementasikan melalui Pustaka Sastrawi. Contohnya, kata “bermain” menjadi “main” dan kata “terindah” menjadi “indah”.

**Tabel 7.** Stemming text

tokens no stopwords	stemmed
['pantai', 'daerah', 'pesawaran', 'pantai', 'pasir', 'putih', 'htm', 'rb', 'dewasa', 'anak <sup>2</sup> ', 'diatas', 'dihitung', 'dewasa', 'dibawah', 'bayarnya', 'rb', 'htm', 'segitu', 'welcome',	['pantai', 'daerah', 'sawar', 'pantai', 'pasir', 'putih', 'htm', 'rb', 'dewasa', 'anak', 'atas', 'hitung', 'dewasa', 'bawah', 'bayar', 'rb', 'htm', 'segitu', 'welcome', 'drink', 'pilih',

'drink', 'pilih', 'air', 'mineral', 'frt', 'tea', 'teh', 'so', 'kamar', 'mandi', 'yg', 'bersih', 'dilengkapi', 'shower', 'air', 'yg', 'lancar', 'pemandangan', 'yg', 'bagus', 'angin', 'sepoi', 'sepoi', 'saran', 'klo', 'akses', 'jalan', 'masuknya', 'diperbagus', 'ya', 'walupun', 'pas', 'kesana', 'juni', 'jalan', 'masuknya', 'padat', 'klo', 'diaspal', 'biar', 'mulus']	'air', 'mineral', 'frt', 'tea', 'teh', 'so', 'kamar', 'mandi', 'yg', 'bersih', 'lengkap', 'shower', 'air', 'yg', 'lancar', 'pandang', 'yg', 'bagus', 'angin', 'sepoi', 'sepoi', 'saran', 'klo', 'akses', 'jalan', 'masuk', 'bagus', 'ya', 'walupun', 'pas', 'kesana', 'juni', 'jalan', 'masuk', 'padat', 'klo', 'aspal', 'biar', 'mulus']
---	---

## 2.4 Pelabelan Tema *Aesthetic Emotion*

Setelah tahap preprocessing, dataset terdiri dari 3.950 data ulasan, namun hanya sebagian kecil yang mempunyai label tema manual. Untuk mengatasi keterbatasan tersebut, penelitian ini menerapkan pendekatan semi-supervised learning dengan mengombinasikan data berlabel dan tidak berlabel [21]. Representasi fitur teks diperoleh memakai metode TF-IDF, kemudian algoritma Label Spreading diterapkan untuk menyebarkan label tema dari data berlabel ke seluruh data berdasarkan kemiripan fitur. Hasil pelabelan memperlihatkan bahwa seluruh data berhasil memperoleh label tema, dengan tema *beach ambience* sebagai tema yang paling dominan, diikuti oleh *nature & creature*, fasilitas, dan atraksi. Hasil ini memperlihatkan bahwa pendekatan semi-supervised learning mampu menghasilkan pelabelan tema yang konsisten dan relevan meskipun jumlah data berlabel manual terbatas.

**Tabel 8.** Pelabelan Tema *Aesthetic Emotions*

nama	text	platform	label
Laras Ayu	Bersih, cozy ala <sup>2</sup> beach club	googlemaps	beach ambience
Dayat Malang	Pantainya Asri & Alami	googlemaps	beach ambience
Riansyah Rian	Harganya tiketnya terlalu mahal 100ibu trus pantai y banyak batu karang dan kulit kerang sehingga kaki saya sakit hampir terluka	googlemaps	fasilitas
Nanang Nurkholis	Kerennn	googlemaps	beach ambience
nhasan.d mp	Too crowded, small sand beach Pantai yang ulasannya bagus, typical Beach Club, tapi masih sangat ramai, parkirannya penuh,	googlemaps	beach ambience

## 2.5 Pelabelan Sentimen

Setelah proses pelabelan tema, pelabelan sentimen dilaksanakan pada dataset yang sama dengan kondisi awal di mana hanya sebagian kecil data mempunyai label sentimen manual. Untuk mengatasi keterbatasan tersebut, penelitian ini kembali menerapkan pendekatan semi-supervised learning dengan mengombinasikan data berlabel dan data tidak berlabel. Representasi fitur teks diperoleh memakai metode TF-IDF, kemudian algoritma Label Spreading digunakan untuk menyebarkan label sentimen dari data berlabel ke seluruh data berdasarkan kemiripan fitur. Hasil pelabelan memperlihatkan bahwa seluruh data ulasan berhasil memperoleh label sentimen akhir yang terdiri dari kelas positif, negatif, dan netral, dengan dominasi sentimen positif. Hal ini mengindikasikan bahwa persepsi pengunjung terhadap objek wisata yang dianalisis cenderung bersifat positif.

**Tabel 9.** Pelabelan Sentimen

cleaned text	label	sentiment
bersih cozy ala beach club pantai asri alami	Beach Ambience	Positif
harga tiket mahal trus pantai y batu karang kulit kerang kaki sakit luka	Beach Ambience	Negatif

## 2.6 Ekstraksi Fitur dengan TF-IDF

Setelah proses pelabelan sentimen diterapkan pada data teks, tahapan selanjutnya dilaksanakan berupa transformasi teks ke dalam bentuk representasi numerik agar dapat diproses oleh algoritma klasifikasi. Pada penelitian ini, metode Term Frequency Inverse Document Frequency dimanfaatkan sebagai teknik ekstraksi fitur teks. Pendekatan ini dipilih karena kemampuannya dalam merepresentasikan karakteristik kata secara kuantitatif berdasarkan frekuensi kemunculannya dalam dokumen serta tingkat kepentingannya terhadap keseluruhan kumpulan teks.

