

Eksplorasi Algoritma *Support Vector Machine* untuk Analisis Sentimen Destinasi Wisata di Indonesia

Junaedi^{1*}, Alexius Hendra Gunawan²⁾, Verri Kuswanto³⁾, Jonathan⁴⁾

¹⁾²⁾³⁾⁴⁾Universitas Buddhi Dharma

Jl. Imam Bonjol No.41, Tangerang, Indonesia

¹⁾junaedi@ubd.ac.id

²⁾alexius.hendra@ubd.ac.id

³⁾verri.kuswanto@ubd.ac.id

⁴⁾jonathan@ubd.ac.id

Article history:

Received 15 Nov 2024;
Revised 19 Nov 2024;
Accepted 25 Nov 2024;
Available online 27 Des 2024

Keywords:

Analisis Sentimen
Sektor Pariwisata
Support Vector Machine
Text Mining
Twitter/X

Abstrak

Penelitian ini mengeksplorasi penerapan algoritma *Support Vector Machine* (SVM) dalam *Text Mining* untuk analisis sentimen sektor pariwisata di Indonesia, menggunakan data dari platform Twitter. Data dikumpulkan melalui API Twitter dan diproses melalui tahapan prapemrosesan teks, termasuk tokenisasi, normalisasi, penghapusan stopword, dan stemming, untuk memastikan kesiapan data dalam analisis. Model SVM diuji dengan tiga kernel berbeda—linear, radial basis function (RBF), dan sigmoid—serta menggunakan rasio data latih-uji 7:3 dan 8:2. Hasil menunjukkan bahwa kernel linear dengan rasio 7:3 menghasilkan kinerja terbaik dengan akurasi 92,89%, precision 92%, recall 74%, dan F1-score 81%. Evaluasi berdasarkan kelas sentimen menunjukkan performa tinggi pada sentimen positif (F1-score 96%) tetapi moderat pada kelas netral (F1-score 67%), mencerminkan pengaruh ketidakseimbangan data. Penelitian ini memberikan kontribusi signifikan dalam mendukung pengambilan keputusan berbasis data untuk pengembangan sektor pariwisata. Temuan ini memungkinkan pengelola destinasi wisata untuk memahami opini wisatawan secara otomatis, menyusun strategi promosi yang lebih efektif, serta meningkatkan kualitas layanan. Dengan menerapkan analisis sentimen berbasis SVM, penelitian ini mendukung pengelolaan pariwisata berbasis data untuk meningkatkan daya saing destinasi wisata di Indonesia. Penelitian lanjutan disarankan untuk mengatasi ketidakseimbangan data melalui teknik resampling atau penerapan algoritma alternatif seperti deep learning, guna meningkatkan akurasi klasifikasi sentimen yang lebih kompleks. Dengan demikian, penelitian ini menjadi langkah strategis dalam memanfaatkan teknologi analitik untuk pengelolaan pariwisata yang lebih inovatif.

I. PENDAHULUAN

Industri pariwisata adalah salah satu sektor utama yang menopang perekonomian Indonesia, memberikan kontribusi besar terhadap Produk Domestik Bruto (PDB) sekaligus menciptakan peluang signifikan untuk pertumbuhan ekonomi, peningkatan lapangan kerja, serta pelestarian budaya dan pengembangan sosial. Walaupun sektor ini mengalami pertumbuhan yang cepat, tantangan tetap muncul, terutama dalam hal pengelolaan dan respons terhadap tanggapan wisatawan secara efektif. Kondisi ini semakin diperparah oleh dampak pandemi COVID-19, yang mengakibatkan penurunan drastis aktivitas pariwisata secara global, termasuk di Indonesia. Pada tahun 2020, kontribusi sektor pariwisata terhadap PDB Indonesia turun dari 4,7% menjadi 4,05%, menekankan kebutuhan mendesak untuk menemukan cara baru guna memulihkan dan meningkatkan sektor ini melalui strategi promosi yang lebih efektif dan berfokus pada pengalaman wisatawan [1].

Media sosial telah menjadi sarana penting bagi wisatawan untuk menyampaikan pengalaman dan pandangan mereka tentang berbagai destinasi wisata. Platform seperti Twitter, dengan lebih dari 18 juta pengguna di Indonesia, menyediakan ruang bagi pengguna untuk memberikan ulasan secara real-time, menyampaikan opini,

* Corresponding author

serta membentuk persepsi publik secara luas. Informasi yang dibagikan, termasuk ulasan positif maupun negatif, memiliki dampak signifikan terhadap keputusan calon wisatawan lain dan berfungsi sebagai masukan berharga bagi pengelola destinasi wisata [2]. Namun, mengelola secara manual ribuan hingga jutaan ulasan di media sosial tidak hanya kurang efisien tetapi juga rentan terhadap bias, sehingga pendekatan analisis otomatis menjadi kebutuhan yang mendesak.

