



Analisis Sentimen Masyarakat Terhadap Objek Wisata Di Kabupaten Lahat Menggunakan Algoritma Support Vector Machine

Nadia Damayanti, Shinta Puspasari, Nazori Suhandi

¹⁻³Program Studi Teknik Informatika, Fakultas Ilmu Komputer dan Sains, Universitas Indo Global Mandiri

Korespondensi penulis: nadiadamayanti710@gmail.com

Abstract. Nature tourism is one of the sectors that plays an important role in supporting the development of regional tourism, including in Lahat Regency, which has significant waterfall tourism potential. Currently, many visitors share their reviews and experiences through digital platforms such as Google Maps. This review can be used as a source of information to understand the public's evaluation of the quality of tourist attractions. This study aims to examine public perception of tourist attractions in Lahat Regency using the Support Vector Machine (SVM) method. Research data were collected through scraping from Google Maps, totaling 500 reviews from five tourist attractions, namely Curup Maung, Curup Buluh, Senyawe Waterfall, Panjang Waterfall, and Green Canyon. The research stages include data preprocessing, consisting of cleaning, case folding, normalization, tokenization, stopword removal, and stemming. After that, feature extraction was carried out using the TF-IDF method and the classification process using the SVM algorithm. Based on the research results, the Support Vector Machine (SVM) method is able to perform sentiment classification quite well, although the accuracy level varies for each tourist attraction. Curup Maung and Panjang Waterfall achieved the highest accuracy level of 90%. Nevertheless, most visitor reviews were dominated by negative sentiments. This indicates that there are still several aspects that need to be improved, particularly related to tourist facilities and services. This research is expected to serve as a consideration for tourism managers and local governments in efforts to improve management quality as well as the development of tourism in Lahat Regency.

Keywords: Sentiment Analysis; Google Maps; Tourist Attractions; Support Vector Machine; Text Mining.

Abstrak. Pariwisata alam adalah salah satu sektor yang berkontribusi besar terhadap mendukung perkembangan pariwisata daerah, termasuk di Kabupaten Lahat yang memiliki potensi wisata air terjun cukup besar. Saat ini, banyak pengunjung memberikan ulasan serta pengalaman mereka pada platform digital seperti Google Maps. Review tersebut bisa digunakan sebagai sumber informasi guna mengetahui penilaian masyarakat pada kualitas objek wisata. Penelitian ini dilakukan untuk mengkaji persepsi masyarakat pada objek wisata di Kabupaten Lahat menggunakan metode Support Vector Machine (SVM). Pengumpulan data penelitian dilakukan melalui proses scraping dari Google Maps sebanyak 500 ulasan yang berasal dari lima objek wisata, yaitu Curup Maung, Curup Buluh, Air Terjun Senyawe, Air Terjun Panjang, dan Green Canyon. Tahapan penelitian mencakup proses prapengolahan data yang terdiri atas cleaning, case folding, normalisasi, tokenization, stopword removal, dan stemming. Setelah itu dilakukan ekstraksi fitur menggunakan metode TF-IDF dan proses klasifikasi menggunakan algoritma SVM. Berdasarkan hasil penelitian menunjukkan bahwa metode SVM bisa melakukan klasifikasi sentimen dengan cukup baik dengan tingkat akurasi yang berbeda pada setiap objek wisata. Curup Maung dan Air Terjun Panjang memperoleh tingkat akurasi tertinggi sebesar 90%, sedangkan sebagian besar ulasan pengunjung didominasi oleh sentimen negatif. Hasil tersebut menunjukkan bahwa masih ditemukan beberapa bagian yang masih diperbaiki, khususnya pada fasilitas dan pelayanan wisata. Penelitian ini diharapkan mampu menjadi bahan pertimbangan bagi pengelola wisata dan pemerintah daerah dalam meningkatkan kualitas pengelolaan dan pengembangan pariwisata di Kabupaten Lahat.

Kata kunci: Analisis Sentimen; Google Maps; Objek Wisata; Support Vector Machine; Text Mining.

1. LATAR BELAKANG

Pariwisata adalah suatu sektor penting yang dapat mendorong pertumbuhan ekonomi daerah. Keberadaan objek wisata tidak hanya meningkatkan jumlah kunjungan

wisatawan, namun juga memberikan dampak terhadap pendapatan daerah serta membuka peluang usaha bagi masyarakat sekitar (Pratama & Nugroho, 2021). Kabupaten Lahat adalah suatu wilayah di Provinsi Sumatera Selatan yang mempunyai potensi wisata alam yang cukup besar, khususnya wisata air terjun atau curup. Beberapa objek wisata yang cukup dikenal antara lain Curup Maung, Curup Buluh, Air Terjun Senyawe, Air Terjun Panjang, dan Green Canyon. Potensi wisata alam tersebut memiliki daya tarik bagi wisatawan, namun pengelolaan dan pengembangannya masih perlu ditingkatkan berdasarkan penilaian dan pengalaman pengunjung.

Menurut (Suhel & Bashir, 2021) Perkembangan teknologi informasi membuat wisatawan semakin aktif memberikan ulasan terhadap tempat yang mereka kunjungi melalui *platform* digital seperti *Google Maps*. *Platform* ini memungkinkan pengguna memberikan komentar serta penilaian pada suatu tempat wisata sehingga bisa berperan sebagai sumber informasi yang bernilai untuk calon pengunjung maupun pengelola wisata. Namun, jumlah ulasan yang semakin banyak menyebabkan proses analisis secara manual menjadi masih belum optimal. Oleh sebab itu, dibutuhkan suatu metode yang mampu mengolah dan menganalisis opini masyarakat secara otomatis.