- Term Frequency* (TF) mengukur tingkat kemunculan sebuah kata dalam suatu dokumen. Semakin sering kata muncul, semakin besar nilai TF-nya.
- Inverse Document Frequency* (IDF) memberikan bobot lebih besar pada kata yang jarang muncul dalam keseluruhan dokumen, dan menurunkan bobot pada kata-kata umum yang sering muncul.

Melalui penggabungan kedua komponen tersebut, tingkat kepentingan sebuah kata terhadap dokumen tertentu dalam keseluruhan kumpulan teks dapat dinyatakan secara kuantitatif memakai pendekatan TF-IDF. Pada tahap implementasi, teks yang telah melewati proses prapengolahan terlebih dahulu dikonversi ke dalam bentuk matriks numerik dengan memanfaatkan pustaka *TfidfVectorizer* yang tersedia pada *scikit-learn*. Matriks yang dihasilkan



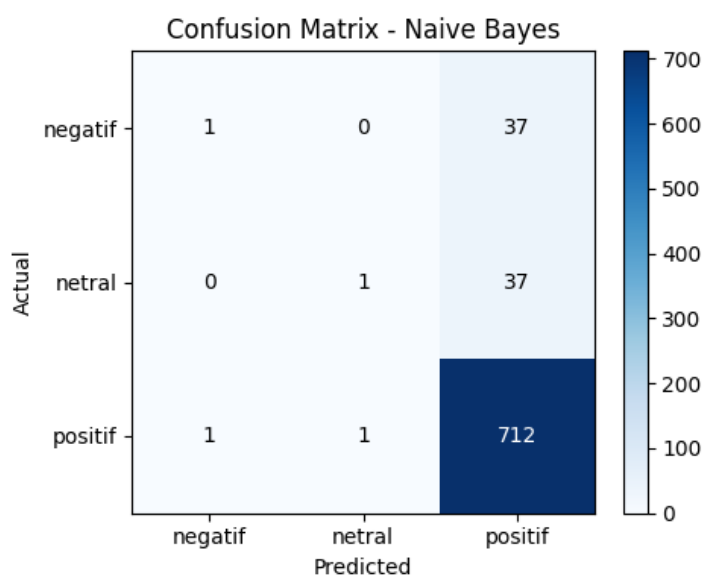
tingkat ketepatan yang tinggi. Kinerja paling optimal terlihat pada pengenalan sentimen positif, di mana nilai recall mencapai 1,00 dan f1-score sebesar 0,95. Capaian tersebut mengindikasikan bahwa hampir seluruh ulasan bernada positif dalam data uji telah berhasil diidentifikasi dan dikelompokkan secara tepat oleh sistem klasifikasi.

Sebaliknya, kemampuan model dalam mengenali sentimen negatif dan netral masih berada pada tingkat yang sangat terbatas. Nilai recall pada kedua kelas tersebut hanya mencapai 0,03, yang memperlihatkan bahwa sebagian besar ulasan dengan sentimen negatif maupun netral tidak dapat dikenali secara akurat dan justru dipetakan ke dalam kelas positif. Pola ini mencerminkan adanya kecenderungan bias yang kuat terhadap kelas sentimen mayoritas. Kondisi tersebut dipengaruhi oleh distribusi data sentimen yang tidak seimbang, sehingga model lebih mudah menyesuaikan diri dengan pola yang paling sering muncul dibandingkan variasi sentimen yang jumlahnya lebih sedikit.

Analisis distribusi hasil prediksi sentimen berdasarkan tema dengan memakai Naive Bayes Classifier dilaksanakan terhadap seluruh 3.950 data ulasan. Hasil klasifikasi memperlihatkan bahwa hampir semua ulasan pada setiap tema diarahkan ke dalam kategori sentimen positif. Tema beach ambience, nature and creature, serta atraksi sepenuhnya didominasi oleh sentimen positif, sedangkan pada tema fasilitas masih ditemukan sedikit variasi dengan keberadaan sejumlah kecil ulasan negatif. Pola temuan ini mengindikasikan bahwa model belum mampu merepresentasikan perbedaan nuansa sentimen antar tema secara mendalam. Variasi makna dan konteks pada setiap tema belum sepenuhnya tertangkap, sehingga hasil klasifikasi cenderung bersifat homogen dan terfokus pada sentimen positif.

### 3.3 Hasil Evaluasi Naive Bayes Classifier

Berdasarkan hasil pengujian pada data uji sebanyak 790 ulasan, Naive Bayes Classifier (NBC) menghasilkan nilai akurasi sebesar 90,38%. Nilai akurasi ini memperlihatkan bahwa secara keseluruhan model mampu mengklasifikasikan sentimen ulasan wisata biru dengan tingkat ketepatan yang cukup tinggi.