Salah satu solusi untuk mengatasi tantangan ini adalah dengan menggunakan analisis sentimen, yaitu metode analisis data yang bertujuan untuk menilai dan memahami perasaan atau opini seseorang yang diekspresikan dalam teks, sering kali didukung oleh teknologi pembelajaran mesin (*Machine Learning*). Melalui analisis sentimen, pengelola destinasi dapat secara otomatis mengenali opini pengguna, apakah berupa sentimen positif, netral, atau negatif. Metode ini memberikan peluang bagi pemangku kepentingan di sektor pariwisata untuk mengidentifikasi pola sentimen pengunjung dan memanfaatkannya sebagai dasar dalam meningkatkan kualitas layanan, memperbaiki infrastruktur, serta merancang strategi pemasaran yang lebih efektif. [3].

Algoritma *Support Vector Machine* (SVM) dikenal sebagai salah satu metode yang efektif dalam klasifikasi teks, termasuk analisis sentimen. SVM bekerja dengan mencari *hyperplane* optimal yang mampu memisahkan data berdasarkan kelasnya, seperti sentimen positif, negatif, atau netral. Keunggulan SVM terletak pada kemampuannya untuk menangani data yang tidak terstruktur, seperti teks ulasan, dan memberikan akurasi tinggi dalam klasifikasi yang kompleks [4]. Dalam hal ini, analisis sentimen berbasis *machine learning* muncul sebagai solusi untuk memahami opini wisatawan secara otomatis. Teknik *Support Vector Machine* (SVM), yang dikenal unggul dalam klasifikasi teks, menawarkan potensi besar dalam mengidentifikasi sentimen positif, negatif, atau netral dari data media sosial. Kombinasi SVM dengan teknik *Text Mining* memungkinkan pengelola destinasi untuk menganalisis opini dengan lebih akurat, memberikan dasar untuk meningkatkan layanan, memperbaiki infrastruktur, dan menyusun strategi pemasaran yang efektif.

Penggunaan *Text Mining* dan SVM untuk menganalisis data ulasan di media sosial memiliki potensi besar untuk memberikan wawasan yang cepat dan akurat kepada pengelola destinasi wisata. Data yang dikumpulkan dari Twitter dianalisis menggunakan metode SVM dengan bantuan beberapa tahapan prapemrosesan teks, seperti tokenisasi, normalisasi, penghapusan kata tidak penting (*stopword removal*), dan *stemming*. Melalui tahapan ini, teks ulasan wisatawan diubah menjadi bentuk yang terstruktur dan siap untuk dianalisis, sehingga model SVM dapat mengklasifikasikan sentimen secara lebih akurat. Penelitian ini juga melakukan pelatihan model SVM menggunakan beberapa jenis kernel, termasuk linear, radial basis *function* (RBF), dan *sigmoid*, untuk memastikan kinerja optimal dalam klasifikasi data sentimen yang beragam.

Penelitian ini bertujuan untuk mengaplikasikan algoritma SVM dalam analisis sentimen wisatawan terhadap destinasi di Indonesia dengan data Twitter sebagai sumber utama. Fokus penelitian diarahkan pada evaluasi performa kernel SVM-linear, RBF, dan sigmoid dengan rasio data latih-uji yang berbeda. Temuan ini diharapkan dapat memberikan kontribusi signifikan dalam mendukung pengambilan keputusan berbasis data di sektor pariwisata Indonesia. Selain itu, hasil dari penelitian ini diharapkan dapat memberikan kontribusi penting dalam pengembangan sektor pariwisata di Indonesia, terutama dalam hal meningkatkan kualitas layanan dan kepuasan pengunjung melalui wawasan yang lebih terarah dan berbasis data. Melalui analisis sentimen otomatis, pengelola destinasi wisata dapat merespons secara proaktif terhadap kebutuhan dan preferensi wisatawan, meningkatkan promosi, dan menarik lebih banyak kunjungan. Dengan demikian, penerapan metode analisis sentimen berbasis SVM ini diharapkan dapat mendukung pengambilan keputusan yang lebih terinformasi dalam industri pariwisata Indonesia.

II. TINJAUAN PUSTAKA

Analisis sentimen adalah metode untuk mengidentifikasi dan mengekstrak informasi subjektif dalam teks, yang sering digunakan untuk menentukan sikap atau emosi penulis terhadap suatu topik. Di sektor pariwisata, analisis sentimen membantu dalam memahami pengalaman dan pandangan wisatawan terhadap suatu destinasi. Menurut Purnamasari [5], analisis sentimen berbasis teks memungkinkan pemetaan sentimen positif, negatif, dan netral dari ulasan wisatawan, memberikan informasi berharga bagi pengelola destinasi dalam meningkatkan layanan. Metode ini biasanya menggunakan teknik *Natural Language Processing* (NLP) dan *Machine learning* (ML) untuk mengevaluasi data teks secara otomatis dan akurat. Dalam penelitian ini, Twitter dipilih sebagai sumber data karena *platform* ini sering digunakan wisatawan untuk berbagi pengalaman, ulasan, dan tanggapan terkait destinasi wisata.