Analisis sentimen merupakan salah satu metode dalam *data mining* yang dimanfaatkan guna menggali opini, sikap, dan penilaian masyarakat berdasarkan data berbasis teks. Teknik ini memungkinkan data ulasan yang tidak terstruktur diklasifikasikan ke dalam kelompok sentimen positif, negatif, maupun netral. Sejumlah penelitian terdahulu telah menerapkan analisis sentimen dalam berbagai bidang, termasuk analisis ulasan wisata. Penelitian menggunakan metode *Support Vector Machine (SVM)* untuk melakukan ulasan wisata pada *Google Maps* dengan tingkat akurasi sebesar 83,8% (Ipmawati et al., 2024). Penelitian lain yang menggunakan metode *Decision Tree* dengan akurasi sebesar 79,6% (Rahman Hakim, 2024). Selain itu, penelitian yang menganalisis sentimen terhadap objek wisata air terjun di Kabupaten Lombok Tengah dengan memanfaatkan algoritma *SVM* menunjukkan akurasi sebesar 89% (Syamsul Hadi, 2025). Hasil penelitian tersebut mengidentifikasi bahwa metode *SVM* mempunyai kemampuan yang efektif saat melakukan klasifikasi sentimen pada data teks.

Walaupun penelitian mengenai analisis sentimen telah berkembang cukup pesat, penelitian yang secara khusus membahas menganalisis ulasan pengunjung terhadap

objek wisata di Kabupaten Lahat masih terbatas. Padahal, informasi mengenai opini pengunjung sangat penting sebagai bahan evaluasi dalam meningkatkan kualitas pengelolaan dan pengembangan destinasi wisata. Berdasarkan uraian tersebut, penelitian ini dilakukan guna mengidentifikasi lebih lanjut sentimen dari ulasan pengunjung pada sejumlah objek wisata di Kabupaten Lahat yang diperoleh melalui *Google Maps* dengan menerapkan metode *Support Vector Machine (SVM)*. Penelitian ini diharapkan mampu memberikan deskripsi terkait pandangan wisatawan serta menjadi bahan pertimbangan dalam pengembangan pariwisata di Kabupaten Lahat.

2. KAJIAN TEORITIS

Analisis Sentimen

Analisis sentimen adalah salah satu metode pada *text mining* yang dimanfaatkan untuk menganalisis opini atau pendapat seseorang terhadap suatu topik yang disampaikan dalam bentuk teks. Menurut (Khalim et al., 2023) data yang dianalisis biasanya berasal dari media sosial, forum diskusi, maupun ulasan pengguna pada *platform* digital. Dengan menggunakan analisis sentimen, pendapat tersebut bisa dikelompokkan ke dalam beberapa kelompok, seperti sentimen positif, negatif, ataupun netral, sehingga mampu memberikan gambaran terkait persepsi masyarakat terhadap suatu objek tertentu (Giovani et al., 2021).

Analisis sentimen digunakan untuk mengekstraksi dan mengolah informasi dari data teks agar dapat diketahui kecenderungan opini atau penilaian masyarakat terhadap suatu produk, layanan, atau aplikasi (Asmara et al., 2024). Dengan menggunakan metode analisis sentimen, data opini yang jumlahnya besar dapat dianalisis secara otomatis sehingga menghasilkan informasi yang lebih mudah dipahami dan dapat digunakan sebagai bahan evaluasi maupun pengambilan keputusan (Herdiansyah et al. 2026).

Sejarah Kabupaten Lahat

Kota Lahat di Provinsi Sumatera Selatan diketahui mempunyai kekayaan alam yang beragam, seperti perbukitan, sungai, dan hutan yang masih alami. Potensi alam tersebut menjadikan sektor pariwisata sebagai salah satu faktor penting dalam mendukung perekonomian daerah (Syamsuar et al., 2022). Dengan keindahan alam yang dimiliki, Lahat memiliki prospek yang menjanjikan untuk dikembangkan menjadi destinasi wisata yang banyak menarik minat wisatawan, baik lokal ataupun internasional.

Selain wisata alam, Kota Lahat juga mempunyai potensi wisata budaya yang dipengaruhi oleh keberagaman latar belakang sejarah dan budaya masyarakatnya. Oleh karena itu, pengembangan sektor pariwisata perlu didukung oleh peningkatan infrastruktur serta promosi yang tepat agar potensi wisata dapat dimaksimalkan. Secara keseluruhan, pengembangan pariwisata di Kota Lahat tidak hanya bertujuan mempromosikan keindahan alam, tetapi juga untuk memberdayakan masyarakat lokal, menjaga kelestarian lingkungan, serta memperkenalkan budaya daerah.

Ulasan Publik

Ulasan publik merupakan pendapat atau tanggapan pengguna terhadap suatu produk, layanan, atau tempat yang disampaikan melalui media daring. Ulasan ini biasanya berisi pengalaman dan penilaian pengguna sehingga dapat menjadi sumber informasi penting untuk mengetahui persepsi masyarakat terhadap suatu objek.

Komentar pada *platform* seperti *Google Maps* dapat dimanfaatkan sebagai data dalam penelitian analisis sentimen karena umumnya bersifat jujur dan spontan (Ramadhan & Putra, 2021). Dalam penelitian ini, ulasan publik digunakan sebagai sumber data untuk menganalisis persepsi masyarakat mengenai objek wisata di Kabupaten Lahat menggunakan metode *text mining*, sehingga dapat diketahui kecenderungan opini positif maupun negatif dari masyarakat (Widodo & Hartati, 2022)..

Text Mining

Text mining merupakan metode yang digunakan supaya menggali dan mengidentifikasi informasi atau pola tertentu dari kumpulan data teks yang bersifat tidak terstruktur. Dalam bidang pariwisata, metode ini dapat dimanfaatkan untuk menganalisis ulasan dan opini masyarakat mengenai destinasi wisata sehingga memberikan informasi yang berguna bagi pengembangan kebijakan maupun strategi pemasaran. Melalui *text mining*, pengelola destinasi wisata dapat mengetahui pandangan pengunjung terhadap fasilitas, layanan, dan daya tarik wisata yang tersedia.

Adapun penelitian yang menunjukkan bahwa analisis sentimen terhadap ulasan wisatawan mampu mengungkap aspek yang disukai maupun kurang disukai oleh pengunjung sehingga dapat menjadi bahan evaluasi dalam pengembangan sektor pariwisata (Irwansyah et al., 2023). Selain itu juga menyatakan bahwa pemanfaatan data dari media sosial dan ulasan daring dapat membantu meningkatkan pengalaman

pengunjung serta menyusun strategi pemasaran yang lebih efektif (Kristian et al., 2021). Dengan demikian, penerapan *text mining* dapat mendukung pengembangan pariwisata berbasis data serta meningkatkan kualitas layanan dan pengelolaan destinasi wisata.

