**Cyberspace**: Jurnal Pendidikan Teknologi informasi
Volume 10, Nomor 2, Oktober 2026, hal. 15 – 32
ISSN 2598-2079 (print) | ISSN 2597-9671 (online)

# ANALISIS SENTIMEN WISATAWAN TERHADAP DESTINASI WISATA UNGGULAN DI ACEH: PENGGABUNGAN ALGORITMA NAIVE BAYES DAN LOGISTIC REGRESSION BERBASIS WEB SCRAPING

**Yaumaliza Putra[1], Raihan Islamadina[2]**

[1,2]Pendidikan Teknologi Informasi, Fakultas Tarbiyah dan Keguruan, Universitas Islam Negeri Ar-Raniry, Jl. Syeikh Abdur Rauf Darussalam, Banda Aceh 23111, Indonesia
E-mail: 220212065@student.ar-raniry.ac.id

**Abstract**

Travelers' opinions posted on digital platforms, especially Google Maps, produce extensive textual information about Aceh's famous tourist spots, including Baiturrahman Grand Mosque, Aceh Tsunami Museum, and Lampuuk Beach. However, the large volume of this unstructured text data poses a challenge for tourism managers and academics to evaluate manually. This research seeks to examine visitor opinions regarding Aceh's primary tourist attractions by merging Naïve Bayes and Logistic Regression algorithms through an ensemble learning technique. Data was collected through *web scraping* techniques from Google Maps, resulting in 2,986 reviews across the three destinations. During *preprocessing*, the text underwent case folding, cleaning, tokenization, stopword removal, and stemming with Sastrawi for Indonesian text handling. TF-IDF was then applied for feature extraction. The dataset was partitioned into 80% for training and 20% for testing. A Soft Voting Classifier served as the ensemble method, integrating predictions from Naïve Bayes and Logistic Regression through averaged probability scores. To evaluate the model, a confusion matrix was utilized, measuring Accuracy, Precision, Recall, and F1-Score. Additionally, 5-fold cross-validation was applied to verify the model's stability. The Soft Voting ensemble yielded the best performance with an accuracy of 96.98%, outperforming the Naïve Bayes model with 96.65% and Logistic Regression with 96.15%. However, Logistic Regression demonstrated superior performance in detecting negative sentiment (65% vs 10%). The ensemble model also demonstrated excellent stability with a 5-fold cross-validation average score of 96.77% and a standard deviation of 0.0017. This study concludes that the Soft Voting ensemble is the most effective model for sentiment analysis of tourist reviews in Aceh, providing a reliable instrument for evaluating tourism services based on digital data.

**Keywords:** *Sentiment Analysis, Naïve Bayes, Logistic Regression, Ensemble Learning, Soft Voting*

**Abstrak**

Ulasan wisatawan yang tersebar di berbagai platform digital, terutama Google Maps, menghasilkan informasi tekstual dalam jumlah besar mengenai tempat-tempat wisata terkenal di Aceh, seperti Masjid Raya Baiturrahman, Museum Tsunami Aceh, dan Pantai Lampuuk. Meskipun demikian, banyaknya data teks yang tidak terstruktur ini

menyulitkan pengelola pariwisata dan akademisi dalam melakukan penilaian secara manual. Penelitian ini bertujuan untuk menelaah opini pengunjung terhadap objek-objek wisata unggulan di Aceh dengan menggabungkan algoritma Naive Bayes dan Logistic Regression melalui pendekatan ensemble learning. Penghimpunan data dilakukan dengan metode *web scraping* dari Google Maps yang berhasil menghimpun 2,986 ulasan dari ketiga destinasi tersebut. Pada tahap preprocessing, teks menjalani serangkaian proses seperti *case folding, cleaning, tokenisasi*, pembersihan kata-kata umum (*stopword*), serta stemming menggunakan Sastrawi untuk penanganan teks berbahasa Indonesia. Selanjutnya, metode TF-IDF diterapkan untuk melakukan ekstraksi fitur. *Dataset* kemudian dibagi menjadi 80% untuk keperluan latih dan 20% untuk uji. Metode *ensemble* yang digunakan adalah Soft Voting Classifier, yang memadukan hasil prediksi dari Naïve Bayes dan Logistic Regression berdasarkan rata-rata skor probabilitas. Evaluasi model dilakukan dengan memanfaatkan *confusion matrix* yang mengukur metrik Akurasi, Presisi, Recall, dan F1-Score. Selain itu, validasi silang 5-fold diterapkan untuk menguji kestabilan model. Model Soft Voting ensemble menunjukkan performa terbaik dengan akurasi mencapai 96,98%, melampaui model Naïve Bayes yang memperoleh 96,65% dan Logistic Regression dengan 96,15%. Namun, Logistic Regression lebih unggul dalam mendeteksi sentimen negatif (65% vs 10%). Model *ensemble* ini juga memperlihatkan stabilitas yang sangat baik dengan rata-rata skor validasi silang 5-fold sebesar 96,77% dan simpangan baku 0,0017. Penelitian ini menyimpulkan bahwa model Soft Voting *ensemble* merupakan pendekatan paling efektif untuk analisis sentimen ulasan wisata di Aceh, sekaligus menyediakan instrumen yang andal bagi evaluasi layanan pariwisata berbasis data digital.

**Kata kunci:** *Analisis Sentimen, Naïve Bayes, Logistic Regression, Ensemble Learning, Soft Voting*

## 1. Pendahuluan

Aceh memiliki sejumlah objek wisata andalan, di antaranya Masjid Raya Baiturrahman, Museum Tsunami Aceh, dan Pantai Lampuuk. Ketiga destinasi ini menjadi ikon pariwisata yang menjadi tujuan favorit wisatawan domestik dan mancanegara. [1]. Seiring dengan perkembangan teknologi digital, Para pengunjung kerap menuangkan pengalaman mereka dalam bentuk ulasan di situs seperti Google Maps. Ulasan-ulasan ini membentuk kumpulan data teks yang besar (*big data*) yang dapat mencerminkan kualitas layanan dan kepuasan pengunjung [2]. Namun, volume data yang besar menjadi kendala bagi pengelola destinasi dan akademisi untuk melakukan evaluasi secara manual [3].

Salah satu teknik pemrosesan bahasa alami (*Natural Language Processing*/NLP) ialah analisis sentimen, yang berfungsi untuk menggali serta mengelompokkan opini-opini subjektif dari teks menjadi data yang lebih terstruktur [4]. Dalam industri pariwisata, analisis sentimen telah banyak digunakan untuk memahami polaritas emosi wisatawan terhadap suatu destinasi, apakah cenderung positif, negatif, atau netral [5]. Penelitian terdahulu membuktikan bahwa algoritma Naïve Bayes mampu memetakan opini publik terhadap sektor pariwisata secara efektif di Aceh pada media sosial X dengan tingkat akurasi mencapai 81%. Namun, penelitian tersebut hanya berfokus pada satu algoritma

dan yang bersumber dari media sosial dengan tingkat gangguan (noise) yang cukup besar [1].

Studi komparatif pada data ulasan aplikasi MyPertamina menunjukkan bahwa Logistic Regression memperoleh akurasi 86%, sedangkan Naïve Bayes hanya mencapai 81%.[6]. Sementara itu, Ada juga penelitian yang menunjukkan bahwa kombinasi TF-IDF dengan Logistic Regression mencapai akurasi tertinggi sebesar 90,4% pada data ulasan marketplace berbahasa Indonesia[7]. Hasil-hasil ini mengindikasikan bahwa Logistic Regression terbukti lebih andal dan presisi ketika diterapkan pada data teks yang rumit, sementara Naïve Bayes lebih efisien untuk data teks skala besar.