Platform media sosial, terutama Twitter, memiliki karakteristik unik dalam penyebaran opini publik. Twitter menyediakan data *real-time* yang kaya dan bersifat dinamis, memungkinkan pengumpulan opini langsung dari wisatawan di berbagai lokasi. Data yang diperoleh dari Twitter, yang mencakup ulasan singkat dan langsung, menjadi sumber data utama untuk analisis sentimen di sektor pariwisata. Rizaty [2] mencatat bahwa di Indonesia terdapat sekitar 18,45 juta pengguna aktif Twitter, menjadikannya *platform* penting untuk memahami opini publik terhadap destinasi wisata. Dengan banyaknya pengguna, *platform* ini mampu mencerminkan tren dan persepsi

secara luas, serta memberikan peluang bagi para peneliti untuk mengeksplorasi opini publik secara lebih mendalam melalui *Text Mining*.

Text Mining adalah proses pengumpulan informasi dari data teks untuk menemukan pola yang bermanfaat, seperti perasaan atau opini [6]. Teknik ini sering diterapkan pada dokumen teks besar untuk mengidentifikasi kata kunci dan mengklasifikasikan konten berdasarkan tema atau perasaan tertentu. Dalam konteks analisis sentimen di pariwisata, *Text Mining* memungkinkan ekstraksi data terkait opini wisatawan, membantu pengelola destinasi memahami tanggapan yang sering muncul dan menentukan strategi promosi yang lebih efektif. Mukhlis [7] menjelaskan bahwa *Text Mining* memiliki beberapa tahapan penting, termasuk prapemrosesan teks yang melibatkan tokenisasi, penghapusan *stopword*, dan *stemming*. Tahapan ini membantu merapikan data teks yang sering kali tidak terstruktur sehingga lebih siap untuk dianalisis dengan algoritma *machine learning*.

Support Vector Machine (SVM) adalah salah satu metode klasifikasi yang efektif dalam analisis sentimen yang berlandaskan pada teori pembelajaran statistik struktural serta dikenal memiliki performa lebih unggul dibandingkan metode *machine learning* lainnya [8], [9]. Algoritma ini pertama kali dikembangkan oleh Cortes dan Vapnik, dengan tujuan untuk menemukan *hyperplane* optimal yang memisahkan data dalam dua atau lebih kelas [4], [7]. SVM bekerja dengan mencari garis atau bidang (*hyperplane*) yang membagi data ke dalam kelas-kelas secara maksimal, sehingga meningkatkan akurasi klasifikasi. Di bidang analisis sentimen, SVM telah terbukti memberikan hasil yang cukup baik dalam memisahkan sentimen positif, negatif, dan netral, membuatnya menjadi pilihan yang populer dalam *Text Mining*. Alcaraz et al [10] menunjukkan bahwa SVM unggul dalam menangani data teks yang tidak terstruktur, karena dapat menyesuaikan dengan baik pada pola-pola dalam data teks.

Beberapa penelitian terdahulu telah menggunakan SVM dalam analisis sentimen di sektor pariwisata. Penelitian oleh Tuhuteru et al [11] menunjukkan bahwa model SVM efektif untuk mengklasifikasikan sentimen wisatawan terkait destinasi wisata di Provinsi Maluku, di mana SVM dibandingkan dengan algoritma lain seperti *Naïve Bayes* dan *Random Forest*. Penelitian ini menemukan bahwa *Random Forest* memiliki akurasi lebih tinggi, tetapi SVM tetap memberikan hasil yang kompetitif dalam analisis sentimen. Selain itu, Era et al [12] membandingkan kinerja algoritma *Naïve Bayes* dan *K-Nearest Neighbor* (KNN) dalam analisis sentimen terhadap kebijakan pembukaan destinasi wisata selama pandemi COVID-19. Hasil penelitian ini menunjukkan bahwa *Naïve Bayes* memiliki akurasi yang lebih baik, namun mereka juga menekankan pentingnya mempertimbangkan SVM dalam klasifikasi sentimen yang kompleks dan data teks yang besar.

Selain itu, *text preprocessing* atau prapemrosesan teks adalah tahap penting sebelum analisis sentimen. Prapemrosesan teks melibatkan beberapa langkah, seperti tokenisasi, normalisasi, dan penghapusan kata tidak berarti (*stopword removal*). Langkah ini bertujuan untuk mempersiapkan data mentah sehingga lebih mudah dianalisis oleh algoritma seperti SVM. Amardita et al [13] menjelaskan bahwa dengan menerapkan prapemrosesan teks, kualitas data dapat ditingkatkan dan akurasi model SVM dalam klasifikasi sentimen dapat ditingkatkan. Rifaldi et al [14] juga menyebutkan bahwa prapemrosesan teks yang baik sangat penting untuk meningkatkan kinerja algoritma *machine learning* dalam analisis sentimen.