Preprocessing

Preprocessing data merupakan tahap awal pada menganalisis teks yang bertujuan agar mempersiapkan data mentah supaya bisa diolah secara efektif dengan bantuan algoritma *machine learning*. Tahapan ini dilakukan menggunakan membersihkan dan menyederhanakan data sehingga informasi yang dihasilkan menjadi lebih relevan serta bebas dari karakter atau elemen yang tidak diperlukan.

Pada data teks, tahapan pra-pemrosesan umumnya mencakup proses *cleaning*, *case folding*, Normalisasi *tokenization*, *stopword removal*, serta *stemming*. Tahapan ini membantu mengurangi kata-kata yang dianggap tidak relevan dihilangkan, sementara kata yang berpengaruh dalam menentukan konteks atau sentimen suatu dokumen.

Adapun langkah-langkah pra-pemrosesan yang diterapkan dalam penelitian ini meliputi:

1. *Cleaning*

Menghilangkan tanda baca, angka, emotikon, serta karakter lain yang tidak diperlukan agar teks menjadi lebih rapi.

2. *Case Folding*

Mengonversi setiap karakter pada teks ke bentuk huruf kecil (*lowercase*) agar penulisan lebih konsisten.

3. Normalisasi

Normalisasi merupakan proses mengubah kata yang tidak baku, singkatan, atau kata yang tidak sesuai dengan ejaan menjadi bentuk kata yang baku sehingga lebih mudah diproses oleh sistem.

4. *Tokenization*

Memisahkan teks ke dalam bentuk kata kecil berupa kata maupun token sehingga bisa dianalisis oleh sistem.

5. *Stopword Removal*

Menghilangkan kata-kata umum yang tidak mempunyai dampak besar terhadap makna teks, seperti “yang”, “dan”, “atau”.

6. *Stemming*

Mengonversi kata berimbuhan menjadi bentuk dasar, seperti “bermain”, “memainkan”, dan “pemain” menjadi “main”.

TF-IDF

TF-IDF (*Term Frequency–Inverse Document Frequency*) adalah metode pemberian bobot pada proses pengolahan teks guna untuk mengubah dokumen ke dalam bentuk representasi numerik. Pendekatan ini mencakup dua bagian utama, yaitu *Term Frequency* (*TF*) yang menandakan tingkat tingkat munculnya kata pada suatu dokumen, serta *Inverse Document Frequency* (*IDF*) yang menunjukkan tingkat kelangkaan kata tersebut untuk keseluruhan dokumen.

Menurut Zhafira et al. (2021) *TF-IDF* memberikan skor penilaian lebih besar terhadap kata yang dominan dalam sebuah dokumen jarang dijumpai pada dokumen lain, sehingga kata tersebut dianggap lebih mewakili isi dokumen. Metode ini kerap diterapkan pada *Natural Language Processing* (*NLP*) seperti analisis sentimen dan pengelompokan teks *TF-IDF* dapat meningkatkan akurasi analisis opini pada media sosial.

Support Vector Machine

Algoritma *Support Vector Machine* (*SVM*) merupakan metode *supervised learning* yang dipakai baik untuk keperluan klasifikasi serta regresi. *SVM* bekerja untuk menentukan *hyperplane* optimal yang mampu membagi dua kelas data dengan jarak pemisah paling besar sehingga menghasilkan pemisahan yang optimal (Saputra, 2024)..

SVM menggunakan data pelatihan yang sebelumnya telah memiliki label untuk membangun model yang dapat mengklasifikasikan data baru ke dalam kelas tertentu. Karena kemampuannya mengenali pola dengan baik, metode ini sering diterapkan dalam penelitian berbasis teks, termasuk pada penerapan analisis sentimen (Kanaka et al., 2024).

Confusion Matrix

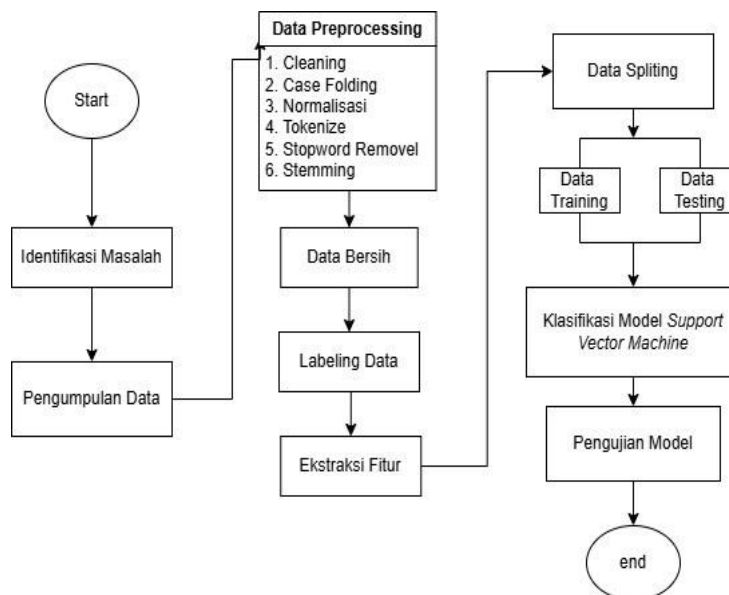
Confusion Matrix merupakan teknik pengujian yang bertujuan untuk menilai performa model klasifikasi melalui perbandingan hasil prediksi dengan data sebenarnya. Matriks ini memiliki empat komponen utama yaitu *True Positive* (*TP*), *True Negative* (*TN*), *False Positive* (*FP*), dan *False Negative* (*FN*) yang

mempresentasikan hasil prediksi model terhadap data sebenarnya (Algoritme et al., 2025)..

Confusion Matrix dimanfaatkan untuk menghitung semua metrik penilaian berupa *accuracy*, presisi, *recall*, dan *F1-score*, serta sering dimanfaatkan pada analisis sentimen untuk mengevaluasi performa algoritma klasifikasi seperti *Support Vector Machine (SVM)*, *Naive Bayes*, dan *K-Nearest Neighbor (KNN)*.

