

Klasifikasi Sentimen Ulasan Turis Terhadap Objek Wisata di Bali Menggunakan Support Vector Machine

Sentiment Classification of Tourist Reviews on Tourist Attractions in Bali Using Support Vector Machine

Raudhah¹, Fajri Profesio Putra²

^{1,2}Jurusan Teknik Informatika, Politeknik Negeri Bengkalis

E-mail: ¹ltlrdh@gmail.com, ²fajri@polbeng.ac.id

Received 26 May 2025; Revised 23 June 2025; Accepted 7 July 2025

Abstrak - Pariwisata di Bali memainkan peran penting dalam perekonomian daerah, dan ulasan wisatawan dapat memberikan wawasan berharga terkait kualitas pengalaman yang dirasakan. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan wisatawan berbahasa Inggris terhadap enam objek wisata di Bali, yaitu Pantai Kuta, Pantai Nusa Dua, Pantai Sanur, Tegalalang Rice Terrace, Pura Luhur Uluwatu, dan Pura Penataran Agung Lempuyang, dengan total data sebanyak 4.872 ulasan, masing-masing 812 ulasan per objek. Data diperoleh dari situs *TripAdvisor* dan *Google Maps*, dan dianalisis melalui tahapan *web scraping*, *text preprocessing*, transformasi menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), serta pemodelan dengan algoritma *Support Vector Machine* (SVM). Hasil evaluasi menunjukkan bahwa algoritma *Support Vector Machine* (SVM) mampu mengklasifikasikan sentimen dengan akurasi sebesar 93,95% saat menggunakan data yang seimbang dengan *precision*, *recall* dan *F1-score* untuk kedua kelas (*Negative* dan *Positive*) masing-masing mencapai 94%, dan akurasi sebesar 90,83% pada data yang tidak seimbang dengan kelas *Positive* memiliki *recall* dan *F1-score* yang lebih tinggi (94% dan 92%), sementara kelas *Negative* memiliki *precision* dan *F1-score* yang sedikit lebih rendah (91% dan 88%). Temuan ini menunjukkan bahwa model bekerja lebih optimal pada data dengan distribusi sentimen yang seimbang. Secara keseluruhan, algoritma *Support Vector Machine* (SVM) terbukti efektif dan akurat dalam menganalisis sentimen ulasan wisatawan terhadap objek wisata di Bali.

Kata kunci: Sentimen, Ulasan Wisatawan, *Support Vector Machine* (SVM), Objek Wisata, Bali

Abstract - Tourism in Bali plays a significant role in the region's economy, and traveler reviews can provide valuable insights into the quality of the experiences. This study aims to classify the sentiment of English-language reviews from tourists on six tourist attractions in Bali: Kuta Beach, Nusa Dua Beach, Sanur Beach, Tegalalang Rice Terrace, Pura Luhur Uluwatu, and Pura Penataran Agung Lempuyang, with a total of 4872 reviews, 812 reviews per attraction. The data was obtained from *TripAdvisor* and *Google Maps*, and analyzed through web scraping, text preprocessing, transformation using *Term Frequency-Inverse Document Frequency* (TF-IDF), and modeling using the *Support Vector Machine* (SVM) algorithm. The evaluation results show that the *Support Vector Machine* (SVM) algorithm is able to classify sentiment with an accuracy of 93.95% when using balanced data, with *precision*, *recall*, and *F1-score* for both classes (*Negative* and *Positive*) each reaching 94%. The accuracy drops to 90.83% when using imbalanced data, with the *Positive* class having higher *recall* and *F1-score* (94% and 92%), while the *Negative* class has slightly lower *precision* and *F1-score* (91% and 88%). These findings indicate that the model performs better on data with a balanced sentiment distribution. Overall, the *Support Vector Machine* (SVM) algorithm proves to be effective and accurate in analyzing the sentiment of tourist reviews toward tourist attractions in Bali.

Keywords: Sentiment, Tourist Reviews, *Support Vector Machine* (SVM) Tourist Attractions, Bali.

1. PENDAHULUAN

Pariwisata adalah sektor yang sangat penting bagi perekonomian Indonesia, dan Bali adalah salah satu tujuan wisata terkenal di dunia. Keberadaan objek wisata juga mendukung kesejahteraan masyarakat sekitar dan menjadi sumber pendapatan bagi pemerintah daerah [1]. Setiap tahun, jutaan wisatawan tertarik datang ke Bali untuk mengunjungi berbagai objek wisata bahari seperti pantai Kuta, Nusa Dua, dan Sanur, objek wisata alam salah satunya Tegalalang Rice Terrace dan objek wisata budaya dan spiritual seperti Pura Luhur Uluwatu dan Pura Penataran. Ini menunjukkan pentingnya Bali dalam perkembangan pariwisata di Indonesia, terutama dalam menarik wisatawan mancanegara [2]. Ulasan wisatawan di *platform* seperti *TripAdvisor* dan *Google Maps* dianggap penting karena dapat mempengaruhi keputusan wisatawan lain dalam merencanakan kunjungan mereka [3].