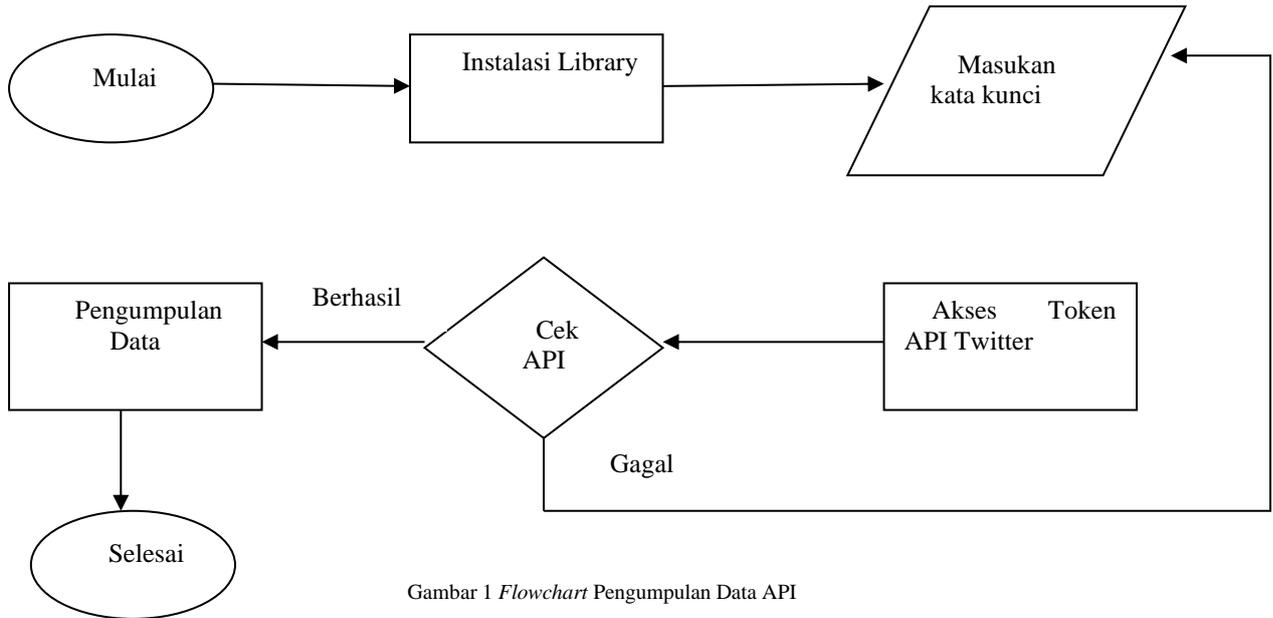
Secara keseluruhan, analisis sentimen menggunakan SVM dalam *Text Mining* menawarkan pendekatan yang menjanjikan untuk memahami persepsi publik terhadap destinasi wisata. Dengan menggunakan data dari media sosial dan mengimplementasikan SVM, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam mendukung pengambilan keputusan yang berbasis data di sektor pariwisata. Penerapan SVM dalam *Text Mining* pada data Twitter memungkinkan pengelola destinasi untuk merespons secara proaktif terhadap opini wisatawan, memperbaiki strategi layanan, dan meningkatkan kepuasan pelanggan.

III. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen wisatawan terhadap destinasi pariwisata Indonesia. Data yang digunakan diperoleh dari Twitter melalui API yang disediakan oleh *platform* tersebut. Tahapan penelitian ini meliputi pengumpulan data, prapemrosesan data teks, pelatihan model SVM dengan berbagai kernel, serta evaluasi model untuk mengukur kinerjanya dalam klasifikasi sentimen.

A. Pengumpulan Data

Data diperoleh dari Twitter menggunakan *Application Programming Interface* (API) yang memungkinkan akses ke *tweet* publik berdasarkan kata kunci yang berkaitan dengan destinasi wisata di Indonesia. Data yang dikumpulkan mencakup *tweet* yang berisi ulasan wisatawan tentang destinasi tersebut. Gambar 1 adalah alur pengumpulan data menggunakan API Twitter:



Gambar 1 Flowchart Pengumpulan Data API

B. Prapemrosesan Teks

Prapemrosesan teks merupakan tahapan penting yang bertujuan untuk memastikan agar data tidak terstruktur dapat diolah secara optimal oleh algoritma *Support Vector Machine* (SVM). Tahapan ini mencakup beberapa langkah utama, dimulai dengan *case folding*, yaitu mengubah seluruh teks menjadi huruf kecil guna menjaga konsistensi data. Langkah berikutnya adalah tokenisasi, yang memecah teks menjadi kata-kata individu atau token agar lebih mudah dianalisis. Setelah itu, dilakukan *stopword removal*, yakni penghapusan kata-kata umum yang tidak memiliki makna signifikan, seperti "di", "dan", atau "yang", sehingga analisis dapat lebih terfokus pada kata-kata yang relevan. Langkah terakhir adalah *stemming*, yaitu mengubah kata-kata menjadi bentuk dasar, misalnya kata "bermain" diubah menjadi "main," yang membantu menyederhanakan variasi kata sehingga data menjadi lebih mudah untuk diklasifikasikan, untuk proses prapemrosesan dapat dilihat pada Tabel 1.

