

ANALISIS SENTIMEN ULASAN EKOWISATA LANGKAT MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)

Rima Asari Br Sitepu¹, Khairul², Kristina Annatasia³
Sistem Informasi, STMIK Kaputama, Binjai
E-mail: *rimaasari04@gmail.com¹

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen pengunjung terhadap ekowisata di Kabupaten Langkat melalui ulasan kuesioner yang diklasifikasikan menggunakan metode Support Vector Machine (SVM). Data diperoleh dari 505 responden dan telah melalui tahap preprocessing, termasuk tokenisasi, stopwords removal, dan stemming. Hasil klasifikasi dibagi menjadi tiga kategori sentimen: positif, netral, dan negatif. Model dilatih menggunakan pendekatan TF-IDF dan penyeimbangan data dengan SMOTE. Hasil pengujian menunjukkan akurasi sebesar 88,12% dengan performa terbaik pada sentimen positif. Temuan ini menunjukkan bahwa sebagian besar pengunjung memiliki pandangan yang positif terhadap pengalaman berwisata di Langkat, terutama di lokasi seperti Bukit Lawang dan Tangkahan. Meskipun demikian, diperlukan perbaikan dalam mendeteksi sentimen netral dan negatif guna meningkatkan keakuratan sistem secara keseluruhan.

Kata kunci

Analisis Sentimen, Ekowisata, Langkat, Support Vector Machine, Kuesioner

ABSTRACT

This study aims to analyze visitors' sentiments toward ecotourism in Langkat Regency through questionnaire-based reviews classified using the Support Vector Machine (SVM) method. The dataset, consisting of 505 respondents, underwent preprocessing stages such as tokenization, stopwords removal, and stemming. Sentiment classification was divided into three categories: positive, neutral, and negative. The model was trained using a TF-IDF approach and data balancing with SMOTE. Evaluation results show an accuracy of 88.12%, with the best performance observed in the positive sentiment category. These findings indicate that most visitors have a favorable perception of their tourism experiences in Langkat, particularly at destinations like Bukit Lawang and Tangkahan. However, improvements are needed in detecting neutral and negative sentiments to enhance overall classification accuracy.

Keywords

Sentiment Analysis, Ecotourism, Langkat, Support Vector Machine, Questionnaire

1. PENDAHULUAN

Ekowisata merupakan salah satu sektor pariwisata yang terus berkembang di Indonesia, termasuk di Kabupaten Langkat, Sumatera Utara, yang memiliki keanekaragaman flora, fauna, dan keindahan alam yang menawan (Nazwin and Hidayat, 2022). Lebih lanjut, konsep ekowisata juga dianggap sebagai penopang ekonomi suatu daerah, yang berasal dari transformasi pariwisata massal (Saputro and Wardani, 2025). Menurut (Angela, 2023), ekowisata juga berperan penting dalam mendukung konservasi alam. Pembangunan ekowisata berbasis masyarakat lokal juga diyakini dapat menjadi strategi yang efektif (Asep Jamaludin et al., 2024).

Keberhasilan ekowisata tidak hanya ditentukan oleh keindahan alam, melainkan juga persepsi dan pengalaman pengunjung (Adnyana, 2015). Ulasan yang diberikan pengunjung, baik positif maupun negatif, dapat memengaruhi citra destinasi dan minat wisatawan lain (Tondang et al., 2023). Hal ini sejalan dengan temuan Tinaliah & Elizabeth (2022) yang menunjukkan pentingnya analisis ulasan dalam evaluasi layanan. Oleh

karena itu, analisis sentimen diperlukan untuk memahami persepsi wisatawan secara lebih sistematis (Sibyan and Hasanah, 2022). Analisis sentimen merupakan bidang ilmu yang berfokus pada pemahaman opini, emosi, dan sikap seseorang terhadap suatu topik (Manullang et al., 2023). Penerapan metode Naive Bayes juga telah terbukti efektif dalam klasifikasi ulasan di sektor e-commerce (Rahmawati and Santoso, 2023).