TripAdvisor dan *Google Maps* memungkinkan wisatawan memberikan ulasan tentang objek wisata yang mereka kunjungi. Ulasan ini bisa menjadi informasi yang sangat berguna bagi wisatawan lain yang ingin merencanakan perjalanan [4]. Namun, dengan banyaknya ulasan yang tersedia menyulitkan analisis manual, sehingga diperlukan metode yang efektif untuk menganalisis ulasan dan mengetahui sentimen wisatawan. Untuk memproses ulasan dari pengguna *TripAdvisor* dan *Google Maps*, klasifikasi sentimen digunakan sebagai pendekatan komputasi untuk mengidentifikasi opini, sentimen, dan emosi yang diekspresikan dalam teks. Melalui analisis ini, informasi dalam bentuk teks dapat diklasifikasikan ke dalam kategori positif atau negatif sehingga kecenderungan opini dapat diketahui. Dalam penelitian ini, analisis sentimen akan dilakukan menggunakan metode data mining dengan algoritma klasifikasi *Support Vector Machine* (SVM) [5].

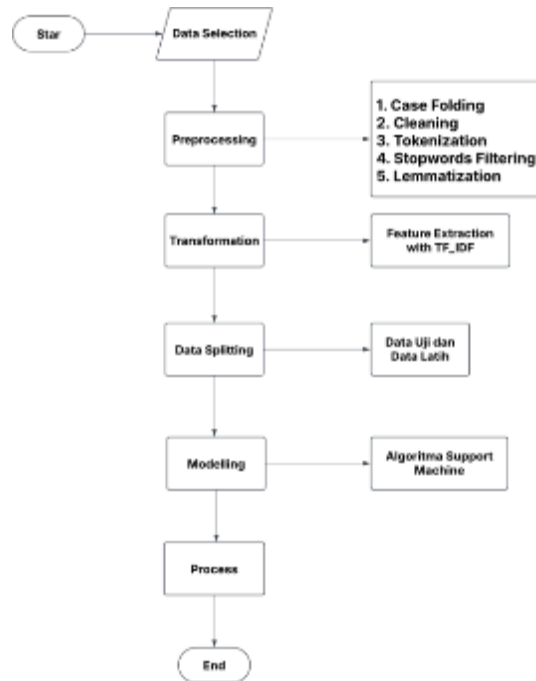
Support Vector Machine (SVM) digunakan sebagai metode *supervised learning* untuk menyelesaikan masalah klasifikasi dan regresi, baik secara *linear* maupun *non-linear*. Metode ini dikenal memiliki dasar matematis yang matang dalam pemodelan klasifikasi. Dalam proses klasifikasi data, *Support Vector Machine* (SVM) bekerja dengan menentukan *hyperplane* terbaik yang memisahkan dua kelas melalui penerapan transformasi matematis menggunakan fungsi kernel, sehingga pemisahan data dapat dilakukan secara optimal [6].

Beberapa penelitian terdahulu yakni penggunaan metode *decision tree* untuk menganalisis sentimen berbasis aspek ulasan pelanggan hotel di Bali yang bertujuan untuk membangun *model aspect-based sentiment analysis* terhadap ulasan pelanggan hotel di Bali menggunakan metode *Decision Tree*. Model menghasilkan performa terbaik pada aspek umum dengan akurasi 82,5%, serta performa yang cukup baik pada aspek layanan dan kebersihan, dengan *F1-Score* masing-masing 76,2% dan 82,4% [7]. Penelitian menggunakan algoritma *Support Vector Machine* (SVM) untuk analisis sentimen wisata alun-alun kota batu guna menganalisis sentimen ulasan pengunjung terhadap wisata Alun-Alun Kota Batu guna membantu pengelola dalam evaluasi layanan dan infrastruktur. Hasilnya, pengujian terbaik diperoleh menggunakan kernel RBF dengan parameter $C = 50$ dan *8-fold cross validation*, menghasilkan akurasi 89,58%, *precision* 90,73%, *recall* 89,48%, *f-measure* 89,45%, dan *specificity* 91,27% [8]. Penggunaan algoritma *Support Vector Machine* (SVM) untuk menganalisis sentimen tempat wisata berdasarkan ulasan pada *Google Maps* yang bertujuan untuk menganalisis sentimen ulasan pengguna *Google Maps* terhadap destinasi wisata di Yogyakarta yang menghasilkan tingkat akurasi rata-rata sebesar 83,8% dalam mengklasifikasikan sentimen positif, negatif, dan netral [9].