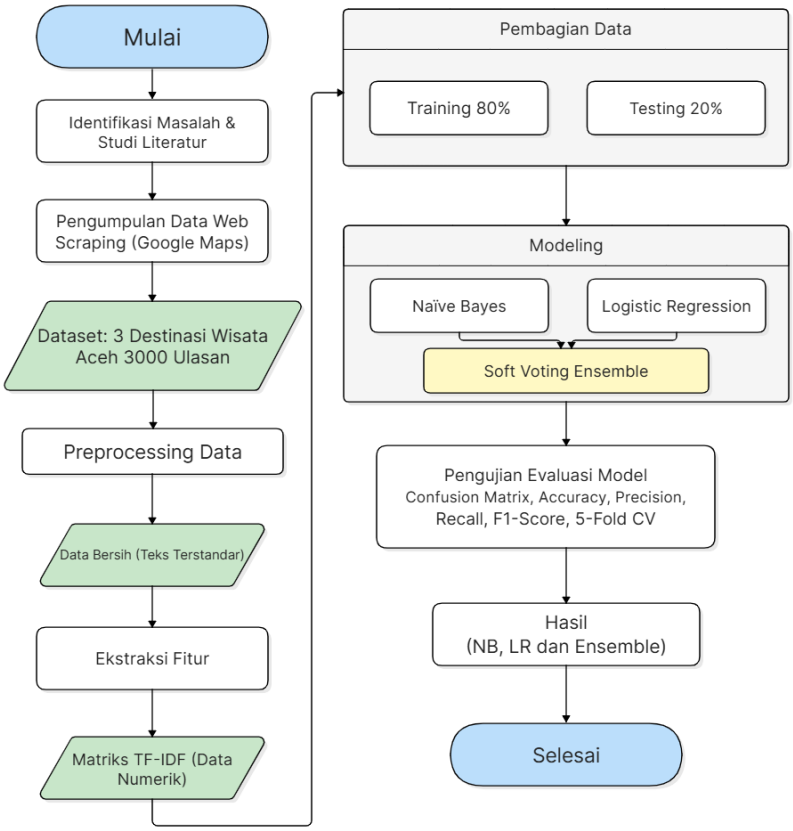
Meskipun kedua algoritma telah banyak diteliti secara terpisah atau dibandingkan, penelitian yang menggabungkan kedua algoritma menggunakan pendekatan ensemble learning khusus untuk analisis sentimen wisata Aceh masih terbatas. Pendekatan ensemble learning terbukti mampu meningkatkan kinerja model dengan menggabungkan kelebihan dari masing-masing algoritma[8]. Penelitian lain menunjukkan bahwa penggabungan SVM dan Naïve Bayes menggunakan Soft Voting berhasil meningkatkan akurasi dari 83,40% menjadi 93,44%[9]. Demikian pula, penelitian terdahulu membuktikan bahwa *Voting Classifier* pada analisis sentimen ulasan aplikasi mencapai akurasi 85%, mengungguli model individual [10].

Tantangan utama dalam analisis sentimen ulasan wisata di Aceh adalah karakteristik bahasa yang digunakan wisatawan, yang sering mencampuradukkan bahasa Indonesia formal, bahasa informal, hingga istilah lokal [1]. Kondisi ini menuntut penggunaan algoritma yang tidak hanya cepat tetapi juga stabil dalam menghadapi variasi bahasa. Penelitian terdahulu yang menelaah sentimen terhadap tempat wisata di Aceh Tengah dengan memakai Naïve Bayes dan *Support Vector Machine*, penelitian tersebut mendapati bahwa sentimen negatif lebih mendominasi, namun akurasi Naïve Bayes masih perlu ditingkatkan[11].

Tujuan dari penelitian ini adalah untuk menelaah opini pengunjung terhadap objek wisata unggulan di Aceh dengan memadukan algoritma Naïve Bayes dan Logistic Regression melalui pendekatan ensemble learning berbasis Soft Voting. Proses pengambilan data dilaksanakan melalui teknik *web scraping* di Google Maps pada tiga lokasi wisata di Aceh. Penelitian ini diinginkan sanggup menyumbangkan model klasifikasi sentimen yang tambah presisi serta konsisten, serta menjadi referensi bagi pengelola pariwisata Aceh guna memperbaiki mutu layanan yang didukung oleh data digital.

## 2. Metode Penelitian

Metode eksperimen komparatif dengan pendekatan kuantitatif diterapkan dalam penelitian ini untuk menguji dan menggabungkan kinerja Naive Bayes dan Logistic Regression dalam menganalisis sentimen pengunjung[6]. Secara keseluruhan, alur penelitian terdiri dari beberapa tahapan utama: Tahapan yang dilakukan mencakup pengambilan data melalui web scraping, pra-pemrosesan teks, ekstraksi fitur dengan TF-IDF, pembagian dataset, pelatihan serta penggabungan model, serta evaluasi performa. Gambar 1 menyajikan diagram alir penelitian secara berurutan[10].




Gambar 1. Flowchart Tahapan Penelitian

## 2.1 Identifikasi Masalah dan Studi Literatur

Penelitian ini dilatarbelakangi oleh masih terbatasnya pemanfaatan data ulasan wisatawan dari platform digital untuk evaluasi layanan pariwisata di Aceh secara otomatis dan objektif. Kendati jumlah ulasan yang ada sangat melimpah, pengelola destinasi dan akademisi masih mengandalkan metode manual yang tidak efisien dalam mengekstraksi informasi berharga dari data tekstual tersebut [1]. Beberapa penelitian sebelumnya telah menunjukkan efektivitas analisis sentimen dalam memetakan persepsi publik terhadap pariwisata Aceh [1], namun kebanyakan masih sekadar memakai satu jenis algoritma dan belum mengoptimalkan pendekatan ensemble learning untuk meningkatkan akurasi klasifikasi [9]. Selain itu, studi yang secara khusus mengkaji tentang penggabungan Naïve Bayes dan Logistic Regression pada data ulasan wisata Aceh dari Google Maps masih sangat terbatas, sehingga penelitian ini diperlukan untuk mengisi celah tersebut.

Kajian literatur memperlihatkan bahwa riset analisis sentimen telah tumbuh pesat dengan beragam pendekatan algoritma, mulai dari Naïve Bayes yang efisien untuk data teks skala besar [1], Logistic Regression yang lebih stabil dalam menangani keterkaitan antar fitur [6], hingga metode ensemble learning yang terbukti mampu meningkatkan akurasi melalui penggabungan model [8]. Penelitian terdahulu yang berhasil menerapkan Naïve Bayes pada data pariwisata Aceh dengan akurasi 81%[1], sementara peneliti terdahulu membuktikan bahwa Logistic Regression mengungguli Naïve Bayes dengan selisih 5% pada data ulasan aplikasi[6]. Hasil Penelitian lain juga

memperlihatkan bahwa metode Soft Voting berhasil menaikkan akurasi dari 83,40% menjadi 93,44% pada analisis sentimen[9]. Namun, penelitian yang mengkombinasikan Naïve Bayes dan Logistic Regression dengan pendekatan ensemble learning untuk wisata Aceh dari data Google Maps belum ditemukan, sehingga penelitian ini hadir untuk mengisi kekosongan tersebut dengan memanfaatkan *web scraping* sebagai teknik pengumpulan data [12] serta memanfaatkan TF-IDF untuk proses ekstraksi fitur[7].

## 2.2 Dataset

Penelitian ini memanfaatkan ulasan pengunjung dari tiga lokasi wisata andalan di Provinsi Aceh, yakni Masjid Raya Baiturrahman, Museum Tsunami Aceh, dan Pantai Lampuuk. Penghimpunan data dilakukan dengan memakai metode *web scraping* di platform Google Maps yang dijalankan menggunakan Python dengan bantuan *library Selenium* dan *BeautifulSoup* [2]. Pendekatan ini diambil karena mampu mengekstrak data secara otomatis dan cepat dalam skala besar[12].