3. METODE PENELITIAN

Penelitian ini dilakukan dengan sejumlah tahapan yang dirancang secara terstruktur agar proses pelaksanaannya berjalan sesuai dengan tujuan yang telah ditetapkan. Rangkaian tahapan penelitian tersebut disajikan dalam bentuk diagram alur yang ditampilkan pada gambar berikut



Gambar 1. Tahap Penelitian

Identifikasi Masalah

Identifikasi masalah merupakan tahapan untuk mengenali dan merumuskan permasalahan yang muncul dari fenomena di lapangan. Pada penelitian ini, objek wisata di Kabupaten Lahat memiliki banyak ulasan dari masyarakat di *Google Maps*, namun informasi tersebut belum dimanfaatkan secara optimal. Komentar yang tersedia sangat beragam, tidak terstruktur, serta memuat pengalaman positif maupun keluhan pengunjung sehingga sulit dianalisis secara manual.

Selain itu, belum terdapat penelitian yang secara khusus menganalisis opini masyarakat terhadap objek wisata di Kabupaten Lahat dengan menerapkan algoritma

Support Vector Machine (SVM). Oleh sebab itu, diperlukan analisis yang lebih terarah untuk mengolah ulasan dari *Google Maps* agar dapat memberikan gambaran yang jelas terkait persepsi masyarakat terhadap objek wisata di Kabupaten Lahat.

Pengumpulan Data

Proses pengumpulan data pada tahap pengumpulan data ini dilaksanakan memakai teknik *scraping* untuk memperoleh komentar serta ulasan masyarakat dari *platform Google Maps* pada objek wisata yang diteliti. Menurut Erama (2025), proses *scraping* dilakukan menggunakan *platform Apify* sebagai alat bantu untuk mengekstraksi data ulasan secara otomatis dan lebih efisien.

Data yang diperoleh kemudian disimpan dalam bentuk *Excel* dan dilakukan pemeriksaan untuk memastikan tidak terdapat duplikasi, spam, atau komentar yang tidak relevan. Dataset yang telah melewati tahapan pemeriksaan selanjutnya digunakan pada proses *preprocessing* dan diolah dengan algoritma *Support Vector Machine (SVM)* agar dapat mengetahui pendapat masyarakat. Dari proses pengumpulan data tersebut diperoleh 500 data ulasan, dengan masing-masing objek wisata sebanyak 100 ulasan.

Data Preprocessing

Data preprocessing adalah tahap awal pada proses pengolahan data yang dilakukan untuk memastikan data siap sebelum proses analisis. Pada tahap ini terdapat beberapa langkah yaitu *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan *stemming*. Tahapan tersebut dilakukan untuk menghilangkan data yang tidak relevan sekaligus mengonversi teks menjadi bentuk yang lebih sesuai sehingga bisa diolah pada tahap analisis selanjutnya.

Data Bersih

Data bersih merupakan data yang telah melalui tahap *preprocessing* sehingga terbebas dari kata-kata yang tidak relevan, duplikasi, serta karakter yang tidak diperlukan. Dalam tahapan ini, teks ulasan telah menggunakan banyak tahapan seperti *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, serta *stemming* sehingga memperoleh data yang lebih terorganisir dan bisa digunakan pada proses analisis selanjutnya.

Labeling Data

Labelling data merupakan proses memberi pelabelan untuk setiap data ulasan untuk menentukan jenis sentimen yang terdapat pada komentar tersebut. Pada penelitian

ini, proses pelabelan dilakukan dengan mengelompokkan ulasan ke dalam dua kelompok yaitu sentimen positif dan sentimen negatif. Ulasan yang berisi pendapat baik, kepuasan, atau penilaian positif terhadap objek wisata diberikan label positif, sedangkan ulasan yang berisi keluhan, kritik, atau penilaian kurang baik diberikan label negatif. Untuk mengetahui distribusi jumlah data pada masing-masing label, digunakan perhitungan persentase dengan rumus sebagai berikut:

$$\text{Sentiment Score} = \sum_{i=1}^n \text{Polarity}(w_i)$$

Berarti:

\sum : Penjumlahan seluruh nilai

$i=1$: Kata pertama

n : Jumlah total kata dalam komentar

$\text{Polarity}(w_i)$: Jumlah kelas positif yang diklasifikasi sebagai negatif

Ekstraksi Fitur

Ekstraksi fitur adalah proses supaya mengonversi data teks ke dalam bentuk angka supaya bisa diolah menggunakan algoritma klasifikasi. Dalam penelitian ini, ekstraksi fitur diterapkan dan memakai metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. Metode ini dimanfaatkan agar menghitung skor tiap kata berdasarkan tingkat seberapa sering tingkat kelangkaannya pada seluruh dokumen.

Perhitungan *Term Frequency (TF)* digunakan untuk memperoleh frekuensi kemunculan setiap kata dalam dokumen. Rumus *TF* adalah sebagai berikut:

$$TF = \frac{\text{Jumlah kemunculan kata pada } D}{\text{Jumlah seluruh data dalam } D}$$

$$IDF = \log \frac{D}{DF}$$

$$TF - IDF = tf * idf$$

Keterangan:

D : jumlah dokumen yang ada dalam data training

DF : jumlah keseluruhan dokumen yang memuat data tersebut

TF : *Term Frequency*/kemunculan suatu kata dalam dokumen

IDF : *Inverse dokumen frequency term*/kata

Data Splitting

Data splitting merupakan proses pembagian dataset dikelompokkan menjadi dua bagian, yakni data latih (*training*) dan data uji (*testing*). Data training dipakai sebagai proses pelatihan model, sedangkan data uji dimanfaatkan untuk menilai kinerja model. Pada penelitian ini, data yang sudah menjalani tahap *preprocessing*, pelabelan, dan ekstraksi fitur kemudian dibagi untuk digunakan pada proses klasifikasi dilakukan dengan menggunakan algoritma *Support Vector Machine (SVM)*.

Model Support Vector Machine

Support Vector Machine (SVM) merupakan metode klasifikasi yang berfungsi memisah data ke dalam dua kelompok dengan menentukan *hyperplane* optimal yang memiliki margin terbesar (Efrizoni, 2022). Pada penelitian ini, *SVM* digunakan untuk mengelompokkan ulasan yang telah melewati tahap ekstraksi fitur yang telah diekstraksi menggunakan *TF-IDF* pada kelompok sentimen positif dan sentimen negatif.