**Gambar 3.** *Confusion Matrix-NBC*

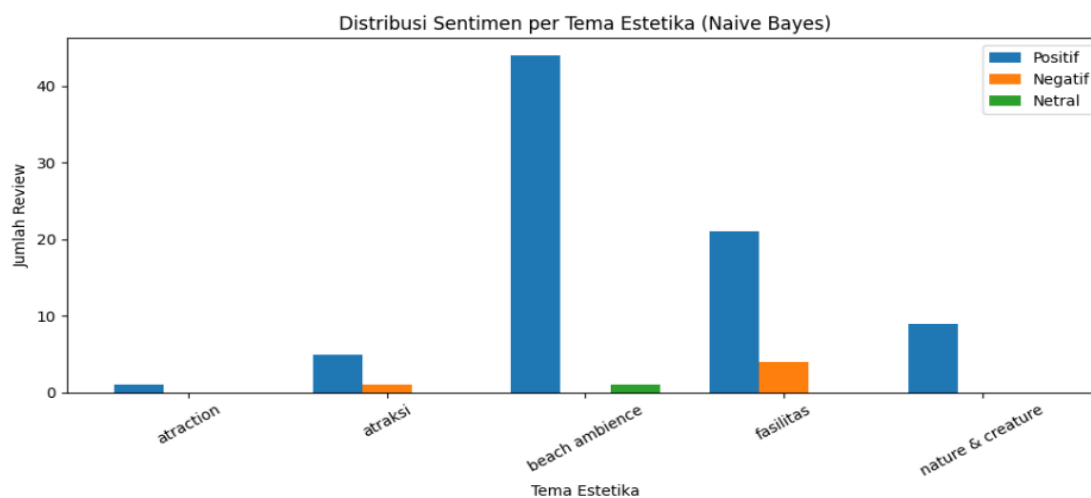
Penilaian kinerja model dilaksanakan dengan memanfaatkan laporan klasifikasi dan confusion matrix untuk menelaah kemampuan sistem dalam membedakan sentimen negatif, netral, dan positif. Proses evaluasi ini difokuskan pada ketepatan prediksi pada setiap kelas sentimen yang diuji. Berdasarkan laporan klasifikasi, kinerja yang sangat tinggi ditunjukkan oleh model Naive Bayes Classifier pada kategori sentimen positif. Pada kelas ini, nilai precision tercatat sebesar 0,91, nilai recall mencapai 1,00, dan f1-score berada pada angka 0,95. Capaian recall yang sempurna mengindikasikan bahwa hampir seluruh ulasan positif dalam data pengujian telah dikenali dan dikelompokkan secara tepat oleh model. Kondisi tersebut mencerminkan bahwa pola kata yang sering muncul pada ulasan bernada positif dapat ditangkap dengan baik oleh sistem klasifikasi.

Sebaliknya, performa yang dihasilkan pada kelas sentimen negatif dan netral berada pada tingkat yang sangat rendah. Pada sentimen negatif, nilai recall hanya tercatat sebesar 0,03 dengan f1-score sebesar 0,05. Angka ini memperlihatkan bahwa dari total 38 ulasan negatif yang diuji, hanya satu ulasan yang berhasil diidentifikasi secara benar sebagai negatif, sementara sebagian besar lainnya justru diprediksi ke dalam kelas positif. Situasi yang hampir sama juga ditemukan pada kelas netral. Nilai recall sebesar 0,03 memperlihatkan bahwa hanya satu dari 38 ulasan netral yang dapat dikenali dengan tepat oleh model.

Hasil confusion matrix mempertegas temuan tersebut. Dari 38 data ulasan negatif, sebanyak 37 ulasan diklasifikasikan secara keliru sebagai positif. Pola kesalahan yang identik juga terjadi pada kelas netral, di mana 37 ulasan salah diprediksi sebagai positif. Sebaliknya, pada kelas positif, tingkat ketepatan klasifikasi sangat dominan.

Sebanyak 712 dari 714 ulasan positif berhasil diprediksi dengan benar. Pola ini memperlihatkan adanya kecenderungan yang sangat kuat dari model untuk mengarahkan prediksi ke kelas sentimen positif.

Secara umum, hasil evaluasi mengindikasikan bahwa Naive Bayes Classifier sangat andal dalam mendeteksi sentimen positif, namun masih menghadapi keterbatasan yang signifikan dalam mengenali sentimen negatif dan netral. Bias terhadap kelas positif dipengaruhi oleh distribusi data uji yang tidak seimbang serta karakteristik asumsi independensi fitur pada metode Naive Bayes. Asumsi ini menyebabkan pola yang dominan lebih mudah ditangkap dibandingkan variasi bahasa yang lebih halus pada sentimen lain. Oleh karena itu, meskipun tingkat akurasi keseluruhan terlihat tinggi, penafsiran hasil klasifikasi perlu dilaksanakan secara lebih hati-hati dengan memperhatikan performa model pada masing-masing kelas sentimen.



**Gambar 4.** Grafik Sentimen per-Tema Estetika-NBC

Grafik distribusi sentimen per tema estetika memperlihatkan hasil klasifikasi sentimen memakai metode Naive Bayes terhadap seluruh 3.950 data ulasan wisata biru. Grafik ini menggambarkan sebaran jumlah ulasan dengan sentimen positif, negatif, dan netral pada setiap tema estetika, yaitu *beach ambience*, *fasilitas*, *nature & creature*, dan *atraksi*. Berdasarkan grafik, tema *beach ambience* merupakan tema dengan jumlah ulasan terbanyak, yaitu 2.467 ulasan. Pada tema ini, sentimen positif sangat mendominasi dengan 2.450 ulasan, sementara sentimen netral hanya berjumlah 15 ulasan dan sentimen negatif sebanyak 2 ulasan. Dominasi sentimen positif memperlihatkan bahwa sebagian besar wisatawan memberikan tanggapan yang sangat baik terhadap suasana pantai, seperti keindahan visual, kenyamanan lingkungan, dan atmosfer yang dirasakan selama berkunjung. Tema *fasilitas* mempunyai total 558 ulasan dengan distribusi sentimen yang masih didominasi oleh sentimen positif sebanyak 551 ulasan. Namun, pada tema ini masih ditemukan 7 ulasan dengan sentimen negatif dan tidak ditemukan sentimen netral. Keberadaan sentimen negatif pada tema *fasilitas* mengindikasikan adanya sejumlah keluhan wisatawan terkait sarana pendukung, seperti kondisi fasilitas umum, aksesibilitas, atau layanan yang tersedia.