Proses *web scraping* dilakukan dengan langkah-langkah sebagai berikut: pertama, identifikasi URL resmi masing-masing destinasi pada Google Maps. Kedua, mengakses halaman destinasi dan menavigasi ke tab ulasan (reviews). Ketiga, melakukan scrolling otomatis untuk memuat seluruh ulasan yang tersedia. Keempat, mengekstrak elemen HTML yang berisi nama pengulas, rating bintang, teks ulasan, dan tanggal ulasan. Kelima, menyimpan data ke dalam format *Comma Separated Values* (CSV) dan XLSX untuk memudahkan proses analisis selanjutnya [9].

Total ulasan yang berhasil dikumpulkan adalah sebanyak 2.986 ulasan dengan rincian per destinasi disajikan pada Tabel 1. Ulasan yang diambil dibatasi pada rentang waktu 2021–2026 dan menggunakan Bahasa Indonesia. Untuk memastikan kualitas data, dilakukan penanganan sebagai berikut: ulasan duplikat (teks ulasan yang sama pada destinasi yang sama) dihapus untuk menghindari bias, ulasan yang tidak lengkap (tanpa teks atau tanpa rating) dikeluarkan dari dataset, dan ulasan dalam bahasa asing atau yang hanya berisi emoji juga difilter pada tahap ini [11].

Tabel 1. Distribusi Data Ulasan per Destinasi

| Destinasi Wisata | Jumlah Ulasan |
|---|---|
| Masjid Raya Baiturrahman | 1.030 |
| Museum Tsunami Aceh | 976 |
| Pantai Lampuuk | 980 |
| **Total** | **2.986** |

## 2.3 *Preprocessing*

Data ulasan yang diperoleh dari *web scraping* masih mengandung elemen pengganggu seperti tanda baca, angka, emoticon, kata tidak baku, dan karakter khusus lainnya. Dengan demikian, tahap *preprocessing* menjadi langkah penting untuk membersihkan dan menyeragamkan teks sebelum diproses oleh algoritma klasifikasi[4]. Studi terdahulu membuktikan bahwa mutu *preprocessing* berdampak besar terhadap tingkat akurasi model, di mana teks yang tidak melalui pembersihan optimal dapat menurunkan performa klasifikasi [7].

Tahap *preprocessing* dalam penelitian ini mengikuti standar *pipeline* yang umum digunakan dalam analisis sentimen teks berbahasa Indonesia [1][9], yang terdiri dari lima tahapan utama. Tabel 2 merangkum seluruh langkah dalam tahap *preprocessing*.

Tabel 2. Ringkasan Tahapan *Preprocessing*

| Langkah | Fungsi | Contoh Input | Contoh Output |
|---|---|---|---|
| ***Case Folding*** | Menyeragamkan huruf menjadi huruf kecil | "Pantai Lampuuk SANGAT Indah" | "pantai lampuuk sangat indah" |
| ***Cleaning*** | Menghapus *noise* (tanda baca, angka, URL, emoticon) | "Pantai, indah! 😊" | "Pantai indah" |
| ***Tokenization*** | Memisahkan kalimat menjadi setiap kata | "pantai lampuuk indah" | ["pantai", "lampuuk", "indah"] |
| ***Stopword Removal*** | Membuang kata-kata yang tidak memberi kontribusi makna | ["pantai", "dan", "indah"] | ["pantai", "indah"] |
| ***Stemming*** | Mengembalikan kata pada bentuk dasar kata tersebut | "mengunjungi", "keindahan" | "kunjung", "indah" |

*Case Folding* merupakan tahap pertama yang Menyeragamkan huruf pada teks menjadi huruf kecil (*lowercase*). Langkah ini bertujuan untuk menyeragamkan bentuk huruf kapital dan huruf kecil agar kata "Indah", "INDAH", dan "indah" dianggap sebagai kata yang sama [1]. Tahap kedua adalah *Cleaning* yang berfungsi membersihkan teks dari elemen-elemen pengganggu, misalnya tanda baca, angka, URL, mention pengguna, serta emoticon [11]. Proses ini menggunakan ekspresi reguler (*regex*) untuk membersihkan teks secara otomatis.

Tahap ketiga Tokenization adalah proses memisahkan kalimat menjadi bagian-bagian kecil yang dinamakan token (kata per kata). Langkah ini penting karena algoritma klasifikasi bekerja berdasarkan kemunculan kata-kata individual dalam dokumen [6]. Tokenisasi dilakukan menggunakan fungsi bawaan Python dengan pemisahan berdasarkan spasi.

Tahap keempat adalah *Stopword Removal* bertugas Membuang kata-kata yang tidak memberi kontribusi makna yang berarti terhadap sentimen, misalnya "dan", "atau", "yang", "di", "ke", serta "dari"[7]. Tujuan dari penghapusan stopword adalah untuk menekan jumlah fitur dan mempercepat proses komputasi. Daftar *stopword* yang digunakan adalah kombinasi dari daftar *stopword* bahasa Indonesia NLTK dan tambahan kata-kata informal seperti "gak", "ga", "sih", dan "banget" [1].

Langkah kelima disebut Stemming, yakni proses mengembalikan kata berimbuhan ke akar katanya dengan membuang afiks (awalan, akhiran, maupun sisipan). Pada teks berbahasa Indonesia, proses ini menggunakan Sastrawi sebagai alat bantu stemming[9].

Contoh penerapan *stemming* antara lain kata "mengunjungi" menjadi "kunjung", "keindahan" menjadi "indah", dan "pelayanan" menjadi "layan"[11]. Implementasi seluruh tahap *preprocessing* Seluruh proses tersebut dijalankan dengan Python di VSCode. Output dari tahapan ini adalah teks yang telah dibersihkan dan siap diekstraksi fiturnya melalui TF-IDF pada tahap berikutnya [7].

## 2.4 Ekstraksi Fitur

Setelah melewati tahap *preprocessing*, langkah berikutnya ialah ekstraksi fitur bertujuan untuk mengonversi data teks dari format kategorikal ke dalam bentuk numerik supaya dapat diolah oleh algoritma *machine learning*[4]. TF-IDF (Term Frequency-Inverse Document Frequency) dipakai pada penelitian ini sebagai pendekatan ekstraksi fitur.

Metode ini merupakan teknik pembobotan kata yang lazim dipakai di bidang temu kembali informasi dan penggalian teks[7]. Metode ini berfungsi guna menilai derajat urgensi suatu kata di dalam sebuah dokumen jika dibandingkan dengan sekumpulan dokumen atau korpus ulasan wisatawan [1]. Nilai TF-IDF dihasilkan dari perkalian dua statistik, yaitu TF-IDF didapat dari hasil perkalian antara *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF)[7]. Adapun formulasi matematis dari TF-IDF adalah sebagai berikut[4]:

$$W_{i,j} = tf_{i,j} \times log\left(\frac{N}{df_i}\right) \quad (1)$$

Pada persamaan (1), *W(i,j)* melambangkan bobot dari kata *i* pada dokumen *j*, yang didapat dari perkalian *tf(i,j)* (frekuensi kata *i* di dokumen *j*) dengan *log(N/df(i))*, di mana *N* adalah total dokumen dan *df(i)* adalah banyaknya dokumen yang memuat kata *i*. Cara kerja TF-IDF adalah memberi bobot besar pada kata yang kerap muncul dalam satu dokumen tetapi jarang ditemukan di dokumen lainnya. Sebaliknya, kata yang hadir di banyak dokumen diberi bobot kecil karena nilai informasinya rendah[1]. Dengan pendekatan ini, sistem dapat mengenali kata-kata yang paling mewakili isi setiap ulasan wisatawan[11].