Dalam ranah analisis data, berbagai algoritma dapat digunakan. (Agustin et al., 2024) menggunakan Naïve Bayes untuk analisis sentimen, sementara (Cahyaningtyas et al., 2021) menerapkan Decision Tree. Selain itu, algoritma Random Forest juga sering digunakan dalam klasifikasi data (Mutoffar and Fadillah, 2022). Di antara algoritma tersebut, SVM menjadi pilihan yang kuat (Wiratama Putra et al., 2022) SVM dikenal unggul dalam klasifikasi teks dengan dimensi tinggi (Ibrahim, 2025) sebuah keunggulan yang juga ditegaskan oleh (Imran et al., 2024) dan (Pane et al., 2021). Beberapa penelitian sebelumnya juga membuktikan efektivitas SVM, misalnya pada analisis sentimen aplikasi JMO (Manullang and Prianto, n.d.) tempat wisata Yogyakarta (Ipmawati et al., 2024), dan aplikasi Wetv (Abdillah et al., 2023). Berdasarkan latar belakang tersebut, penelitian ini dilakukan untuk menganalisis ulasan pengunjung ekowisata di Kabupaten Langkat menggunakan algoritma SVM.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan algoritma Support Vector Machine (SVM). Alur penelitian ini mengacu pada tahapan dalam metode *Knowledge Discovery in Databases* (KDD), sebuah proses sistematis untuk menggali pengetahuan dari data (Ramiro Firjatullah et al., 2025). Tahapan KDD ini meliputi seleksi, *preprocessing*, transformasi, *data mining*, dan evaluasi, yang juga sejalan dengan metodologi yang diusulkan oleh (Subroto et al., 2022).

a. Pengumpulan Data (Data Selection)

Data diperoleh melalui penyebaran kuesioner daring (Google Form) kepada 505 responden pengunjung ekowisata Langkat (Romdona et al., 2025). Kuesioner ini dirancang dengan pertanyaan terbuka untuk mengumpulkan ulasan tekstual mengenai pengalaman mereka. Total data ulasan yang terkumpul adalah 3.528, di mana setiap responden memberikan 7 ulasan. Penggunaan kuesioner online merupakan metode yang efisien untuk mengumpulkan data dalam jumlah besar, seperti yang telah ditunjukkan oleh (Rizka Maya Shova and Hwihanus Hwihanus, 2022) dalam penelitian mereka.

b. Pra-pemrosesan Data (Preprocessing)

Data ulasan tekstual yang telah terkumpul melalui serangkaian tahapan pembersihan, yaitu:

- 1) *Case Folding*: Mengubah seluruh teks menjadi huruf kecil (*lowercase*).
- 2) *Cleaning*: Menghapus karakter khusus seperti emoji, angka, tanda baca, dan spasi ganda.
- 3) *Tokenization*: Memecah kalimat menjadi kata-kata individu (*token*).
- 4) *Stopword Removal*: Menghilangkan kata-kata umum yang tidak relevan dengan sentimen.
- 5) *Stemming*: Mengubah kata menjadi bentuk dasarnya (akar kata) menggunakan algoritma Sastrawi untuk Bahasa Indonesia.

c. Transformasi Data (Transformation)

Setelah data teks bersih, ekstraksi fitur dilakukan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini memberikan bobot

numerik pada setiap kata untuk menunjukkan kepentingannya dalam dokumen relatif terhadap seluruh korpus (Hafizh Mahendra et al., 2023). Rumus yang digunakan adalah sebagai berikut:

1) TF (*Term Frequency*)

$$TF(t,d) \dots\dots\dots(1)$$

(Jumlah kemunculan kata t dalam dokumen d) / (Total kata dalam dokumen d)

Alternatif (*log Normalization*):

$$TF(t,d) = 1 + \log(freq(t,d))$$

jika $freq > 0$

$$TF(t,d) = 0$$

jika $freq = 0$

2) IDF (*Inverse Document Frequency*)

$$IDF(t) = \log(N / DF(t)) \dots\dots\dots(2)$$

Dimana:

N = Total jumlah dokumen

DF(t) = Document Frequency = Jumlah dokumen yang mengandung kata t

3) TF-IDF

$$TF-IDF(t,d) = TF(t,d) \times IDF(t) \dots\dots\dots(3)$$

d. Klasifikasi (*Data Mining*)

Data yang telah diubah menjadi vektor numerik kemudian dibagi menjadi data latih (70%) dan data uji (30%). Algoritma *Support Vector Machine* (SVM) digunakan untuk membangun model klasifikasi sentimen dengan tiga kelas: positif, netral, dan negatif. Model dilatih menggunakan data latih dan kemudian diuji menggunakan data uji untuk mengukur performanya. SVM adalah algoritma *machine learning* yang mencari *hyperplane* terbaik untuk memisahkan data dalam kelas berbeda berikut rumus dasar *hyperplane*:

$$f(x) = w \cdot x + b \dots\dots\dots(1)$$

Dimana :

$f(x)$: Fungsi keputusan yang menentukan kelas suatu data.

w : Vektor bobot yang mempengaruhi arah *hyperplane*.