TABEL 1
PROSES PRAPEMROSESAN

Sebelum Prapemrosesan	Setelah Prapemrosesan
"Luar biasa pemandangan di Borobudur!"	"luar biasa pemandangan borobudur"
"Harga tiket masuk terlalu mahal."	"harga tiket masuk mahal"

C. Pembobotan Kata

Setelah prapemrosesan, dilakukan pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini digunakan untuk menghitung frekuensi kemunculan suatu kata dalam dataset sekaligus frekuensi kemunculannya secara keseluruhan [15]. Pembobotan TF-IDF digunakan untuk mengonversi setiap kata dalam data teks menjadi nilai numerik, sehingga dapat diolah oleh algoritma SVM. Persamaan TF-IDF adalah sebagai berikut:

$$TF-IDF(t,d) = TF(t,d) \times IDF(t) \quad (1)$$

Dimana $TF-IDF(t,d)$ adalah frekuensi kemunculan term t dalam dokumen d dan $IDF(t)$ adalah *invers* frekuensi dokumen yang mengandung term t , dihitung sebagai $\log \frac{N}{n_t}$, dengan N jumlah dokumen dan n_t adalah jumlah dokumen yang mengandung term t .

D. Pelatihan Model SVM

Model SVM dilatih dengan data yang telah melalui proses pembobotan. Penelitian ini menggunakan tiga kernel SVM berbeda—linear, radial basis *function* (RBF), dan *sigmoid*—untuk menemukan model dengan kinerja terbaik. Penggunaan kernel ini memungkinkan peneliti untuk mengevaluasi efektivitas masing-masing kernel dalam mengklasifikasikan data teks yang kompleks.

1) *Kernel Linear*

Kernel linear mencari *hyperplane* yang memisahkan dua kelas secara linier. Fungsi kernel linear adalah sebagai berikut:

$$K(x_i, x_j) = x_i * x_j \tag{2}$$

2) *Kernel Radial Basis Function (RBF)*

Kernel RBF bekerja dengan menghitung jarak antara titik data dan pusat kelas, memberikan hasil yang lebih baik untuk data yang tidak dapat dipisahkan secara linier. Fungsi kernel RBF adalah:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{3}$$

Di mana γ adalah parameter yang mengontrol jarak jangkauan pengaruh dari sebuah titik data.

3) *Kernel Sigmoid*

Kernel *sigmoid* menggunakan fungsi aktivasi *sigmoid* untuk memetakan data ke ruang yang lebih tinggi, membuatnya cocok untuk masalah klasifikasi yang kompleks. Fungsinya adalah:

$$K(x_i, x_j) = \tanh(\alpha x_i \cdot x_j + c) \tag{4}$$

Dengan α dan c adalah konstanta yang ditentukan selama pelatihan.

E. Evaluasi Model

Kinerja model dievaluasi menggunakan matriks konfusi untuk menghitung metrik-metrik utama seperti akurasi, presisi, *recall*, dan *F1-score*. *Confusion matrix* memisahkan prediksi model menjadi *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN), yang kemudian digunakan untuk menghitung metrik sebagai berikut:

Akurasi: mengukur persentase prediksi yang benar dari keseluruhan data.

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \tag{5}$$

Presisi: mengukur ketepatan prediksi positif.

$$Presisi = \frac{TP}{TP+FP} \tag{6}$$

Recall: mengukur kemampuan model dalam mendeteksi semua data positif.

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

F1-score: rata-rata harmonis antara presisi dan *recall*.

$$F1 - Score = 2 * \frac{presisi * recall}{presisi + recall} \tag{8}$$

Evaluasi ini dilakukan untuk setiap kernel, sehingga diperoleh model dengan performa terbaik dalam klasifikasi sentimen.

IV. HASIL

Penelitian ini menunjukkan performa *Support Vector Machine* (SVM) dalam klasifikasi sentimen di sektor pariwisata dengan menggunakan berbagai kernel (linear, RBF, dan *sigmoid*) serta dua rasio pembagian data (7:3 dan 8:2). Model dengan kernel linear pada rasio 7:3 memiliki akurasi tertinggi (92.89%), diikuti oleh kernel RBF (92.64%) dan *sigmoid* (92.40%). Hasil ini mengindikasikan bahwa kernel linear lebih efektif dalam menangkap pola-pola sentimen dibandingkan kernel lainnya, khususnya pada dataset dengan distribusi data training yang lebih besar. Pemilihan rasio data yang tepat juga memengaruhi performa model, di mana rasio 7:3 memberikan hasil lebih baik dibandingkan rasio 8:2. Performa model *Support Vector Machine* dapat dilihat pada Tabel 2.

TABEL 2
PERFORMA MODEL SVM PADA ANALISIS SENTIMEN

Kernel	Rasio Data (Training:Testing)	Akurasi	Precision	Recall	F1-score
Linear	7:3	92.89%	92%	74%	81%
Linear	8:2	92.27%	-	-	-
RBF	7:3	92.64%	-	-	-
RBF	8:2	92.09%	-	-	-
<i>Sigmoid</i>	7:3	92.40%	-	-	-
<i>Sigmoid</i>	8:2	91.72%	-	-	-

Evaluasi metrik per kelas sentimen menunjukkan bahwa model memiliki kemampuan yang sangat baik untuk memprediksi kelas positif (*precision* 93%, *recall* 99%, *F1-score* 96%) dan cukup baik untuk kelas negatif (*precision* 97%, *recall* 68%, *F1-score* 80%). Namun, performa pada kelas netral masih moderat (*precision* 88%, *recall* 54%, *F1-score* 67%). Hal ini mengindikasikan bahwa model lebih cenderung mendeteksi sentimen positif dibandingkan sentimen lainnya, yang kemungkinan disebabkan oleh ketidakseimbangan data latih, di mana kelas positif lebih dominan, bisa dilihat pada Tabel 3 dibawah ini.