Berdasarkan penjelasan latar belakang sebelumnya, penelitian ini bertujuan untuk mengkalsifikasikan sentimen ulasan bahasa Inggris dari turis terhadap objek wisata di Bali yang ada di *TripAdvisor* menggunakan algoritma *Support Vector Machine* (SVM). Dengan menggunakan *Support Vector Machine* (SVM), diharapkan bisa mendapatkan pemahaman yang lebih jelas mengenai pengalaman wisatawan. Hasil dari penelitian ini dapat digunakan untuk memberikan gambaran mengenai sentimen wisatawan terhadap objek wisata di Bali. Penelitian ini juga dapat menunjukkan apakah ulasan wisatawan cenderung positif atau negatif terhadap

berbagai objek wisata. Dengan demikian, penelitian ini dapat memberikan wawasan yang lebih jelas tentang bagaimana wisatawan merasakan pengalaman mereka di Bali.

2. METODE PENELITIAN



Gambar 1 Prosedur Penelitian

Penelitian ini menggunakan metode *Knowledge Discovery in Databases* (KDD) untuk menemukan pola dan informasi dari data terpilih [10]. Proses dimulai dengan pemilihan data yang relevan, kemudian dilakukan pra-pemrosesan berupa case folding, cleaning, tokenization, stopwords filtering, dan lemmatization. Data selanjutnya ditransformasi menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), dibagi menjadi data latih dan data uji, lalu diproses dengan algoritma *Support Vector Machine* (SVM) untuk membentuk model. Setelah pengujian dilakukan, hasil dievaluasi berdasarkan akurasi dan disajikan dalam bentuk laporan atau visualisasi guna mendukung pengambilan keputusan.

Setelah data diproses, tahap transformasi dilakukan dengan menggunakan teknik seperti *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengubah data menjadi format yang bisa dibaca oleh model. Data yang sudah diubah kemudian dibagi menjadi dua bagian, yaitu *training set* untuk melatih model dan *testing set* untuk menguji model. Pada tahap *modelling*, algoritma *Support Vector Machine* (SVM) digunakan untuk membangun model yang dapat mengenali pola dari data. Setelah model diuji, hasilnya dievaluasi berdasarkan metrik seperti akurasi. Terakhir, hasil analisis ini disajikan dalam bentuk laporan atau visualisasi untuk membantu dalam pengambilan keputusan.

2.1 Langkah-Langkah Penelitian

1. Data Selection

Pada tahap pemilihan data (*data selection*) dalam penelitian ini, data yang dipilih adalah ulasan wisata dalam bahasa Inggris terkait objek wisata di Bali yang diambil dari *website TripAdvisor*. Pemilihan bahasa Inggris dilakukan untuk memastikan bahwa data yang digunakan sesuai dengan analisis sentimen yang akan dilakukan. Adapun objek wisata di Bali yang ulasannya diklasifikasi mencakup tiga wisata bahari, yaitu Pantai Kuta, Pantai Nusa Dua,

dan Pantai Sanur, serta satu wisata alam, yaitu Tegallalang Rice Terrace dan dua objek wisata budaya dan spiritual yaitu Pura Luhur Uluwatu dan Pura Penataran.

2. Preprocessing

Preprocessing merupakan langkah awal dalam pengolahan teks yang bertujuan untuk mempersiapkan teks agar dapat diolah lebih lanjut. Karena teks dalam bentuk aslinya tidak bisa langsung diproses oleh algoritma pencarian, maka diperlukan proses *preprocessing* untuk mengubahnya menjadi representasi numerik yang dapat dipahami oleh sistem [11].

Pada *preprocessing*, akan dilakukan beberapa tahapan yaitu, *Case Folding* mengubah huruf menjadi kecil, *Cleaning* menghapus elemen tidak perlu, *Tokenization* memecah teks menjadi token, *Stopwords Filtering* menghilangkan kata-kata umum, dan *Lemmatization* mengubah kata ke bentuk dasarnya.

3. Transformation

Transformation dalam analisis teks mencakup pembobotan *Term Frequency Inverse Document Frequency* (TF-IDF), yaitu dilakukan perhitungan terlebih dahulu nilai *Term Frequency* (TF) per kata dengan bobot masing-masing kata [12]. Hasil dari proses ini memungkinkan data yang telah diproses diubah menjadi format yang dapat dianalisis lebih lanjut, seperti menggunakan *Feature Extraction* dengan TF-IDF.