x : Vektor fitur input.

b : Bias, yang menentukan posisi *hyperplane*. Untuk mendapatkan margin maksimal, SVM menggunakan rumus:

$$M = \frac{2}{\|w\|} \dots\dots\dots(2)$$

SVM juga menggunakan fungsi kernel untuk menangani data non-linear, dengan beberapa jenis kernel seperti:

$$\text{Linear Kernel: } K(x_i, x_j) = x_i \cdot x_j \dots\dots\dots(3)$$

$$\text{Polynomial Kernel: } K(x_i, x_j) = (x_i \cdot x_j + C)^d \dots\dots\dots(4)$$

$$\text{Radial Basis Function (RBF) Kernel: } K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \dots\dots\dots(5) \text{ (Aji et al., 2025).}$$

e. Evaluasi (*Evaluation*)

Performa model diukur menggunakan metrik akurasi, precision, recall, dan F1-score (Astari et al., 2020). Hasil evaluasi ini digunakan untuk menilai seberapa efektif model SVM dalam mengklasifikasikan sentimen ulasan pengunjung. Rumus dari evaluasi *Precision*, *Recall*, *F1-score* sebagai berikut:

1) *Precision*

$$\begin{array}{ll} \text{Positif} & Precision = \frac{TP}{TP+FP} \\ \text{Negatif} & Precision = \frac{TN}{TN+FN} \end{array}$$

Netral	$Precision = \frac{TN_{Netral}}{TN_{Netral} + FN_{Netral}}$
2) Recall	
Positif	$Recall = \frac{TP}{TP + FN}$
Negatif	$Recall = \frac{TN}{TN + FN} =$
Netral	$Recall = \frac{TN_{Netral}}{TN_{Netral} + FN_{Netral}}$
3) F1-Score	
Positif	$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$
Negatif	$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$
Netral	$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$

3. HASIL DAN PEMBAHASAN

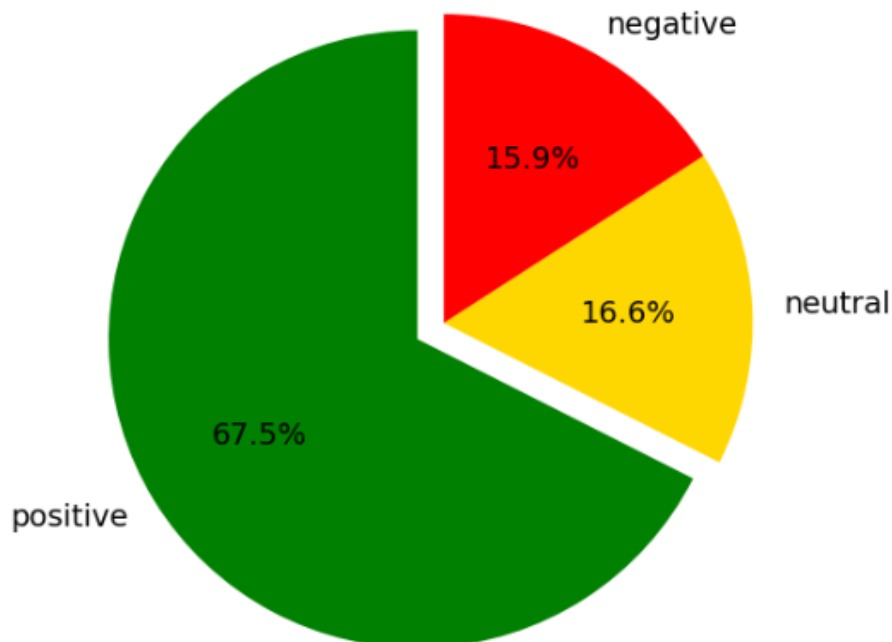
Berdasarkan analisis sentimen terhadap ulasan pengunjung ekowisata Langkat, diperoleh hasil sebagai berikut:

- a. Akurasi Model: Model klasifikasi SVM dengan *linear kernel* menunjukkan akurasi sebesar 88,12%.

Akurasi: 0.88
Akurasi: 88.12%

Gambar 1. Akurasi

- b. Distribusi Sentimen: Mayoritas ulasan pengunjung (67,5%) tergolong sentimen positif, diikuti oleh sentimen netral (16,6%) dan negatif (15,9%).



Gambar 2. Pie Chart

d. Faktor-faktor yang Mempengaruhi Sentimen

Ulasan dengan sentimen negatif umumnya berkaitan dengan masalah fasilitas umum dan infrastruktur. Kata-kata seperti “toilet”, “sampah”, “jalan”, dan “pungli” sering muncul dalam ulasan negatif, hal ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa faktor-faktor seperti kondisi jalan dan fasilitas parkir sering menjadi keluhan utama (Ipmawati et al., 2024).

