

Analisis Sentimen Wisata Air Terjun Di Kabupaten Lombok Tengah Menggunakan Metode Support Vector Machine (SVM)

M. Syamsul Hadi^{1*}, Jihadul Akbar², Muhammad Fauzi Zulkarnain³

^{1,2} Program Studi Teknik Informatika, STMIK Lombok, Praya, Indonesia

³ Program Studi Sistem Informasi, STMIK Lombok, Praya, Indonesia

E-mail: ¹msyhp4lahady@gmail.com, ²jihadul4kbar@gmail.com,

³muhammadfauzizulkarnaen@gmail.com

(* : corresponding author)

Abstrak

Potensi pariwisata air terjun di Kabupaten Lombok Tengah cukup menonjol, akan tetapi belum dimanfaatkan secara optimal. Dengan kemajuan teknologi digital, tinjauan pengguna di platform online bisa digunakan sebagai penanda yang valid untuk menilai kualitas suatu tempat wisata. Tujuan dari penelitian ini adalah untuk menganalisis sentimen ulasan pengunjung terhadap objek wisata air terjun dengan menggunakan algoritma *Support Vector Machine* (SVM). Informasi dikumpulkan melalui proses *web scraping* terhadap ulasan Google Maps menggunakan layanan Apify.com, yang dilakukan pada tanggal 25 Maret 2025. Selanjutnya, data diproses melalui tahap preprocessing, dilabeli secara otomatis dengan menggunakan IndoBERT, dan fitur diekstrak dengan metode TF-IDF. Sebanyak 1.250 ulasan telah dianalisis dan dibagi ke dalam tiga kategori sentimen: positif, netral, dan negatif. Tiga jenis algoritma SVM diuji, yaitu *LinearSVC*, kernel RBF, dan kernel Polynomial. Dari hasil pengujian, *LinearSVC* menunjukkan kinerja terbaik dengan tingkat akurasi 84% dan F1-score 86%. Penemuan ini menunjukkan bahwa pendekatan berbasis *machine learning*, terutama SVM, sangat efektif dalam secara otomatis dan terstruktur mengenali persepsi pengunjung. Data ini juga dapat digunakan sebagai acuan dalam perumusan kebijakan sektor pariwisata yang didasarkan pada fakta.

Kata kunci: Analisis Sentimen, Wisata Air Terjun, SVM, TF-IDF, Lombok Tengah

Abstract

The tourism potential of waterfalls in Central Lombok Regency is quite prominent; however, it has not been optimally utilized. With the advancement of digital technology, user reviews on online platforms can be used as a valid indicator to assess the quality of a tourist attraction. The purpose of this study was to analyze the sentiment of visitor reviews of waterfalls using the Support Vector Machine (SVM) algorithm. Information was collected through web scraping of Google Maps reviews using the Apify.com service, conducted on March 25, 2025. Next, the data was preprocessed, automatically labeled using IndoBERT, and features extracted using the TF-IDF method. A total of 1,250 reviews were analyzed and divided into three sentiment categories: positive, neutral, and negative. Three types of SVM algorithms were tested: LinearSVC, RBF kernel, and Polynomial kernel. From the test results, LinearSVC showed the best performance with an accuracy rate of 84% and an F1-score of 86%. These findings indicate that machine learning-based approaches, especially SVM, are very effective in automatically and structuredly recognizing visitor perceptions. This data can also be used as a reference in formulating tourism sector policies based on facts.

Keywords: Central Lombok, Sentiment Analysis, SVM, TF-IDF, Waterfall Tourism

1. PENDAHULUAN

Pariwisata menjadi salah satu sektor utama yang berkontribusi secara signifikan terhadap peningkatan pendapatan ekonomi masyarakat dan daerah [1]. Di antara berbagai jenis objek wisata yang tersedia, wisata alam seperti air terjun memiliki daya tarik tersendiri karena menawarkan keindahan dan suasana alami yang khas. Kabupaten Lombok Tengah memiliki potensi wisata air terjun yang melimpah, namun pemanfaatannya masih belum optimal. Menurut data dari Badan Pusat Statistik Provinsi Nusa Tenggara Barat, jumlah tamu yang menginap di hotel berbintang pada bulan Juli 2024 mencapai 110.952 orang. Dari jumlah tersebut, 69.796 merupakan wisatawan domestik (62,91%) dan 41.156 merupakan wisatawan mancanegara (37,09%) [2]. Data ini menunjukkan bahwa minat terhadap pariwisata di Lombok cukup tinggi.

Namun, agar potensi wisata terus berkembang secara berkelanjutan, diperlukan upaya peningkatan kualitas layanan dan pemahaman terhadap kebutuhan serta persepsi pengunjung.

Pada era digital saat ini, platform seperti Google Maps yang memuat ulasan dari para wisatawan menjadi sarana yang relevan untuk mengevaluasi kualitas destinasi wisata. [3]. Ulasan tersebut mencerminkan pengalaman dan kepuasan pengunjung, sehingga dapat dianalisis untuk mengetahui persepsi publik secara objektif. Salah satu metode yang efektif dalam menganalisis opini pengalaman pengguna terhadap destinasi wisata adalah melalui pendekatan analisis sentimen. Analisis sentimen merupakan teknik berbasis otomatisasi yang bertujuan untuk memahami makna emosional dalam sebuah pernyataan melalui proses interpretasi, ekstraksi, dan pengolahan informasi dari teks [4]. Dengan menggunakan pendekatan tersebut, ulasan dapat dikelompokkan berdasarkan kategori sentimen, yaitu positif, netral, atau negatif [5]. Dalam penelitian analisis sentimen, *Support Vector Machine* (SVM) sering dipilih sebagai pendekatan karena akurasi yang tinggi dalam klasifikasi teks, termasuk ketika diterapkan pada ulasan dari para wisatawan [6].