Document Frequency (TF-IDF) dijelaskan melalui persamaan berikut:

- 1) Nilai *Term Frequency* (TF) di peroleh dari nilai frekuensi kemunculan fitur t pada dokumen d .

$$TF_t = (t, d) \quad (1)$$

- 2) Nilai *Inverse Document Frequency* di peroleh dari logaritma banyaknya dokumen n dibagi dokumen df yang mengandung fitur t .

$$IDF_t = (\log \frac{n}{df(t)}) \quad (2)$$

- 3) Nilai *Term Frequency Inverse Document Frequency* (W_t) didapatkan dengan mengalihkan nilai TF dengan IDF.

$$W_t = TF_t \times IDF_t \quad (3)$$

4. Data Splitting

Data splitting adalah proses membagi data menjadi data *training* untuk melatih model dan data *testing* untuk menguji akurasi model. Data training adalah data yang sudah dilengkapi dengan label atau kelas, dan digunakan oleh mesin untuk mempelajari karakteristik data guna membentuk suatu pola atau model. Sedangkan data testing merupakan data berlabel yang digunakan untuk menguji sejauh mana model yang telah dibentuk mampu mengklasifikasikan data dengan tepat [13]. Pada penelitian ini, dataset di bagi menjadi data training (80%) dan data testing (20%)

5. Modelling

Pemodelan menggunakan algoritma *Support Vector Machine* (SVM) dilakukan untuk memprediksi sentimen dari ulasan wisatawan. *Support Vector Machine* (SVM) bekerja dengan memaksimalkan jarak antar kelas agar menghasilkan pemisahan yang optimal. Pada pendekatan ini, *Support Vector Machine* (SVM) mampu memberikan generalisasi yang tinggi terhadap data baru yang akan datang [14]. Data dibagi menjadi *training set* dan *testing set*, di mana model dilatih pada data latih dan diuji pada data uji. Fungsi keputusan *Support Vector Machine* (SVM) digunakan dalam proses klasifikasi.

$$f(x) = w \cdot x + b \quad (4)$$

Keterangan:

- x : adalah vektor fitur dokumen baru.
 w : vektor bobot yang menentukan arah dan posisi garis *hyperlane*.
 b : bias yang menentukan jarak garis *hyperlane* dari titik alas.

6. Evaluation

Mengevaluasi kinerja dengan di tampilkan dengan nilai *Accuracy*, dan bentuk table *Confusion Matrix* yaitu merupakan tabel yang menyajikan perbandingan antara hasil klasifikasi yang dihasilkan oleh sistem dengan hasil klasifikasi yang seharusnya atau yang dijadikan acuan [15]. Evaluasi ini membantu menentukan efektivitas model dalam mengkategorikan sentimen ulasan wisatawan.

2.2 Gambaran Hasil Penelitian

Berikut ini merupakan gambaran hasil penelitian analisis sentimen ulasan turis terhadap objek wisata di Bali secara manual:

1. Data Selection

Tabel 1 Pengumpulan Data

No	Username	Review
1.	Erik D	Great hotel with service minded staff. Pool area with available sunbeds. Often live music and barbecue in the afternoon.
2.	Jody262016	Beautiful area to sit and watch the waves, can grab a drink or goods off the many vendors at the beach. Currently the Ukrainian surf championships is on.
3.	Reshi Anahata	Good for surfing, the beach is quite good for chilling and meet with new people around the globe. Warungs on the pedestrian and beach bed quite affordable.
4.	Anna V	Not a good place to go. It is not a beach, but place entirely for surfing. No normal place for sunbathing. Full of small low quality shops and cafe and annoying sellers of different stuff.
5.	Abovethelclouds	So difficult to relax here. we spent one day here and would never return. Constant harassment to buy junk

Tabel 1, menunjukkan data ulasan dari situs *website TripAdvisor*, yang mencakup beberapa kolom, antara lain kolom *username* dan *review*.