4. KESIMPULAN

Penelitian ini menunjukkan bahwa metode *Support Vector Machine* (SVM) efektif dalam melakukan analisis sentimen terhadap ulasan pengunjung ekowisata Langkat, dengan tingkat akurasi model mencapai 88,12%. Mayoritas sentimen pengunjung adalah positif, yang mencerminkan kesan baik mereka terhadap destinasi ekowisata di Langkat. Namun, masih terdapat tantangan dalam mengklasifikasikan sentimen netral dan negatif, yang memerlukan perbaikan model di masa mendatang. Saran perbaikan yang sering muncul, seperti masalah kebersihan, fasilitas, dan aksesibilitas, dapat menjadi masukan berharga bagi pihak pengelola untuk meningkatkan kualitas layanan dan mendukung keberlanjutan sektor pariwisata di Kabupaten Langkat.

5. DAFTAR PUSTAKA

- Abdillah, R., Haerani, E., Candra, R.M., 2023. Analisis Sentimen Ulasan Aplikasi Wetv Untuk Peningkatan Layanan Menggunakan Metode Support Vector Machine. J. Inf. Syst. Res. 4. <https://doi.org/10.47065/josh.v4i3.3353>
- Adnyana, P.B., 2015. Analisis Potensi dan Kondisi Ekosistem Terumbu Karang Pulau Menjangan untuk Pengembangan Ekowisata Bahari Berbasis Pendidikan Terpadu. JST (Jurnal Sains dan Teknol. 3. <https://doi.org/10.23887/jst-undiksha.v3i2.4474>
- Agustin, A.V., Sa'adah, F.N., Umaidah, Y., 2024. Analisis Sentimen Menggunakan Metode Naive Bayes Terhadap Childfree. Dinamik 29. <https://doi.org/10.35315/dinamik.v29i1.9455>
- Aji, M., Maldini, A., Andryana, S., 2025. ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI PERBANKAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAIVE BAYES, Jurnal Mahasiswa Teknik Informatika).
- Angela, vivi F., 2023. Strategi Pengembangan Ekowisata dalam Mendukung Konservasi Alam Danau Tahai. JIM J. Ilm. Mhs. Pendidik. Sej. 8.
- Asep Jamaludin, Sihabudin, Asep Darojatul Romli, Yudi Firmansyah, Wanta, 2024. STRATEGI CAPACITY BUILDING PEMERINTAH DESA DALAM PENGEMBANGAN POTENSI KAMPOENG EKOWISATA BERBASIS MASYARAKAT LOKAL. J. BUANA Pengabdian. 6. <https://doi.org/10.36805/jurnalbuanapengabdian.v6i1.6221>
- Astari, N.M.A.J., Dewa Gede Hendra Divayana, Gede Indrawan, 2020. Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. J. Sist. dan Inform. 15. <https://doi.org/10.30864/jsi.v15i1.332>
- Cahyaningtyas, C., Nataliani, Y., Widiarsari, I.R., 2021. Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE. AITI 18. <https://doi.org/10.24246/aiti.v18i2.173-184>
- Hafizh Mahendra, M., Triantoro Murdiansyah, D., Muslim Lhaksmana, K., 2023. Dike : Jurnal Ilmu Multidisiplin Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer.
- Ibrahim, M.I.K., 2025. Analisis Metode Support Vector Machine (SVM) dan Random Forest