Penelitian ini difokuskan pada ulasan wisatawan terhadap destinasi wisata air terjun di Kabupaten Lombok Tengah, di antaranya Air Terjun Benang Kelambu, Benang Stokel, Babak Pelangi, Batu Janggot, Elong Tune, Goa Batu Emas, dan Pantai Nampung. Destinasi-destinasi tersebut merupakan objek wisata alam yang cukup populer dan sering menjadi bahan ulasan di salah satu platform digital, yaitu Google Maps. Berdasarkan data dari pihak pengelola wisata Benang Kelambu, tercatat sebanyak 1.288 kunjungan wisatawan pada bulan Januari hingga Februari 2024 [7]. Fakta ini menunjukkan pentingnya menganalisis opini pengunjung secara otomatis guna memperoleh pemahaman menyeluruh terhadap persepsi publik. Permasalahan utama dalam penelitian ini adalah bagaimana mengklasifikasikan sentimen ulasan wisatawan secara akurat dan efisien.

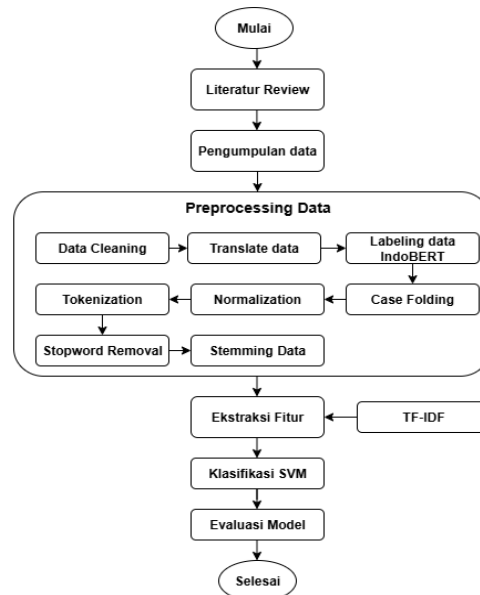
Berdasarkan permasalahan tersebut, *Support Vector Machine* (SVM) berperan dalam penelitian ini sebagai metode klasifikasi guna memetakan ulasan wisatawan ke dalam tiga tipe sentimen: positif, netral, dan negatif. [8].

Pemilihan SVM didasarkan pada hasil sejumlah penelitian sebelumnya yang menunjukkan bahwa algoritma ini unggul dalam klasifikasi teks. Misalnya, penelitian oleh Sinulingga dan Sitorus (2024) menggunakan SVM dan TF-IDF untuk menganalisis opini masyarakat terhadap film horor Indonesia dan berhasil mencapai akurasi 88,3% [9]. Penelitian lain oleh Ahmad Afif (2023) membandingkan metode Naïve Bayes dan SVM dalam klasifikasi ulasan hotel, di mana SVM menunjukkan performa akurasi sebesar 90%, melampaui Naïve Bayes yang hanya mencatat akurasi 82% [10]. Selain itu, Baita et al. (2021) SVM menunjukkan performa yang lebih unggul dalam analisis sentimen vaksin Sinovac dibandingkan K-Nearest Neighbor (KNN), sebagaimana terlihat dari akurasi yang dicapai, yakni 85% untuk SVM dan 80% untuk KNN [11].

Hasil dari penelitian-penelitian sebelumnya menunjukkan bahwa umumnya algoritma SVM memiliki tingkat keakuratan yang lebih tinggi daripada metode lain seperti KNN dan Naïve Bayes dalam analisis sentimen. Dikarenakan keunggulannya tersebut, SVM seringkali dipilih sebagai metode utama dalam klasifikasi sentimen. Tujuan dari penelitian ini adalah untuk menggunakan metode SVM dalam mengevaluasi pandangan wisatawan terhadap destinasi air terjun di Kabupaten Lombok Tengah. Diharapkan bahwa hasilnya dapat memberikan kontribusi untuk mendukung pihak pemerintah dan pengelola pariwisata dalam pengambilan keputusan, terutama dalam meningkatkan kualitas layanan dan pengembangan infrastruktur pariwisata. Di sisi lain, tujuan dari penelitian ini juga sebagai acuan untuk penelitian selanjutnya dalam penerapan analisis sentimen di sektor pariwisata dengan menggunakan *machine learning*.

2. METODE PENELITIAN

Penelitian ini menggunakan metode *Support Vector Machine* (SVM) untuk mengevaluasi pendapat pengunjung terhadap objek wisata air terjun di Kabupaten Lombok Tengah, seperti yang ditunjukkan dalam Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Di tengah perkembangan era digital, pengumpulan data memegang peranan penting dalam menghasilkan keputusan yang tepat dan didasarkan pada fakta.. Salah satu sumber data yang sangat bermanfaat adalah Google Maps, yang menyajikan informasi lokasi, bisnis, ulasan, dan berbagai detail geografis lainnya. Untuk mengoptimalkan pengambilan data ini, kita dapat memanfaatkan Apify.com, sebuah *platform web scraping* yang dapat mengekstraksi data dari berbagai situs, termasuk Google Maps, Twiter, dan Facebook dengan cara yang efisien dan terstruktur.