TABEL 3
EVALUASI BERDASAR KELAS SENTIMEN

Kelas Sentimen	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Negatif	97%	68%	80%
Netral	88%	54%	67%
Positif	93%	99%	96%
Rata-rata	92%	74%	81%

Secara keseluruhan, penelitian ini menunjukkan bahwa SVM adalah algoritma yang efektif untuk analisis sentimen, terutama untuk kelas sentimen positif. Namun, performa untuk kelas netral memerlukan perhatian lebih, seperti penanganan ketidakseimbangan data melalui teknik *oversampling* atau metode alternatif. Hasil ini memberikan gambaran bahwa pengelola sektor pariwisata dapat memanfaatkan analisis sentimen untuk memahami opini wisatawan, meningkatkan layanan, serta menyusun strategi promosi berbasis data yang lebih efektif.

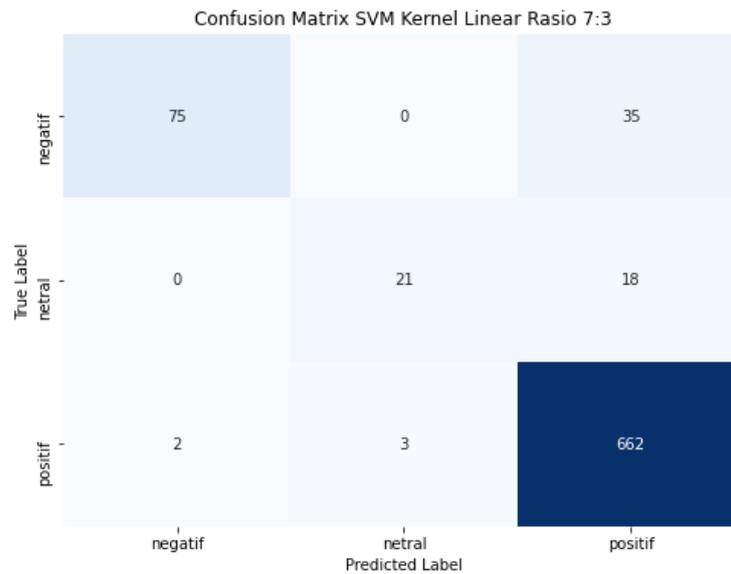
V. PEMBAHASAN

Penelitian ini menunjukkan bahwa algoritma *Support Vector Machine* (SVM) dapat diimplementasikan secara efektif dalam analisis sentimen di sektor pariwisata. SVM mampu mengklasifikasikan sentimen wisatawan dengan akurasi tinggi, terutama pada dataset dengan rasio pembagian data training dan testing 7:3. Kernel linear memberikan performa terbaik dibandingkan kernel RBF dan *sigmoid*, dengan akurasi mencapai 92.89%. Hal ini menunjukkan bahwa kernel linear lebih optimal dalam menangkap pola hubungan antar fitur yang relevan dalam dataset analisis sentimen. Pemilihan rasio data yang tepat juga memengaruhi performa model, di mana rasio 7:3 memberikan hasil lebih baik dibandingkan rasio 8:2, yang menegaskan pentingnya porsi data latih yang lebih besar untuk menghasilkan model yang lebih akurat.

Evaluasi metrik per kelas sentimen menunjukkan bahwa model sangat efektif dalam memprediksi sentimen positif (*F1-score* 96%), cukup baik untuk sentimen negatif (*F1-score* 80%), tetapi memiliki performa moderat pada sentimen netral (*F1-score* 67%). Ketidakseimbangan data latih, di mana sentimen positif mendominasi, menjadi salah satu faktor utama penyebab performa yang lebih rendah pada sentimen netral. Hal ini sejalan dengan temuan Rifaldi [14], yang menyatakan bahwa ketidakseimbangan data dapat menyebabkan bias model terhadap kelas yang lebih dominan. Selain itu, penggunaan prapemrosesan teks yang baik, seperti tokenisasi dan stemming, terbukti meningkatkan akurasi model, sebagaimana disebutkan oleh Amardita [13].

Dibandingkan dengan penelitian Tuhuteru [11], yang menunjukkan Random Forest memiliki akurasi lebih tinggi daripada SVM dalam analisis sentimen wisatawan, hasil ini menegaskan bahwa SVM tetap memberikan performa kompetitif, terutama untuk dataset yang terdistribusi secara linier. Selain itu, penelitian Era [12] menemukan bahwa algoritma Naïve Bayes unggul dalam menangani data dengan distribusi non-linear. Hal ini mengindikasikan bahwa pemilihan algoritma dan kernel harus mempertimbangkan karakteristik dataset.