Dalam analisis sentimen destinasi wisata di Aceh, penggunaan TF-IDF memungkinkan sistem untuk menekan bobot kata-kata yang sering muncul di hampir seluruh ulasan (misalnya kata penghubung) dan memberi bobot lebih besar pada kata-kata yang bermakna khusus dan informatif (seperti "indah", "bersih", "nyaman", atau "macet")[13]. Ulasan wisatawan biasanya memuat istilah-istilah penting yang merefleksikan pengalaman kunjungan mereka ke suatu tempat, sehingga hal ini menjadi penting[1].

Proses TF-IDF pada penelitian ini dikerjakan dengan Scikit-learn di Google Colab. Parameter yang digunakan meliputi max_features=5000 untuk membatasi jumlah fitur, ngram_range=(1,2) untuk menangkap unigram maupun bigram, serta min_df=2 dan max_df=0.95 agar kata yang terlalu jarang atau terlalu sering muncul tidak diikutsertakan[7]. Tahap ekstraksi fitur ini menghasilkan matriks TF-IDF yang kelak menjadi masukan bagi proses klasifikasi[9].

## 2.5 Data Splitting

Setelah proses ekstraksi fitur TF-IDF menghasilkan matriks numerik, Tahap selanjutnya ialah memisahkan data menjadi dua kelompok, yakni data latih (training data) dan data uji (testing data)[6]. Pemisahan ini bertujuan agar model dapat belajar

dari satu set data, lalu diuji dengan data yang belum pernah dilihat sebelumnya, sehingga kemampuannya dalam mengenali pola pada data baru dapat terukur[10].

Data dibagi dengan perbandingan 80:20, di mana 80% dialokasikan sebagai keperluan latih, sedangkan 20% untuk uji[6]. Rasio ini lazim dipakai dalam studi analisis sentimen karena mampu menyeimbangkan porsi data untuk pelatihan model dan pengujian performa[2]. Total data yang digunakan adalah 2.986 ulasan, Dengan demikian, diperoleh 2.389 ulasan untuk keperluan latih dan 597 ulasan untuk uji.

Tabel 3. Pembagian Data Training dan Testing

| Jenis Data | Jumlah Ulasan | Positif | Negatif |
|---|---|---|---|
| Data Latih (Training) | 2.389 | 2.309 | 80 |
| Data Uji (Testing) | 597 | 577 | 20 |
| **Total** | **2.986** | **2.886** | **100** |

Pembagian data dikerjakan dengan memakai fungsi train_test_split milik *Scikit-learn* serta menerapkan random_state=42 agar hasilnya bisa direproduksi[6]. Selain itu, parameter stratify=y diterapkan agar komposisi kelas pada data latih dan uji tetap seimbang, mengingat data yang digunakan memiliki ketidakseimbangan kelas (*imbalance data*) dengan dominasi sentimen positif [2].

## 2.6 Klasifikasi Model

Pada tahap ini, dua algoritma klasifikasi digunakan secara individual sebelum digabungkan dalam metode ensemble.

### 2.6.1 Naive Bayes

Algoritma pertama yang digunakan adalah *Naïve Bayes Classifier* (NBC), yaitu model probabilistik yang bertumpu pada Teorema Bayes dengan anggapan bahwa setiap fitur bersifat independen satu sama lain[1]. Pada analisis sentimen, NBC menghitung probabilitas sebuah dokumen masuk ke dalam kelas sentimen tertentu dengan melihat kata-kata yang muncul di dalamnya[11]. Keunggulan utama NBC adalah efisiensi komputasi dan kemampuannya menangani data teks berdimensi tinggi [6]. Persamaan dasar Teorema Bayes dituliskan sebagai berikut [1]:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \quad (2)$$

Pada persamaan (2), *P(H|E)* menyatakan probabilitas hipotesis *H* setelah mempertimbangkan bukti *E (posterior)*. Sementara itu, *P(E|H)* adalah probabilitas munculnya bukti *E* apabila hipotesis *H* benar *(likelihood), P(H)* adalah probabilitas awal hipotesis *H* sebelum bukti diketahui *(prior*), dan *P(E)* merupakan probabilitas total dari bukti *E*. Dalam penelitian ini, NBC diimplementasikan dengan MultinomialNB dari *Scikit-learn*, yang sesuai untuk data diskrit seperti hasil TF-IDF [6].

### 2.6.2 Logistic Regression

Algoritma kedua adalah *Logistic Regression (LR)* yang merupakan algoritma klasifikasi untuk memprediksi probabilitas variabel terikat kategorikal [6]. Berbeda dengan NBC yang berbasis probabilitas murni, LR menggunakan fungsi logistik *(sigmoid)* untuk mengubah nilai input ke dalam rentang 0 hingga 1[7]. Fungsi sigmoid didefinisikan sebagai berikut [6]:

$$g(z) = \frac{1}{1+e^{-z}} \quad (3)$$

Pada persamaan (3), *g(z)* adalah probabilitas hasil (antara 0 dan 1), *e* adalah bilangan Euler (≈2,718), dan *z* merupakan kombinasi linear antara bobot dan fitur input. Keunggulan utama LR adalah kemampuannya menangani hubungan antar fitur secara lebih fleksibel dibandingkan NBC, sehingga sering memberikan akurasi lebih tinggi pada data teks yang kompleks [7]. Implementasi LR menggunakan *library Scikit-learn* dengan opsi solver='lbfgs', max_iter=1000, serta class_weight='balanced' guna mengatasi ketimpangan distribusi kelas [6].

## 2.7 Ensemble Learning

Setelah kedua model individual dilatih, langkah selanjutnya adalah menggabungkan keduanya menggunakan pendekatan *ensemble learning* dengan metode *Soft Voting Classifie* [9]. Metode ini bekerja dengan merata-ratakan probabilitas prediksi dari setiap model dan memilih kelas dengan probabilitas rata-rata tertinggi sebagai hasil akhir [8]. Tidak seperti *Hard Voting* yang sekadar mengandalkan suara mayoritas, Soft Voting turut mempertimbangkan tingkat kepercayaan *(confidence)* masing-masing model saat menghasilkan prediksi[14].

Keunggulan Soft Voting pada data tidak seimbang (*imbalance*) adalah model yang lebih yakin (misalnya LR pada data negatif) dapat memberikan pengaruh lebih besar dibandingkan model yang kurang yakin [9]. Penelitian oleh Saputra dan Suria [9] membuktikan bahwa Soft Voting berhasil meningkatkan akurasi dari 83,40% menjadi 93,44% pada analisis sentimen. Implementasi ensemble dalam penelitian ini menggunakan *VotingClassifier* dari *library Scikit-learn* dengan parameter voting='soft' [10]. Model ensemble yang dihasilkan kemudian dievaluasi untuk menentukan peningkatan kinerja dibandingkan model individual [8].

## 2.8 Pengujian Evaluasi

Untuk mengevaluasi performa model, digunakan *Confusion Matrix* yang membandingkan hasil predik model terhadap nilai aktual pada data uji[6]. Matriks ini menghasilkan empat nilai pokok, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN)[1]. Dari keempat nilai tersebut, kemudian dihitung metrik *Accuracy, Precision, Recall,* dan *F1-Score* sebagai ukuran performa model secara menyeluruh[2].