2.2. Preprocessing Data

Preprocessing merupakan tahap awal untuk menyaring dan mengolah data teks agar lebih terstruktur, sehingga dapat dimanfaatkan secara optimal dalam analisis sentimen [12]. Adapun tahapan preprocessing yang digunakan meliputi: pembersihan data (*cleaning*), penerjemahan (*translate*), pelabelan (*labeling*), perubahan huruf menjadi format seragam (*case folding*), normalisasi, tokenisasi, penghapusan kata umum (*stopword removal*), dan stemming.

a. Data Cleaning

Tahapan ini merupakan langkah awal yang esensial untuk menjamin kualitas data teks yang akan dianalisis. Proses pembersihan meliputi penghapusan data kosong, data ganda, karakter tak lazim, angka, tanda baca, serta teks tidak relevan seperti URL dan emoji yang berpotensi mengganggu akurasi analisis.

b. Translate Data

Translate data merupakan langkah menerjemahkan semua data kedalam bahasa Indonesia untuk memastikan konsistensi dan kemudahan pemrosesan teks.

c. Labeling Data

Penyebutan kategori adalah langkah penting dalam proses pengklasifikasian sentimen, di mana teks dikelompokkan ke dalam sentimen positif, netral, atau negatif. Salah satu metode yang diterapkan pada tahap ini adalah IndoBERT, sebuah model pra-latih yang dikembangkan dengan memanfaatkan lebih dari 4 miliar kata dalam korpus bahasa Indonesia [13]. Untuk mendukung klasifikasi yang tepat, label pada data ulasan secara otomatis diterapkan dan dikategorikan dalam tiga kelas sentimen utama: positif, netral, dan negatif [14].

d. Case Folding

Case folding melibatkan proses mengubah semua huruf dalam teks menjadi huruf kecil (*lowercase*) agar variasi penulisan seperti 'Bagus', 'bagus', dan 'BAGUS' dianggap sebagai kata yang sama.

e. *Normalization*

Normalisasi bertujuan untuk memperbaiki ejaan dan mengubah kata-kata tidak baku atau singkatan menjadi bentuk formal, seperti mengubah 'tdk' menjadi 'tidak', atau 'gpp' menjadi 'tidak apa-apa'.

f. *Tokenization*

Tokenisasi adalah pengubahan kalimat menjadi kata-kata atau frasa-frasa yang lebih kecil (token). Sebagai contoh, kalimat "Saya sangat senang hari ini!" dapat dipecah menjadi 'saya', 'sangat', 'senang', 'hari', dan 'ini'.

g. *Stopword Removal*

Tindakan ini dilakukan untuk menghilangkan kata-kata umum yang tidak memberikan kontribusi signifikan terhadap makna, seperti 'yang', 'dan', 'ke', atau 'di'. Meski sering muncul, kata-kata tersebut dianggap tidak bermanfaat dalam analisis sentimen karena kurangnya nilai informatif.

h. *Stemming Data*

Stemming adalah proses mengubah kata menjadi bentuk dasarnya dengan menghapus imbuhan. Misalnya, kata 'berlarian' akan berubah menjadi 'lari', dan 'makanannya' menjadi 'makan'.

2.3. Ekstraksi Fitur

Setelah proses pra-pemrosesan selesai, teks yang telah dibersihkan diubah menjadi bentuk numerik dengan mengekstraksi fitur-fiturnya. Salah satu metode yang sering digunakan adalah Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini mengevaluasi kepentingan setiap kata berdasarkan seberapa seringnya kata itu muncul dalam satu dokumen dibandingkan dengan kemunculannya di semua dokumen yang ada. Semakin sedikit kata-kata yang muncul secara utuh tetapi sering muncul dalam satu dokumen, semakin penting nilai mereka dalam analisis [15]. Rumus perhitungannya disajikan sebagai berikut.

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

Keterangan :

- $W_{i,j}$ adalah Bobot TF-IDF untuk term ke-j pada dokumen ke-i
- $tf_{i,j}$ adalah Frekuensi kemunculan term ke-j dalam dokumen ke-i (*Term Frequency*)
- N adalah Jumlah total dokumen
- df_j adalah Jumlah dokumen yang mengandung term ke-j (*Document Frequency*)
- $\log \left(\frac{N}{df_j} \right)$ adalah Inverse Document Frequency (IDF)

2.4. Support Vector Machine (SVM)

Salah satu teknik yang sering digunakan dalam supervised learning untuk memecahkan masalah klasifikasi teks adalah *Support Vector Machine* (SVM). Algoritma ini dapat memisahkan data dengan margin maksimum dan efektif bahkan pada data berdimensi tinggi, yang juga berlaku dalam konteks pariwisata [16]. SVM bertujuan untuk menemukan hyperplane optimal yang mampu memaksimalkan jarak terhadap titik-titik data dari masing-masing kelas. [17]. Rumus untuk menentukan hyperplane tersebut ditunjukkan sebagai berikut:

$$w \cdot x + b = 0 \rightarrow w_0 + w_1x_1 + w_2x_2 \quad (2)$$

Keterangan :

$W = \{w_1, w_2, \dots, w_n\} \rightarrow$ bobot

n = Jumlah atribut

b = Skalar yang digunakan sebagai bias

x = Nilai atribut

2.5. Evaluasi Model

Tahapan evaluasi dilakukan untuk menilai sejauh mana algoritma SVM mampu beroperasi secara optimal. [18]. Penelitian ini memanfaatkan Confusion Matrix sebagai sarana penilaian kinerja model klasifikasi. Dengan menggunakan Confusion Matrix, kita dapat mengetahui seberapa tepatnya model dalam mengklasifikasikan data ke dalam tiga kategori sentimen: positif, netral, dan negatif. Matriks ini mencatat jumlah prediksi yang tepat dan tidak tepat dari setiap kelas, yang memungkinkan perhitungan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Tiga jenis algoritma SVM yang dievaluasi dalam penelitian ini meliputi *LinearSVC*, *SVC* dengan kernel RBF, dan *SVC* dengan kernel Polinomial. Setiap model akan diuji dengan menggunakan 80% data latihan dan sisanya 20% akan digunakan untuk pengujian. Perbandingan antara prediksi hasil dan label asli digunakan untuk membuat Confusion Matrix, yang menjadi dasar dalam mengevaluasi kinerja model. Tabel 1 menunjukkan Struktur Confusion Matrix.