Pada tema *nature & creature*, seluruh 782 ulasan diklasifikasikan sebagai sentimen positif. Tidak ditemukannya sentimen negatif maupun netral pada tema ini memperlihatkan bahwa persepsi wisatawan terhadap aspek alam dan lingkungan cenderung sangat baik dan konsisten. Hal ini mencerminkan kepuasan wisatawan terhadap keindahan alam, kondisi lingkungan, serta ekosistem yang masih terjaga. Tema *atraksi* mempunyai jumlah ulasan paling sedikit, yaitu 143 ulasan, dan seluruhnya diklasifikasikan sebagai sentimen positif. Hasil ini memperlihatkan bahwa aktivitas atau atraksi wisata yang tersedia umumnya memberikan pengalaman yang memuaskan bagi wisatawan, meskipun jumlah ulasan yang membahas tema ini relatif terbatas dibandingkan tema lainnya.

Secara keseluruhan, grafik distribusi sentimen per tema estetika memakai Naive Bayes memperlihatkan dominasi sentimen positif pada seluruh tema yang dianalisis. Namun, pola ini juga mengindikasikan bahwa model cenderung mengklasifikasikan sebagian besar ulasan ke dalam kelas positif, sehingga variasi sentimen negatif dan netral relatif kecil. Oleh karena itu, meskipun grafik memberikan gambaran umum persepsi wisatawan yang sangat positif, interpretasi hasil perlu dilaksanakan secara hati-hati dengan mempertimbangkan karakteristik model dan distribusi data sentimen.

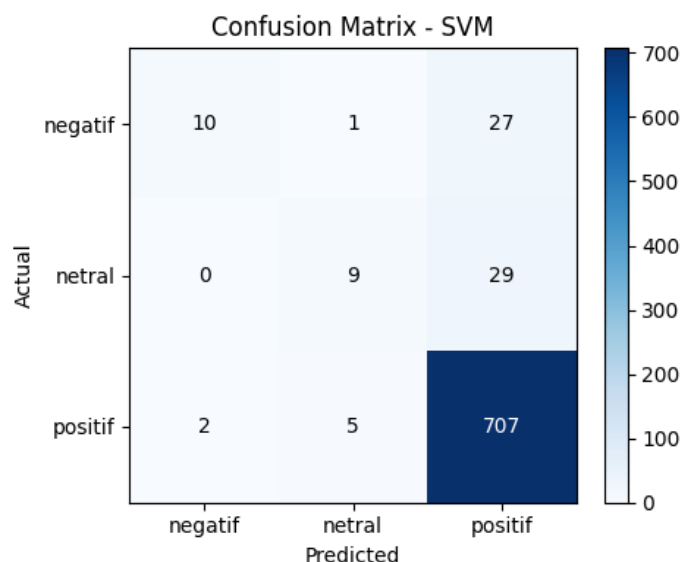
### 3.4 Implementasi Algoritma Support Vector Machine

Support Vector Machine dipahami sebagai algoritma klasifikasi yang dioperasikan dengan menentukan hyperplane paling optimal untuk memisahkan kelas data berdasarkan margin terbesar. Dalam penelitian ini, proses pelatihan dilaksanakan memakai data latih yang telah direpresentasikan melalui TF-IDF, kemudian kinerjanya dievaluasi memakai data uji yang sama seperti yang diterapkan pada Naive Bayes Classifier. Hasil pengujian memperlihatkan bahwa tingkat akurasi sebesar 91,90% berhasil dicapai, sehingga kinerja SVM tercatat lebih tinggi dibandingkan NBC. Kinerja yang sangat baik terlihat pada klasifikasi sentimen positif. Pada kelas ini, nilai precision tercatat sebesar 0,93, nilai recall mencapai 0,99, dan f1-score berada pada angka 0,96. Capaian tersebut memperlihatkan bahwa sebagian

besar ulasan dengan sentimen positif dapat dikenali dan dikelompokkan secara tepat oleh model. Pola bahasa yang dominan pada ulasan positif berhasil ditangkap secara efektif melalui pendekatan SVM. Meskipun demikian, kemampuan model dalam mengidentifikasi sentimen negatif dan netral masih tergolong terbatas. Nilai recall yang diperoleh pada kelas negatif dan netral masing-masing sebesar 0,26 dan 0,24. Angka ini mengindikasikan bahwa sebagian ulasan dengan sentimen negatif dan netral masih cenderung diprediksi sebagai sentimen positif. Namun demikian, jika dibandingkan dengan Naive Bayes Classifier, SVM memperlihatkan kemampuan yang lebih baik dalam mengenali kelas minoritas, sehingga performanya dinilai lebih seimbang dalam konteks klasifikasi sentimen. Distribusi sentimen berdasarkan tema hasil prediksi SVM memperlihatkan variasi yang lebih proporsional. Pada tema *beach ambience*, sebagian besar ulasan bersentimen positif, namun tetap terdapat ulasan negatif dan netral. Tema fasilitas menampilkan distribusi sentimen yang lebih beragam dibandingkan tema lainnya. Sementara itu, tema *nature & creature* dan atraksi tetap didominasi oleh sentimen positif dengan jumlah ulasan negatif yang relatif kecil. Hasil ini memperlihatkan bahwa SVM lebih efektif dalam menangkap variasi sentimen antar tema dibandingkan NBC.