Gambar 2 menunjukkan *confusion matrix* yang digunakan untuk mengevaluasi performa model SVM dalam mengklasifikasikan sentimen wisatawan di sektor pariwisata. Matriks kebingungan ini memperlihatkan hasil prediksi model untuk tiga kelas sentimen: negatif, netral, dan positif. Pada kelas negatif, model berhasil memprediksi 75 data dengan benar sebagai kelas negatif (*True Positive*), tetapi 35 data kelas negatif salah diklasifikasikan ke kelas lain (*False Negative*). Untuk kelas netral, model memprediksi 21 data dengan benar, namun terdapat 18 data kelas netral yang diklasifikasikan secara salah. Sementara itu, kelas positif menunjukkan hasil terbaik dengan 662 data yang diklasifikasikan dengan benar, hanya dengan sedikit kesalahan klasifikasi (*False Positive* = 2, *False Negative* = 5). Hasil ini menunjukkan bahwa model SVM memiliki akurasi tinggi dalam memprediksi kelas positif, yang didukung oleh nilai *precision* dan *recall* yang tinggi pada kelas ini.



Gambar 2. *Confusion Matrix Model*

Confusion matrix ini memberikan gambaran visual mengenai kemampuan dan keterbatasan model dalam mengklasifikasikan data sentimen. Kinerja yang baik pada kelas positif menunjukkan bahwa model SVM dapat diandalkan dalam mendeteksi sentimen positif, yang berpotensi digunakan sebagai masukan dalam promosi pariwisata. Pengelola destinasi wisata dapat menyoro aspek-aspek yang memperoleh ulasan positif, seperti keindahan dan kenyamanan fasilitas, untuk menarik lebih banyak pengunjung. Sebaliknya, sentimen negatif yang teridentifikasi dapat digunakan sebagai bahan evaluasi untuk memperbaiki aspek-aspek yang sering dikritik, seperti kebersihan dan aksesibilitas, guna meningkatkan kualitas layanan wisata.

Di sisi lain, nilai *recall* yang rendah pada kelas netral menunjukkan kesulitan model dalam mengenali pola sentimen yang bersifat netral, yang dapat dipengaruhi oleh ketidakseimbangan jumlah data antara kelas positif dan netral. Model tampaknya lebih "bias" terhadap sentimen positif, yang mungkin disebabkan oleh dominasi data latih pada kelas tersebut. Hal ini menunjukkan bahwa, untuk mendapatkan hasil yang lebih seimbang dan akurat, diperlukan penanganan khusus pada ketidakseimbangan data, misalnya dengan menerapkan teknik resampling atau menggunakan algoritma alternatif yang lebih sensitif terhadap variasi kelas sentimen.

Selain itu, meskipun akurasi model dengan rasio data 8:2 sedikit lebih rendah dibandingkan rasio 7:3, performanya tetap berada pada kisaran tinggi (di atas 92%). Hal ini menunjukkan bahwa SVM cukup robust terhadap variasi pembagian data, meskipun rasio data training yang lebih besar memberikan hasil yang lebih baik secara keseluruhan. Penggunaan kernel lain, seperti RBF dan *sigmoid*, meskipun memiliki akurasi yang mendekati kernel linear, menunjukkan kinerja yang kurang optimal dalam menangani data analisis sentimen pada penelitian ini.

Temuan ini memberikan wawasan berharga tentang pentingnya pemilihan rasio data, kernel, dan penanganan ketidakseimbangan data dalam membangun model *machine learning* yang efektif untuk analisis sentimen. Untuk meningkatkan performa, penelitian lanjutan dapat mempertimbangkan penggunaan teknik pengolahan data tambahan, seperti oversampling atau undersampling, serta penerapan algoritma deep learning untuk menangkap pola yang lebih kompleks dalam data.

Dengan demikian, implementasi SVM dalam analisis sentimen memberikan kontribusi nyata dalam pengelolaan pariwisata berbasis data, membantu pengelola dalam memahami opini wisatawan, meningkatkan kualitas layanan, dan menyusun strategi promosi yang lebih efektif. Hal ini tidak hanya mendukung pengembangan pariwisata tetapi juga memperkuat daya saing destinasi wisata di tengah persaingan global.

VI. KESIMPULAN

Berdasarkan hasil penelitian yang telah dipaparkan, terdapat beberapa poin penting yang dapat disimpulkan. Model *Support Vector Machine* (SVM) dengan kernel linear dan rasio pembagian data 7:3 menunjukkan performa terbaik dibandingkan model SVM dengan kernel RBF dan *sigmoid*. Model ini mencapai akurasi 92,89%, *precision* 92%, *recall* 74%, dan *F1-score* 81%. Evaluasi per kelas sentimen menunjukkan kemampuan tinggi dalam mendeteksi sentimen positif (*F1-score* 96%), cukup baik untuk sentimen negatif (*F1-score* 80%), tetapi moderat untuk sentimen netral (*F1-score* 67%).