Tabel 1. Struktur *Confusion Matrix*

Nilai Prediksi	Nilai Aktual		
	Positif	Netral	Negatif
Positif	TP	FNt	FN
Netral	FP	TNt	FNt
Negatif	FP	FNt	TN

Keterangan:

- TP (*True Positif*): Prediksi positif dan memang positif
- TN (*True Negatif*): Prediksi negatif dan memang negatif
- TNt (*True Netral*): Prediksi netral dan memang netral
- FP (*False Positif*): Prediksi positif tapi bukan positif
- FN (*False Negatif*): Prediksi bukan negatif tapi sebenarnya negatif
- FNt (*False Netral*): Prediksi bukan netral tapi sebenarnya netral

Berdasarkan data pada *Confusion Matrix* di atas, maka evaluasi performa model SVM dapat dihitung menggunakan rumus sebagai berikut:

$$Accuracy = \frac{TP + TNt + TN}{TP + FP + TNt + FNt + TN + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FNt + FN} \quad (5)$$

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

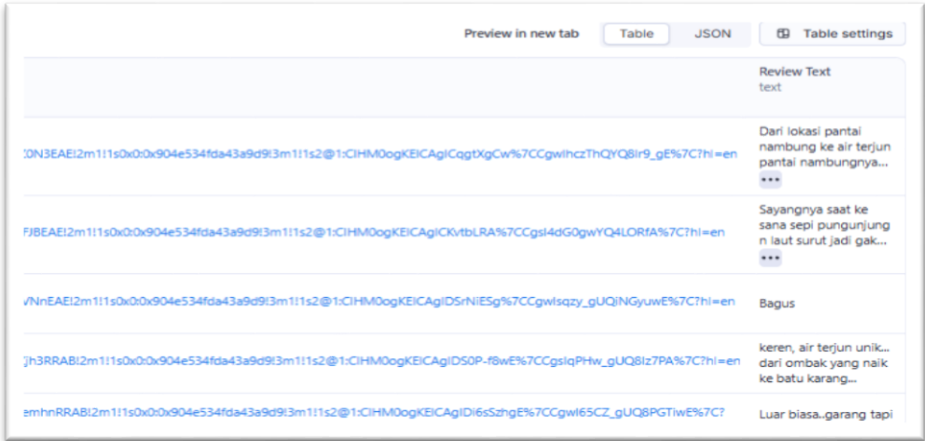
Ulasan pengguna Google Maps yang menyinggung tentang wisata air terjun di Kabupaten Lombok Tengah menjadi dasar dalam proses pengumpulan data pada penelitian ini. Setelah itu dilakukan proses *scraping* data. Scraping data merupakan teknik untuk mengumpulkan atau mengambil data dari situs web atau aplikasi [19].

Proses pengambilan data (*scraping*) dilakukan berdasarkan titik-titik lokasi wisata yang telah ditentukan secara spesifik. Data hasil *scraping* tersebut dapat ditelusuri lebih lanjut melalui tautan yang disajikan pada Tabel 2 berikut.

Table 2. Daftar Air Terjun di Kabupaten Lombok Tengah dan Tautan Google Maps

Air Terjun	Link Google Maps
Benang Stokel dan Benang Kelambu	https://maps.app.goo.gl/MkaVWMHtZSyDVAc7
Benang Setukel	https://maps.app.goo.gl/CCKGdEuRHhy4QqUD7
Babak Pelangi	https://maps.app.goo.gl/zM5QXfyxLxii8LFb7
Batu Janggut	https://maps.app.goo.gl/4LCt6FyG3iKAx9dv7
Gunung Janggut	https://maps.app.goo.gl/btMbTPX4EXbgEXzL8
Elong Tune	https://maps.app.goo.gl/TdwfDT9DR7KGMAxM7
Goa Batu Emas	https://maps.app.goo.gl/RZzXXEsp2w6QNXv89
Pantai Nambung	https://maps.app.goo.gl/U1Wdh8hdAXrsiwpN9
Benang Kelambu	https://maps.app.goo.gl/5UwTZh4WP8gshMZKA

Adapun pengambilan data dilakukan dengan menggunakan teknik *web scraping* melalui Website Apify.com, sebagaimana tergambar pada Gambar 2 berikut ini.



Gambar 2. Proses Scraping data

Setelah dilakukan proses *scraping* dari platform Google Maps terhadap sejumlah tempat wisata air terjun yang berada di wilayah Kabupaten Lombok Tengah, berhasil dikumpulkan sebanyak 2182 data ulasan dari para pengunjung.