### 3.5 Hasil Evaluasi Support Vector Machine

Berdasarkan pengujian yang dilaksanakan pada data uji yang terdiri atas 790 ulasan, tingkat akurasi sebesar 91,89% diperoleh oleh model Support Vector Machine. Capaian ini memperlihatkan bahwa proses klasifikasi sentimen pada ulasan wisata biru dapat dilaksanakan dengan tingkat ketepatan yang tinggi. Hasil tersebut mencerminkan bahwa pola bahasa dan karakteristik sentimen dalam ulasan mampu direpresentasikan dan dipisahkan secara efektif oleh model, sehingga prediksi yang dihasilkan berada pada tingkat keandalan yang baik.



**Gambar 5.** Confusion Matrix-SVM

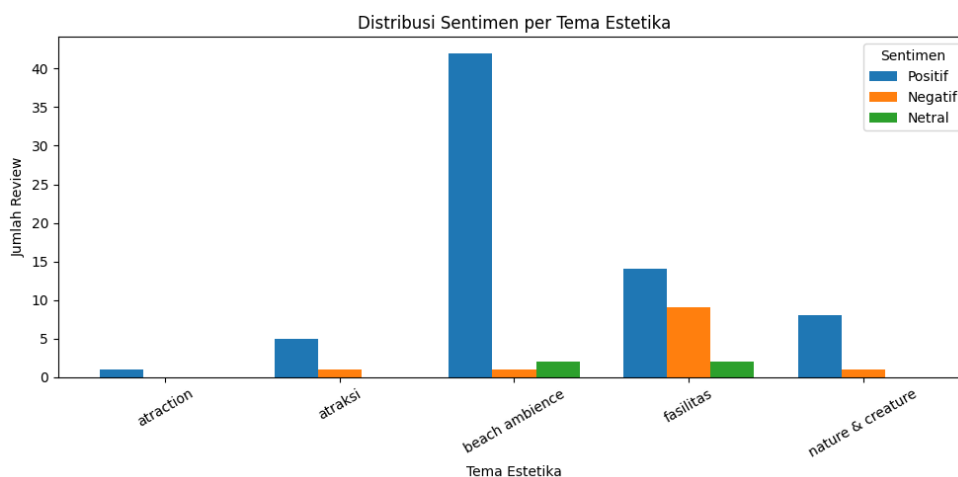
Evaluasi kinerja model dilaksanakan melalui pemanfaatan laporan klasifikasi dan confusion matrix untuk menilai kemampuan sistem dalam mengelompokkan setiap kelas sentimen, meliputi negatif, netral, dan positif. Berdasarkan hasil laporan klasifikasi, performa terbaik ditunjukkan oleh model Support Vector Machine pada kelas positif. Pada kelas ini, nilai precision tercatat sebesar 0,93, nilai recall sebesar 0,99, serta f1-score sebesar 0,96. Tingginya nilai recall mengindikasikan bahwa hampir seluruh ulasan dengan sentimen positif berhasil diidentifikasi secara tepat oleh model. Kondisi tersebut berkaitan erat dengan karakteristik data uji yang didominasi oleh sentimen positif, sehingga pola linguistik pada ulasan positif dapat dipelajari dan dikenali dengan baik oleh SVM.

Pada kelas sentimen negatif, kinerja model masih tergolong terbatas. Nilai precision sebesar 0,83 dan recall sebesar 0,26 menghasilkan f1-score sebesar 0,40. Rendahnya nilai recall memperlihatkan bahwa sebagian besar ulasan negatif belum berhasil dikenali secara optimal dan masih banyak yang diklasifikasikan ke dalam kategori positif atau netral. Situasi ini dipengaruhi oleh ketidakseimbangan jumlah data, di mana ulasan negatif jumlahnya jauh lebih sedikit dibandingkan ulasan positif, sehingga kecenderungan bias terhadap kelas mayoritas tidak dapat dihindari. Kelas netral juga memperlihatkan performa yang belum maksimal, dengan nilai precision sebesar 0,60 dan recall sebesar 0,24. Nilai f1-score sebesar 0,34 mencerminkan bahwa pemisahan antara ulasan netral dan positif masih menjadi tantangan, terutama karena kemiripan pemilihan kata dan struktur kalimat yang digunakan pada kedua jenis sentimen tersebut.

Hasil confusion matrix memberikan penguatan terhadap temuan pada laporan klasifikasi. Dari total 38 ulasan negatif pada data uji, hanya 10 ulasan yang berhasil diklasifikasikan secara tepat sebagai negatif, sementara sebagian besar lainnya diprediksi sebagai sentimen positif. Pada kelas netral, dari 38 ulasan yang tersedia, hanya 9 ulasan yang dikenali dengan benar, sedangkan sisanya sebagian besar kembali terklasifikasi sebagai positif. Sebaliknya, performa yang sangat baik terlihat pada kelas positif, di mana dari 714 ulasan positif, sebanyak 707 ulasan berhasil

diklasifikasikan secara akurat. Hasil ini memperlihatkan kemampuan SVM yang sangat kuat dalam mengenali dan memprediksi sentimen positif secara konsisten.