2. Preprocessing

Tabel 2 Preprocessing

No	Review	Review Case Folding	Review Cleaning	Review Tokenization	Review Stopwords Filtering	Review Lemmatization
1.	Great hotel with service minded staff. Pool area with available sunbeds. Often live music and barbecue in the afternoon.	great hotel with service minded staff. pool area with available sunbeds. often live music and barbecue in the afternoon.	great hotel with service minded staff pool area with available sunbeds often live music and barbecue in the afternoon	[great, hotel, with, service, minded, staff, pool, area, with, available, sunbeds, often, live, music, and, barbecue, in, the, afternoon]	[great, hotel, service, minded, staff, pool, area, available, sunbeds, often, live, music, barbecue, afternoon]	[great, hotel, service, mind, staff, pool, area, available, sunbed, often, live, music, barbecue, afternoon]
2.	So difficult to relax here. We spent one day here and would never return. Constant harassment to buy junk	so difficult to relax here. we spent one day here and would never return. constant harassment to buy junk	so difficult to relax here we spent one day here and would never return constant harassment to buy junk	[so, difficult, to, relax, here, we, spent, one, day, here, and, would, never, return, constant, harassment, to, buy, junk]	[difficult, relax, spent, one, day, would, never, return, constant, harassment, buy, junk]	[difficult, relax, spend, one, day, would, never, return, constant, harassment, buy, junk]

Tabel 2, menunjukkan proses *preprocessing* data ulasan sebelum dianalisis. Tahapan dimulai dari mengubah huruf menjadi kecil (*case folding*), menghapus karakter tak penting (*cleaning*), memecah kalimat menjadi kata-kata (*tokenization*), menghapus kata umum (*stopwords filtering*), dan mengembalikan kata ke bentuk dasarnya (*lemmatization*). Proses ini bertujuan untuk menyederhanakan dan menyiapkan data agar lebih akurat saat dianalisis menggunakan algoritma.

3. Transformation (pembobotan *Term Frequency Inverse Document Frequency*)

Tahapan menghitung *Term Frequency Inverse Document Frequency* (TF IDF) :

- 1) Ada dua ulasan yang sudah melalui tahap *preprocessing* secara manual dan siap untuk dilakukan perhitungan TF-IDF. Dua ulasan tersebut yaitu, satu ulasan sentimen positif satu dan satu ulasan sentimen negatif.

Ulasannya :

- a. Sentiment Positif : “[great, hotel, service, mind, staff, pool, area, available, sunbed, often, live, music, barbecue, afternoon]” → 14 kata
 - b. Sentiment Negatif : “[difficult, relax, spend, one, day, would, never, return, constant, harassment, buy, junk]” → 12 kata
- 2) Menghitung *Term Frequency* (Tf)

$$TF = \frac{\text{jumlah kemunculan kata dalam dokumen}}{\text{total kata dalam dokumen}} \quad (5)$$

- 3) Menghitung *Inverse Document Frequency* (IDF)

$$IDF = \log \frac{N}{df} \quad (6)$$

di mana:

$N = 2$ (jumlah total ulasan)
 df adalah jumlah ulasan yang mengandung kata tersebut.

- 4) Menghitung *Term Frequency Inverse Document Frequency* (TF-IDF)

$$TF - IDF = TF \times IDF \quad (7)$$

Vector TF-IDF D1 dan D2 :

- a. Ulasan Positif [0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.22, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
 - b. Ulasan Negatif [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025]
4. Model dengan *Support Vector Machine* (SVM)

Adapun tahapannya, yaitu:

- 1) Dataset TF-IDF (Vektor)

Dari hasil perhitungan TF-IDF sebelumnya:

- a. Dokumen 1 (positif, label +1)

$X_1 = [0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.22, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$

- b. Dokumen 2 (negatif, label -1)

$X_2 = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025]$

Sekarang vector labelnya ada dua:

$(X_1, y_1 = +1), (X_2, y_2 = -1)$

- 2) Menentukan Bobot Awal (w) dan Bias (b)

- a. Hitung *vector* bobot w dengan menggunakan rumus:

$$w = \sum y_i \cdot X_i \quad (8)$$

Perhitungan :

$$w = (y_1 \times X_1) + (y_2 \times X_2) \quad (9)$$

$$w = (+1 \times [0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.22, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0] + (-1 \times [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025]))$$

Jadi, vector bobot $\mathbf{w}$ adalah :

b. Hitung bias (b) menggunakan rumus:

c. Hitung rata-rata bias kedua dokumen:

Bias Dokumen 2:

Rata-rata Bias (b) :

d. Lalu, gunakan rumus fungsi keputusan:

(14)

Hitung nilai keputusan (x) untuk masing-masing ulasan

a. Untuk Ulasan 1 (positif, $y = +1$):

- $$\begin{aligned} f(x_1) &= [0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.0022, 0.022, 0.195, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025] \times [0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.22, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0] + (-0.0235) \\ &= 0.0487564 + (-0.0235) \\ &= 0.025 \end{aligned}$$

- $$\begin{aligned} f(x_2) &= [0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.022, 0.0022, 0.022, 0.195, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025, -0.025] \times [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025] + (-0.0235) \\ &= -0.003875 + (-0.0235) \\ &= -0.027 \end{aligned}$$

Penyelesaian :

a. Dokument

- e. Model dengan *Support Vector Machine* (SVM)

$$= -0.038$$

Pengumpulan data

Sanur, Tegallalang Rice Terrace, Pura Luhur Uluwatu, dan Pura Penataran Agung Lempuyang. Setiap objek wisata memiliki 812 ulasan, dengan total 4.872 data. Bahasa ulasan difilter menggunakan library *langdetect*.