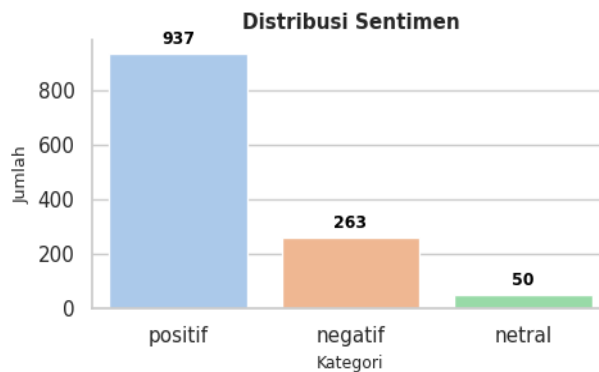
3.2. Preprocessing Data

Dalam tahap ini, data dipersiapkan melalui beberapa proses prapemrosesan, antara lain pembersihan, penerjemahan, pelabelan, *case folding*, normalisasi, tokenisasi, penghilangan *stopword*, serta *stemming*. Proses-proses ini bertujuan untuk menyempurnakan struktur teks sebelum memasuki tahap klasifikasi. Hasil akhirnya dapat dilihat pada Tabel 3.

Table 3. Hasil Preprocessing Data

Proses	Hasil	Label
Text	<i>The waterfall was nice and easy to reach. No guide was needed, you just follow the path.</i>	
Cleaning Data	<i>The waterfall was nice and easy to reach. No guide was needed, you just follow the path.</i>	
Translate Data	Air terjunnya bagus dan mudah dijangkau. Tidak ada panduan yang dibutuhkan, Anda hanya mengikuti jalan setapak.	
Labeling Data	Air terjunnya bagus dan mudah dijangkau. Tidak ada panduan yang dibutuhkan, Anda hanya mengikuti jalan setapak.	Positif
Case folding	air terjunnya bagus dan mudah dijangkau. tidak ada panduan yang dibutuhkan, anda hanya mengikuti jalan setapak.	
Normalization	air terjunnya bagus dan mudah dijangkau. tidak ada panduan yang dibutuhkan, anda hanya mengikuti jalan setapak.	
Tokenisasi	['air', 'terjunnya', 'bagus', 'dan', 'mudah', 'dijangkau', 'tidak', 'ada', 'panduan', 'yang', 'dibutuhkan', 'anda', 'hanya', 'mengikuti', 'jalan', 'setapak']	
Stopword Removal	air terjunnya bagus mudah dijangkau ada panduan dibutuhkan hanya mengikuti jalan setapak	
Stemming Data	air terjun bagus mudah jangkau ada pandu butuh hanya ikut jalan setapak	

Setelah melalui tahapan *preprocessing* yang mencakup pembersihan teks, normalisasi, dan penghapusan duplikasi, jumlah data yang layak untuk dianalisis sebanyak 1.250 ulasan pengunjung. Ulasan tersebut kemudian dikategorikan ke dalam tiga jenis sentimen, yaitu 937 ulasan positif, 263 ulasan negatif, dan 50 ulasan netral. Namun, distribusi data yang tidak merata antar kategori sentimen ini menjadi salah satu batasan dalam penelitian, karena model cenderung memperoleh representasi yang lebih kuat dari sentimen mayoritas. Kondisi ini berpotensi memengaruhi akurasi klasifikasi, khususnya dalam mendeteksi sentimen dengan jumlah data yang terbatas seperti kategori netral. Distribusi ketiga jenis sentimen tersebut ditampilkan dalam bentuk grafik batang pada Gambar 3 berikut.



Gambar 3. Distribusi Sentimen

3.3. Ekstraksi Fitur

Tahap setelah *preprocessing* adalah mengubah teks menjadi representasi numerik menggunakan TF-IDF. Metode ini memberikan bobot kata berdasarkan relevansi dalam dokumen dan kelangkaan dalam korpus. Source Code dan Hasil ekstraksi fitur dapat dilihat pada Gambar 4 dan 5 berikut.

```
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
import math

# Baca data
file_path = "/content/drive/MyDrive/TugasAkhir/data_stemmed.csv"
df = pd.read_csv(file_path)

# Cek kolom
if 'stemmed_text' in df.columns and 'sentiment' in df.columns:
    texts = df['stemmed_text'].fillna('').astype(str).tolist()
    y = df['sentiment']

    # TF
    count_vect = CountVectorizer()
    tf_matrix = count_vect.fit_transform(texts)
    tf_df = pd.DataFrame(tf_matrix.toarray(), columns=count_vect.get_feature_names_out())

    # IDF
    df_counts = (tf_matrix > 0).sum(axis=0).A1
    total_docs = tf_matrix.shape[0]
    idf_values = [math.log((total_docs + 1) / (df_count + 1)) + 1 for df_count in df_counts]
    idf_df = pd.DataFrame([idf_values], columns=count_vect.get_feature_names_out())

    # TF-IDF Manual
    tfidf_manual = tf_df * idf_df.values[0]

    # Simpan
    tfidf_manual.to_csv("/content/drive/MyDrive/TugasAkhir/TFIDF_manual.csv", index=False)
    y.to_csv("/content/drive/MyDrive/TugasAkhir/labels.csv", index=False)
    print("✅ TF-IDF dan label disimpan.")
else:
    print("❌ Kolom tidak ditemukan.")
```

Gambar 4. Hasil Ekstraksi Fitur TF-IDF

Contoh 10 baris pertama TF-IDF sparse format dengan kata:		
(Dokumen 0, Kata: 'air')	TF-IDF: 1.6304	
(Dokumen 0, Kata: 'biasa')	TF-IDF: 3.5674	
(Dokumen 0, Kata: 'luar')	TF-IDF: 3.5991	
(Dokumen 0, Kata: 'terjun')	TF-IDF: 1.7085	
(Dokumen 1, Kata: 'allah')	TF-IDF: 6.7454	
(Dokumen 1, Kata: 'masayaa')	TF-IDF: 7.0331	
(Dokumen 2, Kata: 'cocok')	TF-IDF: 4.4181	
(Dokumen 2, Kata: 'jadi')	TF-IDF: 3.2642	
(Dokumen 2, Kata: 'keren')	TF-IDF: 4.2605	
(Dokumen 2, Kata: 'menyenangkan')	TF-IDF: 4.4428	