Evaluasi lanjutan terhadap distribusi sentimen berdasarkan tema estetika dilaksanakan memakai hasil prediksi SVM pada seluruh 3.950 data ulasan. Hasil analisis memperlihatkan bahwa tema beach ambience muncul sebagai tema yang paling dominan dengan total 2.467 ulasan. Pada tema ini, sentimen positif sangat mendominasi dengan jumlah 2.373 ulasan, disusul oleh sentimen netral sebanyak 90 ulasan dan sentimen negatif yang hanya berjumlah 4 ulasan. Tema fasilitas memperlihatkan variasi sentimen yang lebih beragam, dengan 365 ulasan positif, 67 ulasan netral, dan 126 ulasan negatif. Pola ini mengindikasikan bahwa aspek fasilitas cenderung lebih sering memunculkan ketidakpuasan atau keluhan dibandingkan tema lainnya. Sementara itu, tema nature & creature sepenuhnya didominasi oleh sentimen positif, yang memperlihatkan apresiasi wisatawan terhadap kondisi alam dan keanekaragaman hayati. Tema atraksi juga memperlihatkan dominasi sentimen positif, meskipun masih ditemukan sejumlah ulasan bernada negatif yang mencerminkan adanya aspek atraksi tertentu yang perlu mendapatkan perhatian lebih lanjut.



**Gambar 6.** Grafik Distribusi Sentimen per-Tema Estetik-SVM

Grafik distribusi sentimen per tema estetika memperlihatkan hasil klasifikasi sentimen memakai metode Support Vector Machine (SVM) terhadap seluruh 3.950 data ulasan wisata biru. Grafik ini menggambarkan jumlah ulasan dengan sentimen positif, negatif, dan netral pada setiap tema estetika yang dianalisis, yaitu *beach ambience*, *fasilitas*, *nature & creature*, dan *atraksi*. Berdasarkan grafik, tema *beach ambience* merupakan tema yang paling dominan dengan jumlah ulasan terbanyak, yaitu 2.467 ulasan. Pada tema ini, sentimen positif sangat mendominasi dengan 2.373 ulasan, diikuti oleh sentimen netral sebanyak 90 ulasan dan sentimen negatif sebanyak 4 ulasan. Dominasi sentimen positif memperlihatkan bahwa aspek suasana pantai, seperti keindahan visual, kenyamanan, dan atmosfer lingkungan, menjadi faktor yang paling diapresiasi oleh wisatawan. Tema fasilitas menampilkan distribusi sentimen yang lebih beragam dibandingkan tema lainnya. Dari total 558 ulasan, terdapat 365 ulasan positif, 67 ulasan netral, dan 126 ulasan negatif. Tingginya jumlah ulasan negatif pada tema fasilitas memperlihatkan bahwa aspek fasilitas masih menjadi sumber utama keluhan wisatawan, seperti kondisi toilet, akses jalan, parkir, dan sarana pendukung lainnya. Hal ini memperlihatkan bahwa meskipun fasilitas diapresiasi, masih terdapat ruang perbaikan yang perlu diperhatikan oleh pengelola wisata. Pada tema *nature & creature*, seluruh 782 ulasan diklasifikasikan sebagai sentimen positif. Hal ini memperlihatkan bahwa wisatawan mempunyai persepsi yang sangat baik terhadap aspek alam dan lingkungan, seperti kondisi pasir, ombak, pepohonan, dan keindahan ekosistem alami. Tidak ditemukannya sentimen negatif maupun netral pada tema ini mengindikasikan bahwa pengalaman wisatawan terhadap aspek alam cenderung konsisten dan memuaskan. Tema atraksi mempunyai jumlah ulasan paling sedikit, yaitu 143 ulasan. Meskipun demikian, sentimen positif tetap mendominasi dengan 113 ulasan, sementara sentimen negatif berjumlah 30 ulasan dan tidak ditemukan sentimen netral. Hal ini memperlihatkan bahwa aktivitas dan wahana wisata umumnya dinilai positif, namun masih terdapat sebagian wisatawan yang memberikan ulasan negatif terkait pengalaman atraksi tertentu.

Secara keseluruhan, grafik distribusi sentimen per tema estetika memperlihatkan bahwa sentimen positif mendominasi hampir seluruh tema wisata biru di Lampung. Namun, variasi sentimen paling terlihat pada tema fasilitas, yang menandakan bahwa aspek ini mempunyai pengaruh besar terhadap kepuasan maupun ketidakpuasan wisatawan. Temuan ini memperlihatkan bahwa hasil klasifikasi SVM tidak hanya memberikan gambaran umum sentimen wisatawan, tetapi juga mampu mengidentifikasi tema-tema yang perlu menjadi prioritas dalam pengembangan dan perbaikan destinasi wisata.