*Accuracy* menunjukkan proporsi prediksi benar terhadap total data uji, *Precision* mengukur keakuratan prediksi pada kelas positif, *Recall* mencerminkan kemampuan model dalam mengenali data positif, sedangkan *F1-Score* merupakan rata-rata

harmonik dari *Precision* dan *Recall*[6]. Rumus keempat metrik tersebut disajikan pada persamaan (4) hingga (7) [7]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (7)$$

Keempat indikator pada persamaan (4) hingga (7) yakni TP, TN, FP, dan FN didefinisikan sebagai berikut: TP terjadi saat model tepat mengklasifikasikan data positif ke dalam kelas positif; TN berarti data negatif berhasil dikenali sebagai negatif; FP muncul ketika data negatif justru dikategorikan sebagai positif secara keliru; sedangkan FN terjadi apabila data positif salah ditempatkan ke dalam kelas negatif. Nilai-nilai ini diperoleh dari Confusion Matrix, yang membandingkan prediksi model terhadap data aktual pada proses pengujian[6].

Selain itu, agar model terhindar dari *overfitting* serta mampu beradaptasi dengan data baru, Metode *5-Fold Cross Validation* diterapkan dalam penelitian ini[9]. Metode ini mempartisi data latih menjadi 5 bagian *(fold)* yang berukuran sama, kemudian setiap bagian bergantian menjadi satu bagian dijadikan data uji, sementara empat sisanya dipakai sebagai data latih[10]. Hasil akhir adalah rata-rata akurasi dari 5 putaran tersebut, yang memberikan indikasi stabilitas model [8]. Evaluasi dilakukan pada ketiga model: Naïve Bayes individual, Logistic Regression individual, dan model *Ensemble Soft Voting*, untuk menentukan model terbaik dalam mengklasifikasikan sentimen wisatawan terhadap destinasi wisata unggulan di Aceh [6].

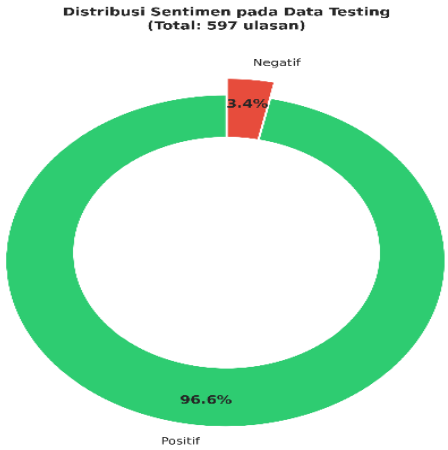
## 3. Hasil dan Pembahasan

Bagian ini memaparkan temuan eksperimen beserta pembahasannya. Hasil yang dipaparkan mencakup statistik dataset, analisis fitur TF-IDF, performa model, confusion matrix, validasi silang, serta diskusi menyeluruh terhadap temuan penelitian.

### 3.1 Data Statistics

Distribusi data ulasan per destinasi telah disajikan pada Tabel 1. Distribusi Data Ulasan per Destinasi di sub bab 2.2. Berdasarkan data tersebut, total ulasan yang berhasil dikumpulkan melalui proses *web scraping* adalah 2.986 ulasan dari tiga destinasi wisata unggulan di Aceh, yaitu Masjid Raya Baiturrahman, Museum Tsunami Aceh, dan Pantai Lampuuk. Ketiga destinasi memiliki distribusi ulasan yang relatif seimbang, dengan selisih kurang dari 2% antara destinasi terbanyak dan terkecil, yang mengindikasikan tingkat popularitas yang sebanding di kalangan wisatawan yang memberikan ulasan pada Google Maps [2].

Setelah melalui proses pelabelan sentimen berdasarkan rating bintang, dengan ketentuan rating 4–5 digolongkan ke dalam sentimen positif, sementara rating 1–3 masuk ke dalam sentimen negatif[6], data ulasan menunjukkan dominasi sentimen positif yang sangat signifikan. Dari total 2.986 ulasan, sebanyak 2.886 ulasan (96,6%) tergolong sentimen positif dan hanya 100 ulasan (3,4%) yang tergolong sentimen negatif. Distribusi sentimen ini disajikan pada Gambar 2. Dominasi sentimen positif ini mencerminkan tingkat kepuasan wisatawan yang tinggi terhadap ketiga destinasi wisata Aceh, hal ini sejalan dengan studi sebelumnya yang mengungkapkan bahwa wisatawan cenderung memberi ulasan positif pada tempat wisata yang sarat akan nilai sejarah dan budaya[1].




Gambar 2. Distribusi Sentimen pada Data Ulasan Wisata Aceh

### 3.2 Hasil Seleksi Ekstraksi Fitur TF-IDF

Usai menjalani tahap ekstraksi fitur dengan TF-IDF, diperoleh bobot untuk setiap kata yang muncul dalam korpus ulasan wisata Aceh. Semakin tinggi bobot TF-IDF suatu kata, semakin besar pula tingkat representasi dan informasinya dalam membedakan antar dokumen[7]. Sepuluh kata dengan skor TF-IDF tertatas dapat dilihat pada Tabel 4.

Tabel 4. Top Kata dengan bobot TF-IDF Teratas

| Ranking | Kata | Bobot TF-IDF |
|---|---|---|
| 1 | aceh | 0,04099 |
| 2 | pantai | 0,03936 |
| 3 | masjid | 0,03656 |
| 4 | tsunami | 0,03305 |
| 5 | indah | 0,03261 |
| 6 | bagus | 0,02887 |
| 7 | bersih | 0,02472 |
| 8 | museum | 0,02393 |
| 9 | banda | 0,01936 |
| 10 | masjid | 0,01899 |

Berdasarkan Tabel 4, kata "aceh" memiliki bobot TF-IDF tertinggi (0,04099), diikuti oleh "pantai" (0,03936), "masjid" (0,03656), dan "tsunami" (0,03305).

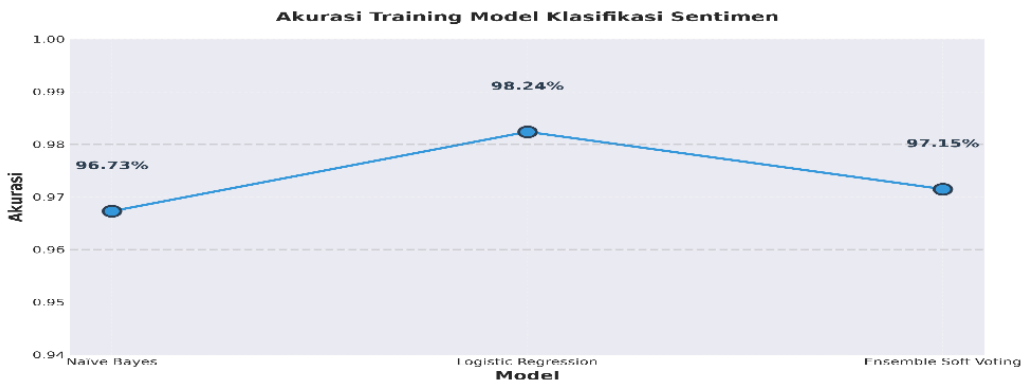
Keempat kata ini merupakan kata kunci yang secara langsung merujuk pada lokasi dan ikon destinasi wisata Aceh, yang menunjukkan bahwa ulasan wisatawan sangat terkait dengan identitas dan ciri khas masing-masing destinasi [6]. Kata "indah" dan "bagus" yang berada pada peringkat kelima dan keenam dengan bobot 0,03261 dan 0,02887 menunjukkan bahwa wisatawan sering mengekspresikan sentimen positif terhadap keindahan destinasi wisata Aceh [11]. Kata "bersih" (0,02472) dan "nyaman" (0,01638, tidak masuk top 10) juga mencerminkan aspek kenyamanan dan kebersihan yang menjadi perhatian wisatawan [13].