Gambar 5. Hasil Ekstraksi Fitur TF-IDF

3.4. Klasifikasi dengan *Support Vector Machine (SVM)*

Klasifikasi dilakukan dengan menggunakan algoritma *Support Vector Machine (SVM)* untuk memisahkan ulasan menjadi tiga kategori sentimen, yaitu positif, netral, dan negatif. Jumlah data yang digunakan dalam proses klasifikasi ini adalah sebanyak 1.250 ulasan hasil dari tahapan preprocessing dan ekstraksi fitur menggunakan metode TF-IDF. Selanjutnya, data tersebut dibagi menjadi dua bagian, yaitu 80% sebagai data latih (sebanyak 1.000 data) dan 20% sebagai data uji (sebanyak 250 data). Pembagian dilakukan dengan menggunakan fungsi *train_test_split* dan pendekatan *stratified sampling* untuk menjaga keseimbangan proporsi setiap kelas sentimen dalam masing-masing subset. Tujuan pembagian ini adalah untuk memungkinkan model belajar dengan baik dari data latih dan mampu mengevaluasi kinerjanya ketika diuji dengan data baru yang belum pernah dikenali sebelumnya [14]. *Source Code* dan Hasil prediksi terhadap data latih dan data uji divisualisasikan pada Gambar 6, 7 dan Gambar 8.

```
# === Split 80:20 ===
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# === Latih Model ===
model = LinearSVC()
model.fit(X_train, y_train)

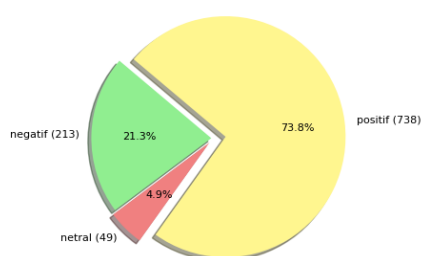
# === Prediksi ===
y_pred_train = model.predict(X_train)
y_pred_test = model.predict(X_test)

# === Mapping label jika perlu ===
label_map = {0: 'negatif', 1: 'netral', 2: 'positif'}

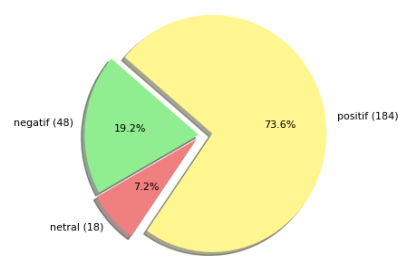
def safe_map_labels(series):
    return series.map(label_map) if series.dtype in ['int64', 'int32'] else series

# === Mapping label asli dan prediksi ===
y_train_mapped = safe_map_labels(pd.Series(y_train))
y_test_mapped = safe_map_labels(pd.Series(y_test))
y_pred_train_mapped = safe_map_labels(pd.Series(y_pred_train))
y_pred_test_mapped = safe_map_labels(pd.Series(y_pred_test))
```

Gambar 6. Source Prediksi Latih dan Uji Data



Gambar 7. Hasil Prediksi Data Latih (80%)



Gambar 8. Hasil Prediksi Pada Data Uji (20%)

3.5. Evaluasi Model

Evaluasi Model dilakukan menggunakan *Confusion Matrix* pada tiga varian algoritma SVM: *LinearSVC*, *RBF*, dan *Polynomial*. Proses ini memanfaatkan beberapa pustaka Python seperti *pandas*, *scikit-learn* untuk klasifikasi dan metrik evaluasi, serta *seaborn* dan *matplotlib* untuk visualisasi. Cuplikan source code pengujian model ditampilkan pada Gambar 9 berikut.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.svm import LinearSVC, SVC
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# Load Data
X = pd.read_csv("/content/drive/MyDrive/TugasAkhir/TFIDF_manual.csv")
y = pd.read_csv("/content/drive/MyDrive/TugasAkhir/labels.csv").squeeze()
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

models = {
    'LinearSVC': LinearSVC(max_iter=10000),
    'SVC_RBF': SVC(kernel='rbf'),
    'SVC_Poly': SVC(kernel='poly', degree=3)
}

for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)

    print(f"\n=== {name} ===")
    print("Akurasi:", round(accuracy_score(y_test, y_pred), 4))
    print("Classification Report:\n", classification_report(y_test, y_pred))
    cm = confusion_matrix(y_test, y_pred, labels=model.classes_ if hasattr(model, "classes_") else model.fit(X_train, y_train).classes_)

    plt.figure(figsize=(5, 4))
    sns.heatmap(cm, annot=True, fmt="d", cmap="Blues",
                xticklabels=model.classes_ if hasattr(model, "classes_") else model.fit(X_train, y_train).classes_,
                yticklabels=model.classes_ if hasattr(model, "classes_") else model.fit(X_train, y_train).classes_)