Hasil ini memiliki relevansi praktis yang signifikan untuk pengelola destinasi pariwisata. Dengan memanfaatkan analisis sentimen berbasis SVM, pengelola dapat memahami opini wisatawan secara lebih

mendalam dan berbasis data. Sentimen positif yang dominan dapat digunakan untuk mempromosikan aspek unggulan destinasi, seperti keindahan alam dan kenyamanan fasilitas. Sebaliknya, sentimen negatif memberikan wawasan penting untuk meningkatkan layanan, memperbaiki infrastruktur, dan mengatasi masalah yang sering dikritik, seperti kebersihan dan aksesibilitas.

Rekomendasi praktis lainnya mencakup integrasi analisis sentimen ke dalam strategi manajemen destinasi berbasis data. Sistem ini dapat digunakan untuk memantau opini wisatawan secara *real-time*, memungkinkan pengelola untuk merespons secara cepat terhadap kebutuhan dan preferensi wisatawan. Dengan demikian, pendekatan ini tidak hanya meningkatkan kualitas pengalaman wisata tetapi juga memperkuat daya saing destinasi pariwisata Indonesia di pasar global.

Penelitian ini menyarankan pengembangan lebih lanjut, seperti penerapan teknik *resampling* untuk mengatasi ketidakseimbangan data atau eksplorasi algoritma lain yang dapat menangkap pola sentimen dengan lebih kompleks. Langkah ini diharapkan dapat meningkatkan keakuratan model dan memberikan wawasan yang lebih kaya bagi sektor pariwisata.

DAFTAR PUSTAKA

- [1] Q. Azzahra, "Kontribusi sektor pariwisata terhadap PDB 2017-2021," *alineia.id*. Accessed: Apr. 01, 2023. [Online]. Available: <https://data.alinea.id/kontribusi-sektor-pariwisata-terhadap-pdb-2017-2021-b2feX9CV9b>
- [2] M. A. Rizaty, "Pengguna Twitter di Indonesia," *DataIndonesia.id*.
- [3] I. Surgawi and Sutopo, "ANALISIS PENGARUH PRODUK WISATA, PERSEPSI HARGA DAN PROMOSI TERHADAP KEPUTUSAN WISATAWAN DALAM MENGUNJUNGI OBJEK WISATA (Studi pada Objek Wisata Puri Maerokoco Kota Semarang)," *Diponegoro J. Manag.*, vol. 5, no. 4, pp. 1–10, 2016.
- [4] C. C. Aggarwal, *Data Classification*. 2015. doi: 10.1007/978-3-319-14142-8_10.
- [5] D. Purnamasari, *Pengantar Metode Analisis Sentimen*. 2023.
- [6] Amna et al., *Data Mining*, vol. 2. 2023.
- [7] I. R. Mukhlis et al., *ALGORITMA PEMBELAJARAN MESIN (Dasar , Teknik , dan Aplikasi)*, no. April. 2024.
- [8] R. Albin Pranata, N. Azmi Verdikha, U. Muhammdiyah Kalimantan Timur, and J. H. Ir Juanda, "Metode Pembobotan Tf-Idf Untuk Klasifikasi Teks Quick Count Pemilihan Wakil Presiden Indonesia 2024 Pada X Twitter Dengan Metode Svm," vol. 18, no. 2, p. 126, 2024, [Online]. Available: <https://doi.org/10.47111/JTIAvailableonlineathttps://e-journal.upr.ac.id/index.php/JTI>
- [9] Septi Putri, Yohanes Agung Apriyanto, and Andri Wijaya, "Analisis Sentimen Ulasan Aplikasi Deepl Pada Google Play Dengan Metode Support Vector Machine (Svm)," *J. Sist. Inf.*, vol. 4, no. 2, pp. 59–66, 2023, doi: 10.32546/jusin.v4i2.2368.
- [10] J. Alcaraz, M. Labbé, and M. Landete, "Support Vector Machine with feature selection: A multiobjective approach," *Expert Syst. Appl.*, vol. 204, p. 117485, 2022, doi: <https://doi.org/10.1016/j.eswa.2022.117485>.
- [11] H. Tuhuteru, L. P. Refialy, M. Laturake, and S. Gildion Pattirane, "Tourist Perceptions Through Sentiment Analysis to Support Tourism Development in Maluku Province," *J. Appl. Informatics Comput.*, vol. 8, no. 1, p. 119, 2024.
- [12] D. Era, S. Andryana, and A. Rubhasy, "Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor pada Analisis Sentimen Pembukaan Pariwisata Di Masa Pandemi Covid 19," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 7, no. 1, pp. 263–272, 2023.
- [13] R. S. Amardita, A. Adiwijaya, and M. D. Purbolaksono, "Analisis Sentimen terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 1, p. 62, 2022, doi: 10.30865/jurikom.v9i1.3793.
- [14] D. Rifaldi, Abdul Fadlil, and Herman, "Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet 'Mental Health,'" *Decod. J. Pendidik. Teknol. Inf.*, vol. 3, no. 2, pp. 161–171, 2023, doi: 10.51454/decode.v3i2.131.
- [15] M. A. A, M. Alamsyah, and M. F. Arif, "Analisis Sentimen Twitter Tentang Pinjaman Online di Indonesia Menggunakan Metode Random Forest".