- Terhadap Perilaku Dan Prestasi Siswa Berbasis Kurikulum Merdeka Di SMK Negeri 1 Stabat. *J. Ris. Sist. Inf. Dan Tek. Inform.* 10, 453–467.
- Imran, B., Karim, M.N., Ningsih, N.I., 2024. KLASIFIKASI BERITA HOAX TERKAIT PEMILIHAN UMUM PRESIDEN REPUBLIK INDONESIA TAHUN 2024 MENGGUNAKAN NAÏVE BAYES DAN SVM. *Din. Rekayasa* 20. <https://doi.org/10.20884/1.dinarek.2024.20.1.27>
- Ipmawati, J., Saifulloh, S., Kusnawi, K., 2024. Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine. *MALCOM Indones. J. Mach. Learn. Comput. Sci.* 4, 247–256. <https://doi.org/10.57152/malcom.v4i1.1066>
- Manullang, O., Prianto, C., n.d. Terbit online pada laman web jurnal: <https://ejurnalunsam.id/index.php/jicom/> Analisis Sentimen dalam Memprediksi Hasil Pemilu Presiden dan Wakil Presiden : Systematic Literature Review.
- Manullang, O., Prianto, C., Harani, N.H., 2023. ANALISIS SENTIMEN UNTUK MEMPREDIKSI HASIL CALON PEMILU PRESIDEN MENGGUNAKAN LEXICON BASED DAN RANDOM FOREST. *J. Ilm. Inform.* 11. <https://doi.org/10.33884/jif.v11i02.7987>
- Mutofar, M.M., Fadillah, A., 2022. KLASIFIKASI KUALITAS AIR SUMUR MENGGUNAKAN ALGORITMA RANDOM FOREST. *Naratif J. Nas. Riset, Apl. dan Tek. Inform.* 4. <https://doi.org/10.53580/naratif.v4i2.160>
- Nazwin, A.H., Hidayat, R., 2022. Evaluasi Pengelolaan Ekowisata: A Systematic Literature Review. *Kolaborasi J. Adm. Publik* 8. <https://doi.org/10.26618/kjap.v8i3.9252>
- Pane, S.F., Owen, A., Prianto, C., 2021. Analisis Sentimen UU Omnibus Law pada Twitter Menggunakan Metode Support Vector Machine. *J. Telekomun. dan Komput.* 11. <https://doi.org/10.22441/incomtech.v11i2.10874>
- Rahmawati, L., Santoso, D.B., 2023. Implementasi Metode Naive Bayes Untuk Klasifikasi Ulasan Aplikasi E-Commerce Tokopedia. *INTECOMS J. Inf. Technol. Comput. Sci.* 6. <https://doi.org/10.31539/intecom.v6i1.5515>
- Ramiro Firjatullah, F., Faqih, A., Dwilestari, G., 2025. OPTIMALISASI STRATEGI PEMASARAN MENGGUNAKAN ALGORITMA AGGLOMERATIVE HIERARCHICAL CLUSTERING PADA HASIL PENJUALAN, *Jurnal Mahasiswa Teknik Informatika*).
- Rizka Maya Shova, Hwihanus Hwihanus, 2022. Peranan Sistem Informasi Akuntansi Dalam Pengolahan Harga dan Kualitas Produk Terhadap Minat Beli Konsumen Pada UMKM Ikan Asap. *Wawasan J. Ilmu Manajemen, Ekon. dan Kewirausahaan* 1. <https://doi.org/10.58192/wawasan.v1i1.379>
- Romdona, S., Senja Junista, S., Gunawan, A., 2025. TEKNIK PENGUMPULAN DATA: OBSERVASI, WAWANCARA DAN KUESIONER 3, 39–47.
- Saputro, L.E., Wardani, D., 2025. Pengembangan Ekowisata Bantaran Sungai Gajah Wong Kapanewon Nologaten Kabupaten Sleman Yogyakarta. *Gemawisata J. Ilm. Pariwisata* 21, 328–350. <https://doi.org/10.56910/gemawisata.v21i1.584>
- Sibyan, H., Hasanah, N., 2022. ANALISIS SENTIMEN ULASAN PADA WISATA DIENG DENGAN ALGORITMA K-NEAREST NEIGHBOR (K-NN). *J. Penelit. dan Pengabd. Kpd. Masy. UNSIQ* 9. <https://doi.org/10.32699/ppkm.v9i1.2218>
- Subroto, G., Sulistiyowati, N., Ridha, A.A., 2022. Klasifikasi Jenis Kekerasan Pada Perempuan Dan Anak Dengan Algoritma Multinomial Naïve Bayes. *INTECOMS J. Inf. Technol. Comput. Sci.* 5. <https://doi.org/10.31539/intecom.v5i1.3598>
- Tondang, B.A., Muhammad Rizqan Fadhil, Muhammad Nugraha Perdana, Akhmad Fauzi, Ugra Syahda Janitra, 2023. Analisis pemodelan topik ulasan aplikasi BNI, BCA, dan BRI menggunakan latent dirichlet allocation. *INFOTECH J. Inform. Teknol.* 4. <https://doi.org/10.37373/infotech.v4i1.601>

Wiratama Putra, T., Triayudi, A., Andrianingsih, A., 2022. Analisis Sentimen Pembelajaran Daring Menggunakan Metode Naïve Bayes, KNN, dan Decision Tree. J. JTIK (Jurnal Teknol. Inf. dan Komunikasi) 6, 20–26. <https://doi.org/10.35870/jtik.v6i1.368>