3.2 Preprocessing

Adapun langkah-langkah *preprocessing* yaitu, *Case Folding* (Mengubah teks menjadi huruf kecil), *Cleaning* (Menghapus karakter non-alfabet), *Tokenization* (Memecah teks menjadi kata-kata), *Stopwords Filtering* (Menghapus kata-kata umum yang tidak berkontribusi dalam analisis), *Lemmatization* (Mengembalikan kata ke bentuk dasarnya).

Original	Tokenized	Stopwords_Removed	Lemmatized
Great hotel with service minded staff. Pool area with available sunbeds. Often live music and barbecue in the afternoon.	great hotel with service minded staff pool area with available sunbeds often live music and barbecue in the afternoon	great hotel service minded staff pool area available sunbeds often live music barbecue afternoon	great hotel service minded staff pool area available sunbeds often live music barbecue afternoon
It was one of the most unpleasant experiences ever lived. I couldn't enjoy the view or the quiet. Every 10 meters a resident came to try to sell you some junk that no one would need. They are lazy, fussy about hygiene, naughty and on top of that insensitive. All the people I met in Bali were of a high degree of civilization and good sense, less this locality. Don't regret coming this far and try to avoid it.	it was one of the most unpleasant experiences ever lived i couldnt enjoy the view or the quiet every meters a resident came to try to sell you some junk that no one would need they are lazy fussy about hygiene naughty and on top of that insensitive all the people i met in bali were of a high degree of civilization and good sense less this locality dont regret coming this far and try to avoid it	one unpleasant experiences ever lived i couldnt enjoy view quiet every meters resident came try sell junk no one would need they lazy fussy hygiene naughty top insensitive all people met bali high degree civilization good sense less locality dont regret coming far try avoid	one unpleasant experience ever lived i couldnt enjoy view quiet every meter resident came try sell junk no one would need they lazy fussy hygiene naughty top insensitive all people met bali high degree civilization good sense less locality dont regret coming far try avoid
I'm sure the rest of Bali is beautiful but as for Kuta (Beach) it was awful, especially the beach, truly awful, so much rubbish etc!!!	im sure the rest of bali is beautiful but as for kuta beach it was awful especially the beach truly awfulso much rubbish etc	im sure rest bali beautiful kuta beach awful especially beach truly awfulso much rubbish etc	im sure rest bali beautiful kuta beach awful especially beach truly awfulso much rubbish etc
There is no beach. Its loaded with driftwoods and trash. I find a little bit of clean space and walk down and the waves bring huge trash and deposit at your feet. There is a walking path in between shacks. The shacks let out seats for payments. Surfing classes are provided. But huge logs are very dangerous.	there is no beach its loaded with driftwoods and trash i find a little bit of clean space and walk down and the waves bring huge trash and deposit at your feet there is a walking path in between shacks the shacks let out seats for payments surfing classes are provided but huge logs are very dangerous	no beach loaded driftwoods trash u find little bit clean space walk waves bring huge trash deposit feet walking path shacks shacks let seats payments surfing classes provided huge logs dangerous	no beach loaded driftwood trash u find little bit clean space walk wave bring huge trash deposit foot walking path shack shack let seat payment surfing class provided huge log dangerous
Dirtiest beach I have ever seen. Piles of trash and wood for as far as your eyes can see. Unsafe to swim as the wood was large. Nothing like the pictures you see of Bali.	dirtiest beach i have ever seen piles of trash and wood for as far as your eyes can see unsafe to swim as the wood was largenothinglike the pictures you see of bali	dirtiest beach ever seen piles trash wood far eyes see unsafe swim wood largenothinglike the pictures see bali	dirtiest beach ever seen pile trash wood far eye see unsafe swim wood largenothinglike the picture see bali

Gambar 2 Hasil *Preprocessing*

Saat melakukan *web scrapping* pada setiap objek wisata, data di ambil sebanyak 812 data. Pada tahap *preprocessing* dilakukan dengan menggabungkan semua data sehingga berjumlah 4.872 data sebelum *preprocessing*, dan setelah *preprocessing* dilakukan jumlah data tetap 4.872.