Berdasarkan matematis, fungsi keputusan pada metode *SVM* dapat dituliskan sebagai berikut:

$$y_i(\omega^T x_i + b) \geq 1$$

Keterangan :

x_i : Data masukan atau vektor fitur ke- i .

y_i : Label kelas data ke- i (+1 untuk positif, -1 untuk negatif).

w : Bobot yang menentukan arah bidang pemisah (*hyperplane*).

b : Nilai bias yang menggeser posisi *hyperplane*.

$(\omega^T x_i + b)$: Persamaan bidang pemisah antara dua kelas.

≥ 1 : Menunjukkan bahwa data harus terletak pada sisi yang sesuai dari *hyperplane* dengan jarak sekurang-kurangnya 1 unit.

Confusion Matrix

Confusion Matrix merupakan metode pengujian yang dilakukan untuk mengukur performa model klasifikasi berdasarkan perbandingan hasil prediksi sistem terhadap data aktual. Matriks ini terdiri atas empat komponen utama yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* yang menunjukkan jumlah performa benar maupun salah yang dihasilkan oleh model.

Berdasarkan hasil tersebut, dapat dihitung beberapa ukuran evaluasi untuk mengetahui performa model klasifikasi. Adapun metrik yang dipakai pada penelitian ini yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Rumus *Accuracy* digunakan untuk menilai tingkat akuratan model dalam melakukan klasifikasi seluruh data.

$$Precision\ Positif = \frac{TP}{TP + FP}$$

$$Precision\ Negatif = \frac{TN}{TN + FN}$$

Rumus *Precision* digunakan untuk mengetahui seberapa akurat model dalam memprediksi data yang termasuk ke dalam kelas positif.

$$Recall\ Positif = \frac{TP}{TP + FN}$$

$$Recall\ Negatif = \frac{TN}{TN + FP}$$

Rumus *Recall* dipakai untuk mengetahui kemampuan model dalam mengklasifikasikan data yang termasuk kelas positif secara tepat.

$$F_1 = 2 \times \frac{Precision \cdot Recall}{Precision + Recall}$$

$$= \frac{2 \cdot TP}{2 \cdot TP + FN + FP}$$

Rumus *F1-Score* merupakan nilai yang diperoleh dari kombinasi antara *Precision* dan *Recall* yang berfungsi untuk menunjukkan keselarasan performa model dalam proses klasifikasi.

4. HASIL DAN PEMBAHASAN

A. Pengumpulan Dataset

Pada Penelitian ini data yang digunakan didapat melalui proses *scraping* ulasan dari *Google Maps* menggunakan *platform Web Apify*. Dari proses *scraping* tersebut diperoleh sebanyak 100 data ulasan pada setiap objek wisata. Hasil data *scraping* tersebut ditampilkan pada tabel berikut

Tabel 1. Contoh data komentar

No.	Data

0	Akses jalan menuju air terjun sangat jelek
1	Lumayan bagus tetapi akses kesana agak jauh
2	Tempat yang keren dan sangat seru untuk menghabiskan waktu liburan

B. Hasil Preprocessing

Cleaning

```

text = re.sub(r'http\S+|www\S+|https\S+', '', text)
text = re.sub(r'@\w+|\#\w+', '', text)
emoji_pattern = re.compile(
    "[
    u"\U0001F600-\U0001F64F"
    u"\U0001F300-\U0001F5FF"
    u"\U0001F680-\U0001F6FF"
    u"\U0001F1E0-\U0001F1FF"
    u"\U00002700-\U000027BF"
    u"\U000024C2-\U0001F251"
    ]+", flags=re.UNICODE)
text = emoji_pattern.sub(r'', text)
text = re.sub(r'^a-zA-Z\s', ' ', text)
text = ' '.join([word for word in text.split() if len(word) > 2])
text = re.sub(r'\s+', ' ', text).strip()
    
```

Gambar 2. Source Code Cleaning

Tabel 2. Cleaning

Text	Sesudah di <i>Cleaning</i>
Curup ini walaupun perjalanan yg sangat ekstrim, tetapi nikmat, begitu indah dan alami.ðŸ‘ðŸ‘ðŸ‘ kunjungi yaaa agar tidak penasaran.	Curup ini walaupun perjalanan yang sangat ekstrim, tetapi nikmat, begitu indah dan alami. Kunjungi ya agar tidak penasaran.

Case Folding

```

def case_folding(text):
    return text.lower()

# Terapkan ke DataFrame
df['case_folding'] = df['cleaning'].apply(case_folding)
    
```

Gambar 3. Source Code Case Folding

Tabel 3. Case Folding

Text	Sesudah <i>Case Folding</i>
Curup ini walaupun perjalanan yang	curup ini walaupun perjalanan yg sangat

sangat ekstrim, tetapi nikmat, begitu indah dan alami. Kunjungi ya agar tidak penasaran.	ekstrim tetapi nikmat begitu indah dan alami kunjungii yaaa agar tidak penasaran
--	--

Normalisasi

```
def normalisasi_text(text):
    words = text.split()
    normalized_words = []

    for word in words:
        if word in kamus_dict:
            normalized_words.append(kamus_dict[word])
        else:
            normalized_words.append(word)
```

Gambar 4. Source Normalisasi

Tabel 4. Normalisasi

Text	Sesudah Normalisasi
curup ini walaupun perjalanan yg sangat ekstrim tetapi nikmat begitu indah dan alami kunjungii yaaa agar tidak penasaran	curup ini walaupun perjalanan yang sangat ekstrim tetapi nikmat begitu indah dan alami kunjungii ya agar tidak penasaran

Tokenize

```
def tokenizing(text):
    return text.split()

# Terapkan ke DataFrame
df['tokenizing'] = df['case_folding'].apply(tokenizing)
```

Gambar 5. Source Tokenize

Tabel 5. Tokenize

Text	Sesudah Tokenize
curup ini walaupun perjalanan yang sangat ekstrim tetapi nikmat begitu indah dan alami kunjungii ya agar tidak penasaran	[curup, iini, walaupun, perjalanan, yang, sangat, ekstrim, tetapi, nikmat, begitu, indah, dan, alami, kunjungii, ya, agar, tidak, penasaran]