Frasa *bigram* yang berhasil ditangkap melalui parameter ngram_range=(1,2) juga menunjukkan hasil yang informatif. Frasa seperti "banda aceh", "tsunami aceh", dan "pantai indah" muncul dengan bobot yang signifikan. Hal ini penting karena frasa dapat menangkap konteks yang lebih luas dibandingkan kata tunggal (*unigram*), seperti "pantai indah" yang secara langsung mengindikasikan sentimen positif terhadap pantai, atau "museum tsunami" yang merujuk pada salah satu destinasi utama [9]. Kemunculan kata "mesjid" sebagai varian ejaan dari "masjid" dengan bobot 0,01899 menunjukkan bahwa wisatawan menggunakan variasi bahasa dalam ulasan mereka, yang mengindikasikan bahwa proses *preprocessing* telah berhasil mempertahankan kata-kata dengan makna penting meskipun terdapat variasi ejaan [1].

### 3.3 Model Performance Training

Pada tahap ini, Pengujian terhadap data training dilakukan pada ketiga model untuk menilai kemampuannya dalam menangkap pola dari data yang dilatihkan. Hasil evaluasi pada data training disajikan pada Tabel 5. Berdasarkan Tabel 5, Hasil evaluasi pada data training memperlihatkan bahwa seluruh model memiliki kinerja yang sangat baik, dengan akurasi di atas 96,50%. Logistic Regression mencapai akurasi training tertinggi sebesar 98,24%, diikuti oleh *Ensemble Soft Voting* (97,15%) dan Naïve Bayes (96,73%).

Hal ini menandakan bahwa ketiga model mampu menyerap pola-pola dari data training secara optimal. meskipun Logistic Regression menunjukkan kecenderungan *overfitting* yang lebih tinggi dibandingkan kedua model lainnya [6]. *Ensemble Soft Voting* menunjukkan keseimbangan yang baik antara akurasi training (97,15%) dan testing (96,98%) dengan selisih hanya 0,17%, Kondisi tersebut juga mencerminkan bahwa model-model tersebut mempunyai kemampuan generalisasi yang memadai[9].




Gambar 3. Hasil Akurasi Training

### 3.4 Model Performance Testing

Tahap ini memuat penilaian performa terhadap tiga model yang sudah dikembangkan, yakni Naïve Bayes, Logistic Regression, dan *Ensemble Soft Voting*. Proses evaluasi memakai data uji berjumlah 597 ulasan dengan metrik *Accuracy, Precision, Recall, serta F1-Score*[6]. Tabel 5 menyajikan rekapitulasi hasil evaluasi dari ketiga model tersebut.

Tabel 5. Perbandinagn Kinerja Model

| Model | *Accuracy* | *Precision* | *Recall* | *F1-Score* |
|---|---|---|---|---|
| **Naive Bayes** | 96,65% | 96,65% | 100,00% | 98,30% |
| **Logistic Regression** | 96,15% | 98,77% | 97,23% | 97,99% |
| **Ensemble Soft Voting** | 96,98% | 96,97% | 100,00% | 98,46% |

Berdasarkan Tabel 5, model *Ensemble Soft Voting* terbaik yang dimana mencapai akurasi tertinggi sebesar 96,98%, mengungguli *Naïve Bayes* (96,65%) dan *Logistic Regression* (96,15%). Hal ini menunjukkan bahwa penggabungan kedua model melalui pendekatan Soft Voting berhasil meningkatkan kinerja dibandingkan model individual [9]. Peningkatan akurasi sebesar 0,33% dari *Naïve Bayes* dan 0,83% dari *Logistic Regression* membuktikan bahwa metode ensemble efektif dalam mengkombinasikan kelebihan masing-masing algoritma [8].

Dari sisi *Recall*, baik Naïve Bayes maupun *Ensemble Soft Voting* mencapai nilai sempurna (100%), yang berarti kedua model mampu mendeteksi seluruh ulasan positif dengan benar tanpa ada yang terlewat [6]. Sementara itu, Logistic Regression memiliki *Recall* 97,23%, yang berarti masih terdapat 16 ulasan positif yang keliru diprediksi sebagai negatif. Nilai *Precision* tertinggi dicapai oleh Logistic Regression (98,77%), diikuti oleh *Ensemble Soft Voting* (96,97%) dan Naïve Bayes (96,65%) [7]. Nilai *F1-Score*, sebagai rata-rata harmonik dari *Precision* dan *Recall*, memperlihatkan bahwa *Ensemble Soft Voting* memiliki nilai tertinggi (98,46%), diikuti Naïve Bayes (98,30%) dan Logistic Regression (97,99%) [2].

Temuan ini memperkuat hasil studi terdahulu yang menyatakan bahwa metode ensemble melalui Soft Voting mampu menaikkan akurasi dari 83,40% menjadi 93,44% pada analisis sentimen [9]. Penelitian lain juga membuktikan bahwa penggabungan model melalui Voting Classifier berhasil mencapai akurasi 85%, mengungguli model individual [10]. Dengan demikian, dapat disimpulkan bahwa model *Ensemble Soft Voting* merupakan model terbaik dalam rangka menganalisis sentimen pengunjung terhadap objek wisata unggulan di Aceh. Namun, Logistic Regression lebih unggul dalam mendeteksi sentimen negatif (65% vs 10%).

Selain metrik akurasi keseluruhan, evaluasi per kelas juga dilakukan untuk melihat kemampuan model dalam mengenali masing-masing kelas sentimen. Hasil evaluasi per kelas disajikan pada Tabel 6.

Tabel 6. Evaluasi Per Kelas pada Data Testing

| Metrik | Naive Bayes | Logistic Regression | Ensemble Soft Voting |
|---|---|---|---|
| **Akurasi Kelas Positif** | 100,00% | 97,23% | 100,00% |
| **Akurasi Kelas Negatif** | 0,00% | 65,00% | 10,00% |
| **Macro-F1** | 0,58 | 0,82% | 0,58 |
| **Balanced Accuracy** | 50,00% | 81,12% | 55,00% |

Berdasarkan Tabel 6, Logistic Regression menunjukkan kinerja terbaik dalam mengenali kelas negatif dengan akurasi 65%, jauh lebih tinggi dibandingkan Naïve Bayes (0%) dan Ensemble Soft Voting (10%). Hal ini tercermin pada nilai Macro-F1 dan Balanced Accuracy, di mana Logistic Regression memperoleh nilai tertinggi (0,82 dan 81,12%). Dengan demikian, meskipun Ensemble Soft Voting memiliki akurasi keseluruhan tertinggi, Logistic Regression lebih unggul dalam hal keseimbangan kinerja antar kelas, terutama dalam mendeteksi sentimen negatif yang merupakan kelas minoritas.

### 3.5 Confusion Matrix Analysis

Untuk mengetahui sejauh mana masing-masing model dapat membedakan sentimen positif dan negatif, Confusion Matrix digunakan sebagai alat analisis[6]. Dari matriks ini, diperoleh empat indikator, yaitu TP, TN, FP, dan FN, yang mencerminkan kinerja klasifikasi setiap model[1]. Ringkasan Confusion Matrix dari ketiga model tersaji pada Tabel 7, Tabel 8, dan Tabel 9.

Tabel 7. Confusion Matrix Naive Bayes

| | Prediksi Positif | Prediksi Negatif |
|---|---|---|
| **Aktual Positif** | 577 | 0 |
| **Aktual Negatif** | 20 | 0 |

Tabel 6 memperlihatkan bahwa Naïve Bayes berhasil mengklasifikasikan seluruh data positif dengan benar, ditandai dengan nilai TP = 577 dan FN = 0[6]. Di sisi lain, model ini sama sekali tidak mampu mendeteksi ulasan negatif; seluruh 20 ulasan negatif diprediksi sebagai positif (FP = 20), sementara tidak ada satu pun yang teridentifikasi dengan benar (TN = 0). Hal ini menunjukkan bahwa model sangat bias terhadap kelas mayoritas (positif) akibat ketidakseimbangan data[6].