### 3.6 Analisis Perbandingan Metode Naive Bayes dan Support Vector Machine

Dalam penelitian ini, dua algoritma klasifikasi diterapkan untuk menganalisis sentimen pada ulasan wisata biru, yaitu Naive Bayes Classifier dan Support Vector Machine. Evaluasi kinerja kedua metode dilakukan secara komprehensif

dengan mempertimbangkan beberapa indikator, meliputi tingkat akurasi, laporan klasifikasi yang mencakup nilai precision, recall, dan f1-score, analisis confusion matrix, serta distribusi sentimen berdasarkan tema estetika. Hasil pengujian menunjukkan bahwa Naive Bayes menghasilkan akurasi sebesar 90,38% pada data uji, sedangkan Support Vector Machine memperoleh akurasi yang lebih tinggi, yaitu 91,90%. Pada laporan klasifikasi, Naive Bayes menunjukkan performa yang sangat baik dalam mengklasifikasikan sentimen positif dengan f1-score sebesar 0,95. Namun demikian, kinerja model ini sangat rendah pada kelas negatif dan netral karena sebagian besar data pada kedua kelas tersebut diklasifikasikan sebagai positif. Sementara itu, Support Vector Machine juga memperlihatkan performa tinggi pada kelas positif dengan f1-score sebesar 0,96, serta mampu mengenali sentimen negatif dan netral dengan lebih baik dibandingkan Naive Bayes. Hasil analisis confusion matrix memperlihatkan bahwa kesalahan klasifikasi pada Naive Bayes didominasi oleh bias terhadap kelas mayoritas, sedangkan Support Vector Machine menunjukkan distribusi kesalahan yang lebih seimbang pada setiap kelas. Selain itu, analisis distribusi sentimen per tema estetika menunjukkan bahwa kedua metode menghasilkan pola yang relatif serupa, dengan tema *beach ambience* mendominasi jumlah ulasan bersentimen positif. Namun, hasil klasifikasi menggunakan Support Vector Machine memperlihatkan pembagian sentimen negatif dan netral yang lebih proporsional pada beberapa tema, dibandingkan dengan Naive Bayes yang cenderung mengklasifikasikan hampir seluruh ulasan sebagai sentimen positif.

**Tabel 10.** Perbandingan Model NBC dan SVM

Metrik	Naive Bayes (NBC)	Support Vector Machine (SVM)
Akurasi (Data Uji)	0.9038	0.9130
Precision – Negatif	0.50	0.83
Recall – Negatif	0.03	0.26
F1-Score – Negatif	0.05	0.40
Precision – Netral	0.50	0.60
Recall – Netral	0.03	0.24
F1-Score – Netral	0.05	0.34
Precision – Positif	0.91	0.93
Recall – Positif	1.00	0.99
F1-Score – Positif	0.95	0.96

#### 4. KESIMPULAN

Berdasarkan temuan penelitian yang telah diperoleh, gambaran persepsi wisatawan terhadap wisata biru di Provinsi Lampung berhasil diidentifikasi melalui analisis sentimen ulasan. Sebanyak 3.950 data ulasan dikompilasi dari berbagai platform daring, yaitu Twitter atau X, Instagram, serta Google Maps Review. Dari hasil pengolahan data tersebut, pola distribusi sentimen terlihat tidak proporsional karena sentimen positif muncul lebih dominan dibandingkan sentimen negatif dan netral. Pola ini memperlihatkan bahwa pengalaman wisatawan terhadap wisata biru di Lampung dinilai secara umum memuaskan. Persepsi positif tersebut terutama banyak muncul pada tema *beach ambience*, yang tercatat sebagai topik paling sering dibahas dalam ulasan pengunjung. Pengujian kinerja model klasifikasi memperlihatkan adanya perbedaan performa antara algoritma yang digunakan. Support Vector Machine memperlihatkan tingkat ketepatan klasifikasi yang lebih tinggi dibandingkan Naive Bayes Classifier. Pada data uji, tingkat akurasi sebesar 91,90 persen diperoleh oleh SVM, sedangkan NBC menghasilkan akurasi sebesar 90,38 persen. Selain mempunyai akurasi yang lebih tinggi, SVM juga mampu mengklasifikasikan variasi sentimen secara lebih stabil, terutama pada kategori negatif dan netral yang jumlah datanya lebih terbatas. Sebaliknya, hasil klasifikasi memakai NBC memperlihatkan kecenderungan berat sebelah ke kelas mayoritas, sehingga sebagian besar ulasan diarahkan ke sentimen positif. Berdasarkan keseluruhan hasil tersebut, efektivitas algoritma Support Vector Machine dinilai lebih unggul untuk diterapkan dalam analisis sentimen ulasan wisata biru di Provinsi Lampung. Temuan ini dapat dimanfaatkan sebagai dasar pengembangan kajian analisis sentimen berbasis ulasan daring, khususnya sebagai alat evaluasi dan pendukung pengambilan keputusan dalam perencanaan, pengelolaan, dan peningkatan kualitas destinasi wisata.

#### REFERENCES

- [1] H. Azizah, F. Syuhada, and Y. Sa, "Sentimen Analisis Tempat Wisata Berdasarkan Ulasan Google Maps Menggunakan Metode Naive Bayes ( Studi Kasus Bukit Merese )," *SIJ(SainsTech Innov. Journal)*, vol. 7, no. no .2, pp. 467–476, 2024, doi: <https://doi.org/10.37824/sij.v7i2.2024.753>.
- [2] A. M. Ndapamuri, D. Manongga, A. Iriani, U. Kristen, S. Wacana, and J. D. No, "Analisis Sentimen Ulasan Aplikasi Tripadvisor Dengan Metode Support Vector Machine , K-Nearest Neighbor , Dan Naive Bayes," *J. INOVTEK POLBENG - SERI Inform.*, vol. 8, no. No.1, pp. 127–140, 2023, doi: 10.35314/isi.v8i1.3260.
- [3] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment Analysis in Tourism: Capitalizing on Big Data," *Tour. Manag.*, vol. 58, no. 2, pp. 98–115, 2019, doi: [doi.org/10.1177/0047287517747753](https://doi.org/10.1177/0047287517747753).
- [4] D. S. Utami, A. Erfina, S. S. Indormasi, and U. N. Putra, "Analisis Sentimen Objek Wisata Bali Di Google Maps Menggunakan Algoritma Naive Bayes Pada Dasarnya Indonesia," *J. Sains Komput. Inform.*, vol. 6, no. 1, pp. 418–427, 2022.
- [5] H. Liu, W. Zhang, and M. Chen, "Public Sentiment and Perception of Blue Tourism Destinations on Social Media," *J.*