3.3 Transformation Term Frequency Inverse Document Frequency TF-IDF

Ulasan ubah ke vektor angka menggunakan *Term Frequency Inverse Document Frequency* TF-IDF.

```
from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer(stop_words='english', min_df=2, max_df=0.95)
X_tfidf = vectorizer.fit_transform(df['Processed Review'])
```

3.4 Data Splitting

Membagi dataset menjadi data *Testing Set* dan *Training Set* dengan rasio 80:20.

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder

# Encoding label sentimen
le = LabelEncoder()
df['Sentiment'] = le.fit_transform(df['Sentiment'])

# Pisahkan fitur (X) dan target (y)
X = X_tfidf
y = df['Sentiment']

# Split data (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                                    random_state=42)

print(f'Jumlah Data Latih: {X_train.shape[0]}, Jumlah Data Uji: {X_test.shape[0]}')
```

3.5 Modelling

Pemodelan dilakukan menggunakan algoritma klasifikasi *Support Vector Machine* (SVM) untuk membangun model prediksi sentimen dari teks ulasan wisatawan.

```
from sklearn.svm import SVC

# Training model SVM
model = SVC(kernel='linear', C=1.0, random_state=42)
model.fit(X_train_balanced, y_train_balanced)
print("Model SVM berhasil dilatih!")
```

3.6 Evaluation

Model dievaluasi menggunakan beberapa metrik, yaitu *Accuracy Score*, *Confusion Matrix*, *Classification Report*.

```
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report

# Prediksi data uji
y_pred = model.predict(X_test)

# Evaluasi model
accuracy = accuracy_score(y_test, y_pred)
cm = confusion_matrix(y_test, y_pred)
report = classification_report(y_test, y_pred, target_names=le.classes_)
```

Hasilnya :

```
=== Evaluasi Model untuk 80:20 ===
Akurasi: 0.9394871794871795
Confusion Matrix:
[[458  29]
 [ 30 458]]
Classification Report:
              precision    recall  f1-score   support

   Negative         0.94         0.94         0.94         487
   Positive         0.94         0.94         0.94         488