Tabel 8. Confusion Matrix Logistic Regression

| | Prediksi Positif | Prediksi Negatif |
|---|---|---|
| **Aktual Positif** | 561 | 16 |
| **Aktual Negatif** | 7 | 13 |

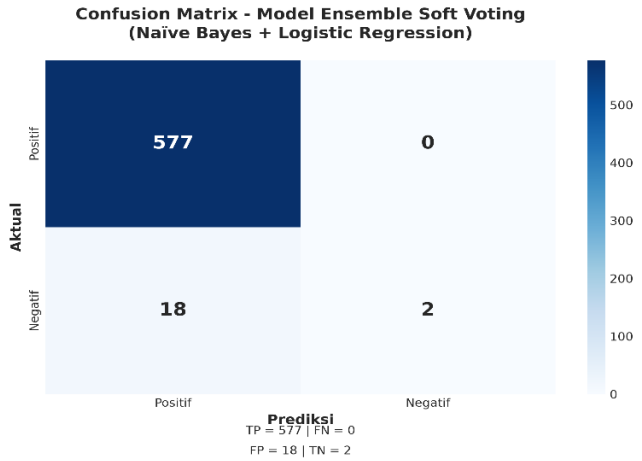
Pada Tabel 8, Logistic Regression memperlihatkan kinerja yang lebih proporsional dibanding Naïve Bayes. Model ini berhasil mengklasifikasikan 561 ulasan positif dengan benar (TP) dan 13 ulasan negatif dengan benar (TN) [7]. Namun, masih terdapat 16 ulasan positif yang terlewat (FN) dan 7 ulasan negatif yang salah diprediksi sebagai positif (FP). Akurasi kelas negatif mencapai 65% (13 dari 20 ulasan

negatif terdeteksi dengan benar), yang jauh lebih baik dibandingkan Naïve Bayes yang tidak mampu mendeteksi ulasan negatif sama sekali [6].

Tabel 9. Confusion Matrix Ensemble Soft Voting

| | **Prediksi Positif** | **Prediksi Negatif** |
|---|---|---|
| **Aktual Positif** | 577 | 0 |
| **Aktual Negatif** | 18 | 2 |

Berdasarkan Tabel 9, model *Ensemble Soft Voting* berhasil mengklasifikasikan seluruh ulasan positif dengan benar (TP = 577) dan tidak ada ulasan positif yang terlewat (FN = 0), sama seperti Naïve Bayes [9]. Model ini juga berhasil mendeteksi 2 dari 20 ulasan negatif dengan benar (TN = 2), meskipun masih ada 18 ulasan negatif yang keliru diklasifikasikan sebagai positif. (FP = 18). Dibandingkan dengan Naïve Bayes yang memiliki TN = 0, *Ensemble Soft Voting* menunjukkan peningkatan dalam kemampuan mendeteksi sentimen negatif [8].




Gambar 4. Heatmap Confusion Matrix Model Ensemble Soft Voting

Hasil Confusion Matrix menunjukkan bahwa meskipun *Ensemble Soft Voting* memiliki akurasi tertinggi, kemampuannya dalam mendeteksi sentimen negatif masih terbatas (10%). Hal ini disebabkan oleh ketidakseimbangan data yang ekstrem, di mana hanya 3,4% data yang berlabel negatif [1]. Kondisi serupa juga ditemukan dalam penelitian sebelumnya yang menunjukkan bahwa model cenderung kecenderungan model terhadap kelas dominan pada data yang tidak proporsional[6]. Meskipun demikian, *Ensemble Soft Voting* berhasil meningkatkan deteksi sentimen negatif dari 0% (Naïve Bayes) menjadi 10%, menunjukkan bahwa pendekatan ensemble memberikan perbaikan meskipun masih terbatas [9].

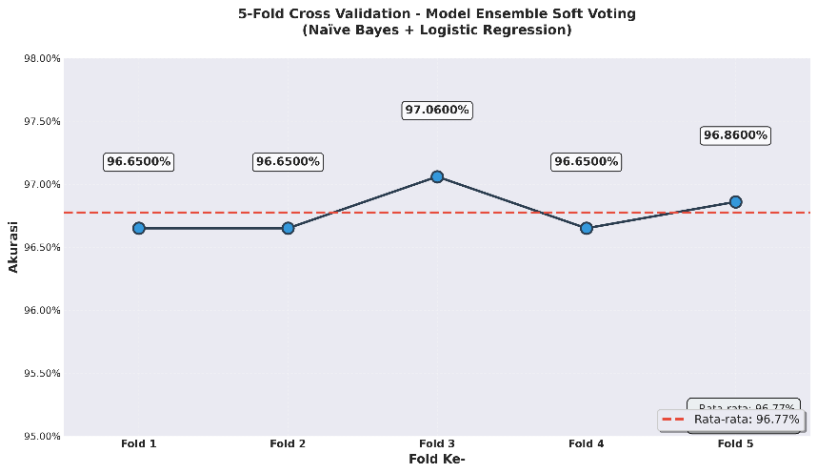
### 3.6 Hasil Cross Validation

Agar model terhindar dari overfitting dan mampu beradaptasi dengan data baru, dilakukan validasi silang 5-fold (*5-Fold Cross Validation*) terhadap model *Ensemble Soft Voting*[9]. Metode ini membagi data latih menjadi 5 bagian (fold) berukuran sama. Secara bergiliran, masing-masing fold dipakai sebagai data uji, sementara empat fold lainnya digunakan untuk pelatihan[10]. Hasil validasi silang disajikan pada Tabel 10.

Tabel 10. Hasil 5-Fold Cross Validation Model Ensemble Soft Voting

| Fold | Akurasi |
|---|---|
| Fold 1 | 96,65% |
| Fold 2 | 96,65% |
| Fold 3 | 97,06% |
| Fold 4 | 96,65% |
| Fold 5 | 96,86% |
| **Rata-rata** | **96,77%** |
| **Standar Deviasi** | **0,0017** |

Berdasarkan Tabel 10, model *Ensemble Soft Voting* menampilkan performa yang konsisten pada setiap *fold* dengan rata-rata akurasi sebesar 96,77% dan standar deviasi yang sangat kecil yaitu 0,0017. Nilai standar deviasi yang rendah hal ini menandakan bahwa model memiliki stabilitas yang sangat baik serta tidak mengalami overfitting terhadap data training tertentu [8]. Akurasi tertinggi dicapai pada Fold 3 sebesar 97,06%, sedangkan akurasi terendah pada Fold 1, Fold 2, dan Fold 4 sebesar 96,65%.




Gambar 5. Grafik 5-Fold Cross Validation Model Ensemble Soft Voting

Hasil validasi silang ini juga mendukung temuan sebelumnya bahwa model *Ensemble Soft Voting* memiliki kinerja yang stabil dan dapat diandalkan untuk klasifikasi sentimen wisatawan [9]. Konsistensi akurasi di seluruh *fold* menunjukkan bahwa model tidak sensitif terhadap perubahan komposisi data training, sehingga dapat digeneralisasikan dengan baik pada data baru [10].

## 3.7 Discussion

Berdasarkan hasil penelitian, model Ensemble Soft Voting terbaik yang dimana memperoleh akurasi tertinggi yakni 96,98%, mengungguli Naïve Bayes (96,65%) dan Logistic Regression (96,15%). Peningkatan ini membuktikan bahwa penggabungan kedua model melalui pendekatan Soft Voting efektif dalam mengkombinasikan kelebihan masing-masing algoritma [8].

Keunggulan *Ensemble Soft Voting* terlihat dari kemampuannya mempertahankan *Recall* sempurna (100%) untuk sentimen positif, sambil tetap mampu mendeteksi sebagian kecil sentimen negatif (10%) yang tidak dapat dideteksi oleh Naïve Bayes

sama sekali [6]. Namun, kemampuan deteksi sentimen negatif yang masih rendah mengindikasikan bahwa ketidakseimbangan data yang ekstrem (96,6% positif vs 3,4% negatif) tetap menjadi tantangan utama dalam analisis sentimen wisata Aceh [1].