Stopword Removal

```
def remove_stopwords(tokens):
    return [word for word in tokens if word not in stop_words]
```

Gambar 6. Source Code Stopword Removal

Tabel 6. Stopword Removal

Text	Sesudah <i>Stopword Removal</i>
[curup, ini, walaupun, perjalanan, yang, sangat, ekstrim, tetapi, nikmat, begitu, indah, dan, alami, kunjungi, ya, agar, tidak, penasaran]	[curup, perjalanan, ekstrim, nikmat, indah, alami, kunjungi, penasaran]

Stemming

```
def stemming_text(tokens):
    return [stemmer.stem(word) for word in tokens]
df['stemming'] = df['stopword_removal'].apply(stemming_text)
```

Gambar 7. Source Stemming

Tabel 7. Stemming

Text	Sesudah <i>Stemming</i>
[curup, perjalanan, ekstrim, nikmat, indah, alami, kunjungi, penasaran]	[curup, buluh, jalan, ekstrim, nikmat, indah, alami, kunjung, penasaran]

C. Pelabelan Data

Tahap berikutnya adalah proses pelabelan untuk menentukan kategori sentimen pada setiap komentar wisatawan. Pada penelitian ini, pelabelan diterapkan berbasis pendekatan berbasis *lexicon-based* dengan memanfaatkan kamus kata positif serta negatif sebagai dasar penentuan sentimen. Proses pelabelan data bisa dilihat pada Gambar 8. *Source Code* Pelabelan Data.

Penentuan sentimen dilakukan dengan menghitung jumlah kata positif dan kata negatif pada setiap komentar. Nilai score diperoleh dari selisih antara jumlah kata positif dan negatif. Jika skor bernilai positif, maka komentar dikategorikan sebagai sentimen positif, sedangkan apabila skor bernilai negatif dikategorikan sebagai sentimen negatif. Contoh hasil pelabelan data ditampilkan pada Tabel 8. Pelabelan Data.

```

positive_lexicon = set(pd.read_csv("positive.tsv", sep="\t", header=None)[0])
negative_lexicon = set(pd.read_csv("negative.tsv", sep="\t", header=None)[0])

# Fungsi untuk menentukan sentimen dan menghitung skornya
def determine_sentiment(text):
    if isinstance(text, str):
        positive_count = sum(1 for word in text.split() if word in positive_lexicon)
        negative_count = sum(1 for word in text.split() if word in negative_lexicon)
        sentiment_score = positive_count - negative_count

        # Jika skor 0, ubah menjadi positif atau negatif secara acak
        if sentiment_score == 0:
            sentiment_score = random.choice([1, -1])

        if sentiment_score > 0:
            sentiment = "Positif"
        else:
            sentiment = "Negatif"

        return sentiment_score, sentiment
    return 0, "Netral"
    
```

Gambar 8. Source Pelabelan Data

Tabel 8. Pelabelan Data

Stemming	Score	Sentiment
[curup, buluh, jalan, ekstrim, nikmat, indah, alami, kunjung, penasaran]	3	Positif

D. Ekstraksi Fitur

Tahap ekstraksi fitur dilakukan guna memanfaatkan metode *TF-IDF* untuk mengonversi teks menjadi bentuk angka, dengan bantuan *TfidfVectorizer* yang menetapkan bobot untuk tiap kata sesuai dengan tingkat kemunculannya pada dokumen. Proses ini ditunjukkan pada Gambar 9. *Source Code* Ekstraksi Fitur.

Hasil dari proses ekstraksi fitur berupa nilai bobot pada setiap kata yang merepresentasikan komentar wisatawan. Contoh hasil dari ekstraksi fitur dapat dilihat pada Tabel 9. Ekstraksi Fitur.

```

vectorizer = TfidfVectorizer()
tfidf_matrix = vectorizer.fit_transform(df['stemming_str'])
tfidf_df = pd.DataFrame(
    tfidf_matrix.toarray(),
    columns=vectorizer.get_feature_names_out()
)
top_n = 10
top_words = tfidf_df.apply(
    lambda x: x.nlargest(top_n).index.tolist(),
    axis=1
)
top_values = tfidf_df.apply(
    lambda x: x.nlargest(top_n).values.tolist(),
    axis=1
)
hasil_tfidf = pd.DataFrame({
    'Top Words': top_words,
    'Top Values': top_values
})
    
```

Gambar 9. Source Code Ekstraksi Fitur

Tabel 9. Ekstraksi Fitur

Top Words	Top Values
curup, buluh, jalan, ekstrim, nikmat,	0.33, 0.33, 0.33, 0.33, 0.33, 0.33,
indah, alami, kunjung, penasaran	0.33, 0.33, 0.33

E. Klasifikasi Support Vector Machine

Berikut tahapan terakhir yang digunakan pada proses pengklasifikasian sentimen menggunakan *Support Vector Machine (SVM)*. Data ini dipakai dalam penelitian ini hanya ada sentimen positif dan sentimen negatif tanpa sentimen netral yang diperoleh dari hasil pelabelan data sebelumnya.

Sebelum melakukan klasifikasi dataset sebelumnya dibedakan menjadi data latih (*training data*) dan data uji (*testing data*) dengan berbagai skenario pembagian data mulai dari 90:10 sampai 50:50. Pembagian data ini bertujuan agar dapat melihat kinerja algoritma *SVM* dalam mengklasifikasikan sentimen pada ulasan wisatawan.

Dari beberapa skenario pembagian data tersebut, kemudian dipilih hasil dengan nilai akurasi tertinggi untuk mengetahui seberapa baik algoritma *SVM* untuk mengelompokkan klasifikasi sentimen terhadap dataset yang dipakai pada penelitian ini.

F. Hasil Evaluasi

Pengujian ini memberi hasil nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Setiap objek wisata memiliki hasil evaluasi yang berbeda pada setiap skenario pembagian data yang digunakan. Oleh karena itu, dipilih hasil klasifikasi dengan nilai akurasi tertinggi sebagai acuan dalam menilai kinerja model. Berikut ditampilkan hasil

pengujian pada objek wisata Curup Maung serta *classification report* dengan nilai tertinggi yang diperoleh dari proses klasifikasi tersebut.