 accuracy                   0.94         0.94         0.94         975
 macro avg              0.94         0.94         0.94         975
 weighted avg           0.94         0.94         0.94         975
```

Gambar 3 Evaluation

3.7 Implementasi Pada Website

Website ini dibangun untuk melakukan klasifikasi sentimen terhadap ulasan wisata menggunakan algoritma *Support Vector Machine* (SVM). Sistem ini memungkinkan pengguna untuk memasukkan ulasan dalam bahasa Inggris, dan sistem akan memprediksi apakah ulasan tersebut memiliki sentimen positif atau negatif. Adapun tampilan website dapat dilihat pada gambar berikut:

- 1) Tampilan website sebelum memasukkan ulasan



Gambar 4 Tampilan Website sebelum memasukkan ulasan

- 2) Tampilan website ulasan Positif



Gambar 5 Tampilan Website ulasan Positif

3) Tampilan website ulasan Negatif



Gambar 6 Tampilan Website ulasan Negatif

Website ini dirancang dengan antarmuka yang sederhana dan mudah digunakan. Pengguna cukup memasukkan ulasan dalam bahasa Inggris, kemudian klik **Proses**, dan sistem akan menampilkan hasil analisis apakah ulasan tersebut bersifat **positif** atau **negatif**.

3.8 Pengujian Model

1. Pengujian *Split Ratio Sensitivity Analysis*

Tabel 3 Pengujian *Split Ratio Sensitivity Analysis*

	50:50	60:40	70:30	80:20
Akurasi	92,40%	93,17%	93,50%	93,95%

Tabel 3, menunjukkan akurasi model cukup stabil meski rasio data diubah. Akurasi tertinggi (93,95%) terjadi saat data dibagi 80:20. Artinya, semakin banyak data latih, akurasi model cenderung meningkat.

2. Pengujian *Class Imbalance Evaluation*

Tabel 4 Pengujian *Class Imbalance Evaluation*

	Dataset Tidak Seimbang	Dataset Seimbang
Akurasi	90,83%	93,95%
Precision	Negative 91%, Positive 91%	Negative 94%, Positive 94%
Recall	Negative 86%, Positive 94%	Negative 94%, Positive 94%
F1-Score	Negative 88%, Positive 92%	Negative 94%, Positive 94%

Tabel 4, menunjukkan distribusi data memengaruhi akurasi. Saat data tidak seimbang, akurasi hanya 90,83%. Namun saat data seimbang, akurasi naik menjadi 93,95%. Ini

menunjukkan algoritma *Support Vector Machine* (SVM) bekerja lebih baik jika data seimbang antara sentimen positif dan negatif

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menunjukkan efektivitas algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen ulasan wisatawan terhadap objek wisata di Bali, dengan akurasi tertinggi mencapai 93,95% pada data seimbang. Dengan *Precision*, *recall* dan *F1-score* untuk kedua kelas (*Negative* dan *Positive*) masing-masing mencapai 94%, membuktikan keunggulan *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen ulasan dengan distribusi kelas yang baik, konsisten dan akurat. Untuk itu, saran yang dapat diberikan adalah memperluas cakupan data, mencoba algoritma pembandingan seperti *Deep Learning*, serta mengimplementasikan sistem ini secara luas untuk mendukung pengambilan keputusan strategis oleh pengelola destinasi wisata. Selain itu, penting untuk melakukan analisis lebih mendalam mengenai faktor-faktor yang mempengaruhi sentimen ulasan, seperti lokasi, fasilitas, dan pengalaman wisatawan, untuk meningkatkan akurasi model.

DAFTAR PUSTAKA

- [1] B. R. Atmadja, "Analisis Sentimen Bahasa Indonesia Pada Tempat Wisata di Kabupaten Sukabumi Dengan Naïve Bayes," vol. 15, no. 2, pp. 371–382, 2022, [Online]. Available: <http://journal.stekom.ac.id/index.php/elkom/page/371>
- [2] A. Maulana *et al.*, "POLA MUSIMAN KUNJUNGAN WISATAWAN MANCANEGARA KE BALI Seasonal Pattern of Foreign Tourist Arrivals to Bali," 2020.
- [3] R. Koen Iswandari and E. Setiyowati, "Pengaruh Media Sosial terhadap Keputusan Berkunjung di Pantai Panrita Lopi Muara Badak," *Edu tourism Journal Of Tourism Research*.
- [4] M. N. Muttaqin and I. Kharisudin, "Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor," *UNNES Journal of Mathematics*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [5] Kelvin, J. Banjarnahor, E. Indra, and S. H. Sinurat, "ANALISIS PERBANDINGAN SENTIMEN CORONA VIRUS DISEASE- 2019 (COVID19) PADA TWITTER MENGGUNAKAN METODE LOGISTIC REGRESSION DAN SUPPORT VECTOR MACHINE (SVM)," *JUSIKOM PRIMA (Jurnal Sistem Informasi dan Ilmu Komputer Prima)*, vol. 5, pp. 47–52, Feb. 2022.
- [6] S. Azza Amira, S. Utama, and dan Muhammad Hanif Fahmi, "Edu Komputika Journal Penerapan Metode Support Vector Machine untuk Analisis Sentimen pada Review Pelanggan Hotel," *Edu Komputika*, vol. 7, no. 2, 2020, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/edukom>
- [7] N. P. A. Dewi, N. A. Sanjaya ER, A. E. Karyawati, I. B. M. Mahendra, I. B. G. Dwidasmar, and I. G. A. Wibawa, "Analisis Sentimen Berbasis Aspek Ulasan Pelanggan Hotel di Bali Menggunakan Metode Decision Tree," *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, vol. 11, no. 3, p. 625, Jul. 2022, doi: 10.24843/jlk.2023.v11.i03.p19.
- [8] G. Fikri Baihaqi, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Wisata Alun-Alun Kota Batu menggunakan Algoritma Support Vector Machine," 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] J. Ipawati, S. Saifulloh, and K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/malcom.v4i1.1066.
- [10] M. Khoirul, U. Hayati, and O. Nurdian, "ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES," 2023.
- [11] Y. Findawati and M. A. Rosid, *BUKU AJAR TEXT MINING*. UMSIDA PRESS.

- [12] N. Arifin, U. Enri, and N. Sulistiyowati, "PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DENGAN TF-IDF N-GRAM UNTUK TEXT CLASSIFICATION," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 6, Dec. 2021.
- [13] W. Musu, A. Ibrahim, and Herjadi, "Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4.5," vol. X, pp. 186–195, Feb. 2021.
- [14] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, "PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER," *JOISM: JURNAL OF INFORMATION SYSTEM MANAGEMENT*, vol. 3, pp. 16–21, 2021.
- [15] F. Rahmad, Y. Suryanto, and K. Ramli, "Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, Aug. 2020. doi: 10.1088/1757-899X/879/1/012076.