- Outdoor Recreat. Tour.*, vol. 44, p. 100620, 2023, doi: 10.1016/j.jort.2023.100620.
- [6] 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,” *J. Interkom*, vol. no. 16, no. 4, pp. 1–10, 2022, doi: <https://doi.org/10.35969/interkom.v16i4.192>.
- [7] D. Wicaksono and I. M. A. Agastya, “Perbandingan Algoritma Support Vector Machine , Decision Tree , Naïve Bayes , dan Neural Network dalam Klasifikasi Email,” *BITS J. Teknol. Inf.*, vol. 6, no. 4, pp. 2559–2572, 2025, doi: 10.47065/bits.v6i4.6949.
- [8] D. Siregar, F. Ladayya, N. Z. Albaqi, and B. M. Wardana, “Penerapan Metode Support Vector Machines ( SVM ) dan Metode Naïve Bayes Classifier ( NBC ) dalam Analisis Sentimen Publik Terhadap Konsep Child-free di Media Sosial Twitter,” *J. Stat. dan Apl.*, vol. 7, no. 1, pp. 93–104, 2023, doi: <https://doi.org/10.21009/JSA.07109>.
- [9] S. Heristian, M. Napiyah, and W. Erawati, “Analisis Sentimen Ulasan Pelanggan Menggunakan Algoritma Naive Bayes pada Aplikasi Gojek,” *Comput. Sci.*, vol. 5, no. 1, pp. 35–41, 2025, doi: <https://doi.org/10.31294/coscience.v5i1.7775>.
- [10] I. Kurniawan, A. L. Hananto, S. S. Hilabi, and A. Hananto, “Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter,” *J. Sist. Inf.*, vol. 10, no. 1, pp. 731–740, 2023, doi: <https://doi.org/10.35957/jatasi.v10i1.3582>.
- [11] Y. A. Singgalen, “Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier,” *Build. Informatics Technol. Sci.*, vol. 4, no. 3, pp. 1343–1352, 2022, doi: 10.47065/bits.v4i3.2486.
- [12] O. I. Gifari, M. Adha, I. R. Hendrawan, F. Freddy, S. Durrand, and A. S. Literature, “Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine,” *JIFOTECH (Journal Informayion Technol.*, vol. 2, no. 1, pp. 36–40, 2022, doi: <https://doi.org/10.46229/jifotech.v2i1.330>.
- [13] L. Budi and A. Mude, “Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 154–161, 2020, doi: [doi.org/10.33096/ilkom.v12i2.597.154-161](https://doi.org/10.33096/ilkom.v12i2.597.154-161).
- [14] W. Ningsih, B. Alfiana, and D. Wulandari, “Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter Pada Penggunaan Mobil Listrik di Indonesia,” *Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 556–562, 2024, doi: <https://doi.org/10.57152/malcom.v4i2.1253>.
- [15] V. No, Z. A. Mukharyahya, Y. P. Astuti, and O. N. Cahyani, “Perbandingan Naive Bayes dan Support Vector Machine dalam Klasifikasi Tingkat Kemiskinan di Indonesia,” *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, pp. 119–128, 2025, doi: 10.29408/edumatic.v9i1.29512.
- [16] P. Nandwani and R. Verma, “A Review on Sentiment Analysis and Emotion Detection From Text,” *Soc. Netw. Anal. Min.*, vol. 11, no. 1, p. 81, 2021, doi: 10.1007/s13278-021-00776-6.
- [17] A. Novanto, D. Indra, W. Astuti, U. M. Indonesia, and T. Rumah, “Analisis Pre-processing Sentimen Terhadap Komentar Layanan Indihome Pada Twitter,” *Lit. ilformatika Komput.*, vol. 1, no. 2, pp. 145–152, 2024.
- [18] A. M. Yolanda and R. T. Mulya, “Implementasi Metode Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Aplikasi Sayurbox di Google Play Store,” *J. Stat. Its Appl. Teach. Res.*, vol. 6, no. 2, pp. 76–83, 2024, doi: 10.35580/variensium258.
- [19] M. F. Saleh and R. Imanda, “Public Sentiment Analysis of the Free Meal Program : A Comparison of Naive Bayes and Support Vector Machine Methods on the Twitter ( X ) Social Media Platform,” *J. Appl. Informatics Comput.*, vol. 9, no. 1, pp. 131–139, 2025, doi: <https://doi.org/10.30871/jaic.v9i1.8895>.
- [20] A. F. Hidayatullah, “Pengaruh Stopword Terhadap Performa Klasifikasi Tweet Berbahasa Indonesia,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 1, pp. 1–4, 2016, doi: 10.14421/jiska.2016.11-01.
- [21] P. Ayuningtyas, S. Khomsah, T. Informatika, F. Informatika, S. Data, and F. Informatika, “Pelabelan Sentimen Berbasis Semi-Supervised Learning Menggunakan Algoritma LSTM dan GRU,” *J. Ilmu Komput. dan Inform.*, vol. 9, no. 3, pp. 217–229, 2024, doi: [doi.org/10.14421/jiska.2024.9.3.217-229](https://doi.org/10.14421/jiska.2024.9.3.217-229).