Tabel 4.10 Tabel Hasil Pengujian SVM Curup Maung

No	Pembagian Data	Hasil Akurasi	Hasil Presentasi
1.	90:10	0.9000	90.00%
2.	80:20	0.8000	80.00%
3.	70:30	0.7667	76.67%
4.	60:40	0.7250	72.50%
5.	50:50	0.7000	70.00%

	precision	recall	f1-score	support
Negatif	0.857	1.000	0.923	6.000
Positif	1.000	0.750	0.857	4.000
accuracy	0.900	0.900	0.900	0.900
macro avg	0.929	0.875	0.890	10.000
weighted avg	0.914	0.900	0.897	10.000

Gambar 4. 10 Hasil *Clasification Report* Curup Maung

Selanjutnya ditampilkan hasil pengujian pada objek wisata Curup Buluh beserta *classification report* dengan nilai akurasi tertinggi yang diperoleh dari proses klasifikasi.

Tabel 4.11 Tabel Hasil Pengujian SVM Curup Buluh

No	Pembagian Data	Hasil Akurasi	Hasil Presentasi
1.	90:10	0.9000	90.00%
2.	80:20	0.7000	70.00%
3.	70:30	0.8000	80.00%
4.	60:40	0.7500	75.00%
5.	50:50	0.8000	80.00%

	precision	recall	f1-score	support
Negatif	0.875	1.000	0.933	7.000
Positif	1.000	0.667	0.800	3.000
accuracy	0.900	0.900	0.900	0.900
macro avg	0.938	0.833	0.867	10.000
weighted avg	0.912	0.900	0.893	10.000

Gambar 4. 11 Hasil *Classification Report* Curup Buluh

Berikut disajikan hasil pengujian klasifikasi pada objek wisata Curup Senyawe beserta *classification report* dengan nilai performa terbaik yang diperoleh dari proses pengujian model.

Tabel 4.12 Tabel Hasil Pengujian SVM Curup Senyawe

No	Pembagian Data	Hasil Akurasi	Hasil Presentasi
1.	90:10	0.7000	70.00%
2.	80:20	0.7000	70.00%
3.	70:30	0.7000	70.00%
4.	60:40	0.7250	72.50%
5.	50:50	0.7000	70.00%

	precision	recall	f1-score	support
Negatif	0.730	0.964	0.831	28.000
Positif	0.667	0.167	0.267	12.000
accuracy	0.725	0.725	0.725	0.725
macro avg	0.698	0.565	0.549	40.000
weighted avg	0.711	0.725	0.662	40.000

Gambar 4.12 Hasil *Classification Report* Curup Senyawe

Selanjutnya ditampilkan hasil evaluasi klasifikasi pada objek wisata Air Terjun Panjang serta *classification report* dengan hasil akurasi tertinggi dari skenario pembagian data yang dilakukan.

Tabel 4.13 Tabel Hasil Pengujian SVM Air Terjun Panjang

No	Pembagian Data	Hasil Akurasi	Hasil Presentasi
1.	90:10	0.9000	90.00%

2.	80:20	0.8500	85.00%
3.	70:30	0.8333	83.33%
4.	60:40	0.8250	82.50%
5.	50:50	0.8200	82.00%

	precision	recall	f1-score	support
Negatif	0.889	1.000	0.941	8.000
Positif	1.000	0.500	0.667	2.000
accuracy	0.900	0.900	0.900	0.900
macro avg	0.944	0.750	0.804	10.000
weighted avg	0.911	0.900	0.886	10.000

Gambar 4.13 Hasil *Clasification Report* Air Terjun Panjang

Pada bagian ini ditunjukkan hasil pengujian klasifikasi pada objek wisata Green Canyon yang dilengkapi dengan *classification report* dengan nilai terbaik yang dihasilkan oleh model SVM.

Tabel 4.14 Tabel Hasil Pengujian SVM Gree Canyon

No	Pembagian Data	Hasil Akurasi	Hasil Presentasi
1.	90:10	0.6000	80.00%
2.	80:20	0.7000	70.00%
3.	70:30	0.7333	73.33%
4.	60:40	0.7000	70.00%
5.	50:50	0.7000	70.00%

	precision	recall	f1-score	support
Negatif	0.857	1.000	0.923	6.000
Positif	1.000	0.750	0.857	4.000
accuracy	0.900	0.900	0.900	0.900
macro avg	0.929	0.875	0.890	10.000
weighted avg	0.914	0.900	0.897	10.000

Gambar 4.13 Hasil *Clasification Report* Green Canyon

Berdasarkan hasil pengujian terhadap lima objek yang diteliti, diperoleh nilai akurasi tertinggi sebesar 90% untuk objek Curup Maung, Curup Buluh dan Air Terjun Panjang dengan *splitting data* 90:10, yang memberitahu bahwa algoritma *Support Vector Machine (SVM)* bisa mengelompokkan data sentimen dengan tingkatan ketepatan yang cukup baik. Hasil ini mengindikasikan bahwa sebagian besar data uji dapat diklasifikasikan dengan benar sesuai dengan pola yang diperoleh dari data latih, sehingga metode *SVM* dapat mengenali karakteristik umum data teks secara efektif, khususnya pada kelas sentimen yang memiliki jumlah data lebih dominan.

Meskipun demikian, hasil klasifikasi tersebut belum sepenuhnya optimal karena masih terdapat sekitar 10% data yang mengalami kesalahan klasifikasi. Kondisi ini diduga disebabkan oleh keterbatasan metode dalam memahami konteks bahasa yang kompleks, seperti penggunaan makna ganda, serta sentimen yang tidak disampaikan secara eksplisit. Selain itu, variasi penggunaan bahasa tidak baku pada data ulasan juga turut memengaruhi tingkat akurasi hasil klasifikasi.