Kelemahan utama penelitian ini terletak pada ketimpangan distribusi data yang cukup besar. Untuk riset ke depan, disarankan menggunakan pendekatan resampling, misalnya SMOTE, agar data menjadi lebih seimbang atau class weighting untuk menangani ketidakseimbangan data [6]. Di samping itu, penggunaan Soft Voting yang mempertimbangkan probabilitas terbukti lebih baik daripada Hard Voting, sehingga direkomendasikan untuk penelitian serupa di masa mendatang [14]. Selain itu, pemanfaatan algoritma deep learning seperti LSTM atau BERT dapat dijadikan alternatif guna meningkatkan performa analisis sentimen pada teks berbahasa Indonesia [11].

Sekian, penelitian ini memberikan sumbangan terhadap pengembangan model analisis sentimen untuk pariwisata Aceh melalui pendekatan ensemble learning. Model *Ensemble Soft Voting* yang dihasilkan dapat dijadikan instrumen evaluasi layanan pariwisata berbasis data digital bagi pengelola destinasi dan pemerintah daerah [2].

## 4. Kesimpulan

Merujuk pada temuan penelitian ini, model Ensemble Soft Voting berhasil mencapai akurasi testing terbaik sebesar 96,98% dalam mengklasifikasikan sentimen wisatawan terhadap destinasi wisata unggulan di Aceh, mengungguli Naïve Bayes (96,65%) dan Logistic Regression (96,15%). Pada evaluasi data training, Logistic Regression mencatat akurasi tertinggi sebesar 98,24%, diikuti Ensemble Soft Voting (97,15%) dan Naïve Bayes (96,73%). Model Ensemble Soft Voting menunjukkan stabilitas yang sangat baik dengan rata-rata validasi silang 5-fold sebesar 96,77% dan standar deviasi 0,0017. Temuan ini menegaskan bahwa adopsi pendekatan ensemble learning untuk memadukan kedua algoritma efektif dalam meningkatkan kinerja klasifikasi, dengan Ensemble Soft Voting sebagai model dengan akurasi keseluruhan terbaik, meskipun Logistic Regression lebih unggul dalam mendeteksi sentimen negatif.

Meskipun Ensemble Soft Voting memiliki akurasi keseluruhan yang tinggi (96,98%), model ini masih memiliki kelemahan mendasar karena hanya mampu mengenali 10% dari sentimen negatif yang ada (2 dari 20 ulasan). Hal ini disebabkan oleh ketidakseimbangan data yang ekstrem (96,6% positif vs 3,4% negatif). Oleh karena itu, pembaca dan pengelola objek wisata perlu memahami bahwa model ini lebih cocok untuk analisis sentimen positif secara umum, bukan untuk deteksi keluhan wisatawan secara spesifik.

Selain itu, dominasi sentimen positif (96,6%) pada data ulasan menunjukkan tingkat kepuasan wisatawan yang tinggi terhadap destinasi wisata Aceh [2].Secara keseluruhan, model *Ensemble Soft Voting* yang dihasilkan dapat dijadikan instrumen evaluasi layanan pariwisata berbasis data digital bagi pengelola objek wisata maupun pemerintah daerah guna memperbaiki mutu layanan berdasarkan umpan balik wisatawan yang terkumpul di platform digital.

## Daftar Pustaka

[1] Susilawati and S. Wahyuni, "Analisis Sentimen Publik Terhadap Pariwisata Aceh di Media Sosial X Menggunakan Algoritma Naive Bayes Classifier," *Bull. Inf. Technol.*, vol. 5, no. 4, pp. 269–278, 2024, doi: 10.47065/bit.v5i2.1700.

[2] S. M. Sa'adah, K. Umam, M. R. Handayani, and M. I. Mustofa, "Mapping the Polarity of Tourist Opinions on Indonesian Destinations through Google Maps Reviews Using Supervised Learning Methods," *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2845–2853, 2025, doi: 10.30871/jaic.v9i5.9836.

[3] W. H. DeLone and E. R. McLean, "The DeLone and McLean model of information systems success: A ten-year update," *J. Manag. Inf. Syst.*, vol. 19, no. 4, pp. 9–30, 2003, doi: 10.1080/07421222.2003.11045748.

[4] Z. Rifa'i and B. P. Mukti, "Weakly Supervised Sentiment Analysis of Indonesian Rural Tourism Reviews: A TF-IDF Baseline for Melung Tourism Village," *Edu Komputika J.*, vol. 12, no. 1, pp. 48–60, 2025, doi: 10.15294/edukom.v12i1.31893.

[5] M. L. Methods, "Sentiment Analysis of Visitor Reviews on Baturaden Tourist Attraction Using Machine Learning Methods," *Edu Komputika J.*, vol. 11, no. 1, pp. 57–64, 2024, doi: 10.15294/edukom.v11i1.10561.

[6] D. Y. Saraswati, M. R. Handayani, K. Umam, and M. I. Mustofa, "Sentiment Classification of MyPertamina Reviews Using Naïve Bayes and Logistic Regression," *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1319–1325, 2025, doi: 10.30871/jaic.v9i4.9723.

[7] V. B. Lestari and C. A. Hutagalung, "Evaluation of TF-IDF ExtractionTechniques in Sentiment Analysis ofIndonesian-Language MarketplacesUsing SVM, Logistic Regression, andNaive Bayes," *J. Comput. Sci. Appl.*, vol. 8, no. 1, pp. 22–2025, 2025, [Online]. Available: https://doi.org/10.21009/j-

[8] V. T. Malya, S. Erlinda, H. Hamdani, and R. Yanti, "Analisis Sentimen Terhadap Masjid Raya an-Nur Provinsi Riau Menggunakan Teknik Stacking Machine Learning," *J. Tek. Inf. dan Komput.*, vol. 8, no. 1, p. 37, 2025, doi: 10.37600/tekinkom.v8i1.2305.

[9] F. P. Saputra and O. Suria, "Combining SVM and Naive Bayes Models using a Soft Voting Approach for Sentiment Analysis of Tong Tji Tea House," *Sistemasi*, vol. 14, no. 5, p. 2479, 2025, doi: 10.32520/stmsi.v14i5.5481.

[10] R. M. Sasmita, A. Meiriza, and H. Novianti, "An Ensemble Learning Approach for Sentiment Analysis of Maxim Application Reviews Using SVM, KNN, and Random Forest," *J. Appl. Informatics Comput.*, vol. 9, no. 6, pp. 3697–3705, 2025, doi: 10.30871/jaic.v9i6.11447.

[11] Uswatun Khasanah, "Analisis sentimen terhadap tempat wisata menggunakan metode naïve bayes classifier dan support vector machine," pp. 1–112, 2023.

[12] H. T. Nugroho, "Analisis Sentimen Tempat Wisata di Kabupaten Bandung Jawa Barat dengan Visualisasi Website," vol. 12, no. 6, pp. 8476–8482, 2025.

[13] F. Widyastuti and E. Mailoa, "Analisis Sentimen Ulasan Konsumen Menggunakan Algoritma Tf-Id Untuk ( Studi Kasus : Gunthem Premium Coffee )," *JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 16, no. 2, pp. 8–16, 2025.

[14] S. Jindal, M. Sachdeva, and A. K. S. Kushwaha, "Performance evaluation of machine learning based voting classifier system for human activity recognition Dept . of Computer Science and Engineering I . K . Gujral Punjab Technical University , Jalandhar , India Guru Ghasidas Vishwavidyalaya , Bilaspur , I," pp. 1–13, 2022.