    plt.title(f"Confusion Matrix - {name}")
    plt.xlabel("Predicted Label")
    plt.ylabel("True Label")
    plt.tight_layout()
    plt.show()
```

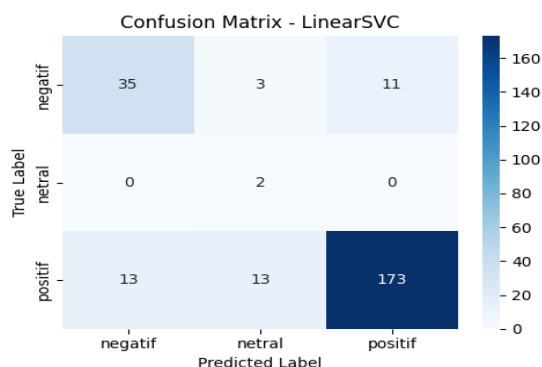
Gambar 9. Source Code Evaluasi Model

Berdasarkan Tabel 5 dibawah, model *LinearSVC* menunjukkan performa paling stabil dengan akurasi sebesar 84% dan F1-score 86%. Model *RBF* sedikit lebih tinggi dalam akurasi (85%), namun F1-score-nya lebih rendah (82%), yang menunjukkan ketidakseimbangan antara presisi dan *recall*. Model *Polynomial* menghasilkan performa terendah dengan F1-score sebesar 74%.

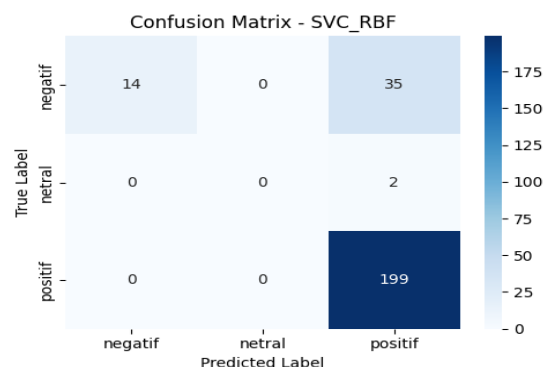
Table 4. Hasil Evaluasi Model (80:20)

Model	Akurasi	Presisi	Recall	F1-Score
LinearSVC	0.84	0.89	0.84	0.86
RBF	0.85	0.87	0.85	0.82
Polynomial	0.80	0.78	0.80	0.74

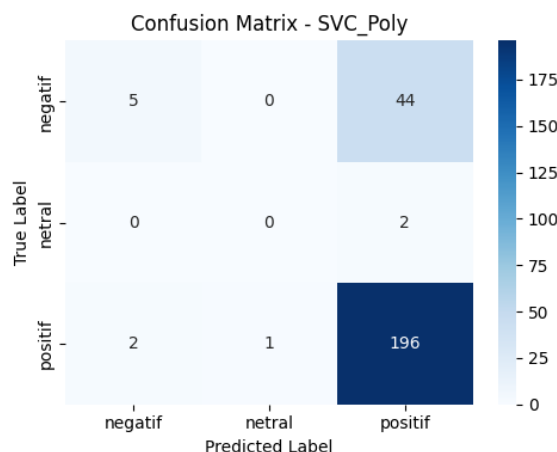
Visualisasi *Confusion Matrix* untuk ketiga model SVM Sebagaimana tertera pada gambar 10, 11, dan 12 berikut ini.



Gambar 10. Confusion Matrik LinierSCV



Gambar 11. Confusion Matrik RBF

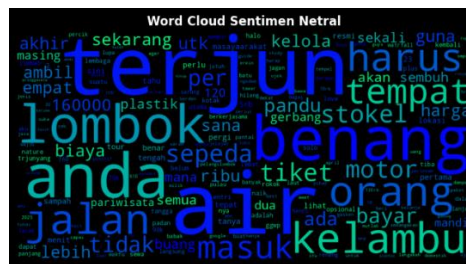


Gambar 12. Confusion Matrix Polynomial

Untuk menunjukkan kata-kata yang paling umum muncul dalam ulasan wisatawan, kami membuat *word cloud* yang didasarkan pada kategori sentimen masing-masing: positif, negatif, dan netral. Gambarnya menunjukkan hasil visualisasi *word cloud* pada Gambar 13, 14, dan 15 berikut.



Gambar 13. Word Cloud Positif



Gambar 14. Word Cloud Netral



Gambar 15. Word Cloud Negatif

4. KESIMPULAN DAN SARAN

Algoritma *Support Vector Machine* (SVM) berhasil diterapkan dalam klasifikasi sentimen ulasan wisatawan terhadap destinasi air terjun di Kabupaten Lombok Tengah. Dari 1.250 ulasan yang diperoleh melalui Google Maps, data diolah melalui tahap preprocessing, ekstraksi fitur TF-IDF, dan pelatihan model SVM dengan beberapa kernel. Hasil terbaik dicapai oleh model LinearSVC dengan akurasi 84% dan F1-score 86%, melebihi performa kernel RBF dan Polynomial. Temuan ini memperkuat efektivitas SVM dalam menganalisis sentimen ulasan berbahasa Indonesia.

Kelebihan pendekatan ini terletak pada kemampuannya menangani data besar secara sistematis, dimulai dari pengumpulan data, pelabelan otomatis dengan IndoBERT, hingga evaluasi menggunakan *Confusion Matrix*. Namun, keterbatasan masih muncul pada sebaran sentimen yang tidak merata dan jumlah data netral yang rendah, yang berpotensi mempengaruhi akurasi sistem klasifikasi.

Untuk penelitian selanjutnya, disarankan untuk memperhatikan distribusi data yang lebih seimbang, memperluas cakupan sumber ulasan, serta menguji algoritma lain seperti Random Forest, XGBoost, atau pendekatan deep learning. Selain itu, pendekatan *aspect-based sentiment analysis* juga dapat menjadi alternatif untuk analisis yang lebih mendalam terkait persepsi wisatawan.

DAFTAR PUSTAKA

- [1] A. Hanan, P. Gadu, and A. Sriwi, "Strategi Pengebangan Daya Tari Wisata Air Terjun Babak Pelangi Desa Lantan Kabupaten Lombok Tengah," *J. Responsible Tour.*, vol. 3, no. 1, pp. 65–76, 2023, doi: 10.47492/jrt.v3i1.2720.
- [2] BPS NTB, "Perkembangan Pariwisata Nusa Tenggara Barat," no. 51, 2024, [Online]. Available: <https://ntb.bps.go.id/id/pressrelease/2024/05/02/988/ntb--maret-2024--perkembangan-pariwisata.html>
- [3] C. P. Helena Panjaitan and C. Supriadi, "Analisis Sentimen Pengunjung Wisata Heritage Kota Semarang Menggunakan Naive Bayes Pada Ulasan Google Maps," *J. Elektron. dan Teknol. Inf.*, vol. 4, no. 2, pp. 2721–9380, 2023, doi: <https://doi.org/10.5201/jet.v4i2.412>.