Maka karena itu penelitian ini masih memiliki banyak peluang agar dikembangkan lebih lanjut dengan melakukan optimalisasi pada tahap prapemrosesan data, terutama pada proses normalisasi, *stemming*, dan pelabelan data. Peningkatan kualitas prapemrosesan diharapkan dapat menghasilkan representasi fitur yang lebih baik sehingga mampu meningkatkan performa metode *SVM* dalam klasifikasi sentimen.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menerapkan metode *Support Vector Machine (SVM)* untuk menganalisis sentimen ulasan pendatang terhadap objek wisata di Kabupaten Lahat berdasarkan data ulasan dari *Google Maps*. Hasil pengujian diketahui bahwa metode *SVM* berhasil mengklasifikasi sentimen dengan cukup baik, sehingga objek wisata Curup Maung, Curup Buluh, dan Air Terjun Panjang mendapatkan hasil akurasi tertinggi sebesar 90%, Green Canyon sebesar 80%, dan Curup Senyawe sebesar 72,50%. Berdasarkan hasil analisis yang dilakukan, sebagian besar ulasan pengunjung cenderung bersentimen negatif. Hal tersebut menunjukkan bahwa masih ada sejumlah aspek yang perlu diperhatikan dan ditingkatkan oleh pengelola wisata, terutama terkait fasilitas dan pelayanan kepada pengunjung. Dengan hasil akurasi yang diperoleh, metode *SVM* dinilai cukup efektif untuk digunakan dalam analisis sentimen pada data ulasan wisata berbasis teks.

Penelitian ini masih memiliki keterbatasan pada jumlah data ulasan dan sumber data yang hanya berasal dari *Google Maps*. Dengan demikian, penelitian selanjutnya disarankan untuk menambahkan jumlah data serta memanfaatkan data yang berasal dari media sosial lain seperti *Instagram*, *Facebook*, maupun *platform* lainnya agar hasil analisis menjadi lebih beragam dan akurat. Selain itu, penelitian selanjutnya juga bisa membandingkan algoritma *SVM* dengan metode klasifikasi lainnya untuk mengetahui metode yang memiliki performa terbaik dalam analisis sentimen. Penambahan kategori sentimen seperti sentimen netral juga dapat dilakukan agar hasil klasifikasi menjadi lebih detail. Hasil penelitian ini diharapkan bisa dijadikan sebagai pertimbangan dalam pengelola wisata maupun pemerintah daerah dalam upaya peningkatan mutu pelayanan dan pengembangan objek wisata di Kabupaten Lahat.

DAFTAR REFERENSI

- Algoritme, J., et al. (2025). Analisis sentimen supporter Sriwijaya FC berbasis machine learning. *Algoritme*, 6(1), 37–49. <https://doi.org/10.35957/algoritme.v6i1.11288>
- Asmara, Y., Arief, M. R., & Kusriani. (2024). Analisis sentimen berbasis aspek pada ulasan aplikasi RuangGuru menggunakan support vector machine dan naive Bayes. *Februari*, 2024(2), 209–215
- Efrizoni, L. (2022). Komparasi ekstraksi fitur pada klasifikasi teks menggunakan algoritma machine learning. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 21(3), 653–666
- Erama, R. (2025). Pemanfaatan *platform cloud Google Colab* untuk *scraping* komentar TikTok pada konten Gorontalo sebagai dasar analisis respons warganet. *Journal of Applied Engineering Science (JAES)*, 1(2), 123–134
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2021). Sentiment analysis of Ruang Guru application on Twitter using classification algorithm. *Jurnal Teknoinfo*, 14(2), 115
- Hadi, S. (2025). Analisis sentimen wisata air terjun di Kabupaten Lombok Tengah menggunakan metode support vector machine. *JOM Fakultas Teknologi Informasi*
- Ipmawati, J., Saifulloh, S., & Kusnawi, K. (2024). Analisis sentimen tempat wisata berdasarkan ulasan pada *Google Maps* menggunakan algoritma support vector machine. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 247–256
- Irwansyah, M., Suhardi, M., & Ali. (2023). Analisis sentimen untuk pengembangan pariwisata: Studi kasus pada destinasi wisata di Sumatera Barat. *Jurnal Teknologi Informasi dan Pendidikan*, 77–86
- Kanaka, S. P. N. A., Heriansyah, R., & Syahreal, K. (2024). Perbandingan algoritma decision tree dan support vector machine dalam pemilihan calon mahasiswa penerima KIP-K. *Terapan Informatika Nusantara*, 613–619

- Khalim, K. A., Hayati, U., & Bahtiar, A. (2023). Perbandingan prediksi penyakit hipertensi menggunakan metode random forest dan naïve bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 498–504
- Kristian, Y. H., Prilianti, K. R., & Tirma, P. L. (2021). Implementasi text mining untuk analisis tempat wisata di Indonesia. *Jurnal SimanteC*, 73–82
- Pratama, R. D., & Nugroho. (2021). Penerapan algoritma support vector machine untuk klasifikasi opini masyarakat pada media sosial. *Jurnal Teknologi Informasi dan Komunikasi*
- Herdiansyah, Nazori Suhandi, & Dwi Asa Verano. (2026). Analisis sentimen masyarakat terhadap polemik ijazah Joko Widodo menggunakan natural language processing dengan algoritma multi layer perceptron
- Rahman Hakim, A. (2024). Analisis sentimen objek wisata di *Google Maps* menggunakan metode decision tree. *Computer Based Information System Journal*, 12(1), 122–130. <http://ejournal.upbatam.ac.id/index.php/cbis>
- Ramadhan, D. Y., & Putra, A. P. (2021). Analisis sentimen komentar pengguna Instagram terhadap pariwisata Indonesia menggunakan support vector machine (SVM). *Jurnal Teknologi Informasi dan Komputer*
- Saputra, R. A. (2024). Analisis sentimen review skincare Skintific dengan algoritma support vector machine. 12(2)
- Suhel, S., & Bashir, A. (2021). The role of tourism toward economic growth in the local economy. *Economic Journal of Emerging Markets*, 10(1), 32–39. <https://doi.org/10.20885/ejem.vol10.iss1.art4>
- Syamsuar, D., Negara, E. S., Andryani, R., Triadi, D., Kusmeta, Y., & Sintia, M. (2022). Sistem informasi mobile untuk objek wisata. 1, 35–40
- Widodo, A., & Hartati, S. (2022). Penerapan analisis sentimen pada ulasan pelanggan untuk evaluasi kualitas layanan wisata menggunakan metode text mining. *Jurnal Informatika*, 55–64
- Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021). Sentiment analysis of independent campus policy using naive Bayes and TF-IDF weighting based on comments on YouTube. *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi*, 2(1), 55–63