- [4] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, "Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine," *JifoTech: Journal Information Technology*, vol. 2, no. 1, pp. 36–40, 2022, doi: <https://doi.org/10.46229/jifotech.v2i1.330>.
- [5] Alfandi Safira and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *Zo. J. Sist. Inf.*, vol. 5, no. 1, pp. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.
- [6] M. S. Syahlan, D. Irmayanti, and S. Alam, "Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Support Vector Machine (Svm)," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 8, no. 2, pp. 315–319, 2023, doi: 10.51876/simtek.v8i2.281.
- [7] N. Ikhsan, S. P. Damayanti, and P. Gadu, "Konsep Daya Tarik Wisata Berbasis Fasilitas Dalam Meningkatkan Minat Kunjungan Wisatawan Ke Wisata Benang Kelambu Di Desa Aik Berik," vol. 4, no. 2, pp. 1–23, 2016, doi: <https://doi.org/10.47492/jrt.v4i2.3664>.
- [8] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [9] J. E. Br Sinulingga and H. C. K. Sitorus, "Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF," *J. Manaj. Inform.*, vol. 14, no. 1, pp. 42–53, 2024, doi: 10.34010/jamika.v14i1.11946.
- [10] A. Afif, "Analisis Metode Klasifikasi Data Naïve Bayes dan SVM Dalam Menentukan Keunikan Hotel," *JUTEKINF (Jurnal Teknol. Komput. dan Informasi)*, vol. 11, no. 1, pp. 38–46, 2023, doi: 10.52072/jutekinf.v11i1.526.
- [11] A. Baita, Y. Pristyanto, and N. Cahyono, "Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (SVM) Dan K-Nearest Neighbor (KNN)," *Infos*, vol. 4, no. 2, pp. 42–42, 2021, doi: <https://doi.org/10.24076/infosjournal.2021v4i2.687>.
- [12] H. Sibyan and N. Hasanah, "Analisis Sentimen Ulasan Pada Wisata Dieng Dengan Algoritma K-Nearest Neighbor (K-NN)," *J. Penelit. dan Pengabd. Kpd. Masy. UNSIQ*, vol. 9, no. 1, pp. 38–47, 2022, doi: 10.32699/ppkm.v9i1.2218.
- [13] N. Surhasanah, R. Dwiyanaputra, S. I. Murpratiwi, and A. Aranta, "Analisis Sentimen Pengguna Platform Media Sosial X Pada Topik Pemilihan Presiden 2024 Menggunakan Perbandingan Model," vol. 9, no. 1, pp. 626–634, 2025, doi: <https://doi.org/10.36040/jati.v9i1.12430>.
- [14] T. Tarwoto, R. Nugroho, N. Azka, and W. S. R. Graha, "Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 9, no. pp. 495–505, 2025, doi: <https://doi.org/10.35870/jtik.v9i2.3340>.

- [15] K. Tri Putra, M. A. Hariyadi, and C. Crysdiyan, “Perbandingan Feature Extraction TF-IDF dan Bow untuk Analisis Sentimen Berbasis SVM,” *Jurnal Mandalika*, vol. 3, no. 2, pp. 1449–1463, 2023, doi: <https://doi.org/10.36312/jcm.v3i2>.
- [16] F. Rifaldy, Y. Sibaroni, and S. S. Prasetyowati, “Effectiveness of Word2vec and TF-IDF in Sentiment Classification on Online Investment Platforms Using Support Vector Machine,” vol. 10, no. 2, pp. 863–874, 2025, doi: <https://doi.org/10.29100/jipi.v10i2.6055>.
- [17] J. Ipmawati, S. Saifulloh, and K. Kusnawi, “Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 247–256, 2024, doi: [10.57152/malcom.v4i1.1066](https://doi.org/10.57152/malcom.v4i1.1066).
- [18] S. Rabbani, D. Safitri, and N. Rahmadhani, “Comparative Evaluation of SVM Kernels for Sentiment Classification in Fuel Price Increase Analysis Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM,” vol. 3, no. 2, pp. 153–160, 2023, doi: <https://doi.org/10.57152/malcom.v3i2.897>.
- [19] F. N. Salsabilla, and A. Witanti., “Analisis asentimen Akhir Masa Jabatan Presiden Jokowi Pada Media Sosial X Menggunakan Naïve Bayes,” *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 8, no. 1, pp. 106–115, 2025, doi: <https://doi.org/10.36080/skanika.v8i1.3331>.