

ANALISIS SENTIMEN: PREDIKSI *RATING* TERHADAP *REVIEWS* WISATAWAN TANJUNG PUTING PADA TRIPADVISOR MENGGUNAKAN *SUPPORT VECTOR MACHINE*

Quratul Ain^{*1)}, Ema Utami²⁾, Asro Nasiri³⁾

1. Universitas Amikom Yogyakarta, Indonesia
2. Universitas Amikom Yogyakarta, Indonesia
3. Universitas Amikom Yogyakarta, Indonesia

Article Info

Kata Kunci: *Online Reviews*; Pariwisata; *Rating*; SVM; TripAdvisor.

Keywords: *Online Reviews*; Pariwisata; *Rating*; SVM; TripAdvisor.

Article history:

Received 12 June 2024

Revised 10 July 2024

Accepted 14 August 2024

Available online 1 September 2024

DOI :

<https://doi.org/10.29100/jipi.v9i3.5430>

* Corresponding author.

Corresponding Author

E-mail address:

ainquratul@students.amikom.ac.id

ABSTRAK

Di era digital, sebagian besar data di internet berbentuk teks mentah. Data tambang emas ini sangat berharga karena berisi banyak informasi mendasar yang dapat diekstraksi menggunakan natural language processing dengan analisis sentimen. Termasuk data *Reviews* dan *rating* online di era digital menjadi hal pertimbangan penting dan terpercaya bagi wisatawan sebelum berkunjung ke sebuah objek wisata, salah satunya situs TripAdvisor. Penelitian ini akan memproses data teks ulasan wisatawan pada TripAdvisor pada objek Taman Nasional Tanjung Puting untuk menentukan sentimen wisatawan berdasarkan *rating* 1 hingga 5. Data penelitian yang digunakan adalah sebanyak 390 *reviews* yang telah di-export dengan TripAdvisor *Reviews Scrapper*. Penelitian ini menggunakan Metode Support vector machine (SVM) yang dikombinasikan dengan feature extraction dan Truncate singular-value decomposition (TSVD) pada preprocessing data. Penelitian ini dapat menampilkan visualisasi wordcloud top 10 kata yang paling banyak muncul berdasarkan *rating* sebagai pengetahuan dan informasi bagi pelaku industri pariwisata. Berdasarkan pengujian dengan pembagian data set 30:70 tingkat akurasi memperoleh rata-rata 80%.

ABSTRACT

In the digital era, most of the data on the internet is in the form of raw text. This gold mine data is very valuable because it contains a lot of fundamental information that can be extracted using natural language processing with sentiment analysis. Including online review and rating data in the digital era is an important and reliable consideration for tourists before visiting a tourist attraction, one of which is the TripAdvisor site. This research will process text data from tourist reviews on TripAdvisor on Tanjung Puting National Park objects to determine tourist sentiment based on ratings from 1 to 5. The research data used is 390 reviews that have been exported with TripAdvisor *Reviews Scrapper*. This research uses the Support vector machine (SVM) method combined with feature extraction and Truncate singular-value decomposition (TSVD) in data preprocessing. This research can display a wordcloud visualization of the top 10 words that appear most frequently based on ratings as knowledge and information for tourism industry players. Based on testing with a 30:70 dataset division, the accuracy level obtained an average of 80%.

I. PENDAHULUAN

KONSUMEN atau pelanggan di era digital saat ini akan mencari informasi yang relevan untuk mengurangi ketidakpastian pilihan sebelum membeli barang atau jasa. *Rating* dan *review* merupakan data yang paling terpercaya dalam proses pencarian informasi konsumen.[1]. Hal ini mendorong industri pariwisata mempromosikan produk dan layanannya berdasarkan ulasan (*reviews*) yang ditulis oleh wisatawan pada website travel seperti TripAdvisor.com[2]. Platform ulasan TripAdvisor menjadi salah satu sumber informasi paling banyak digunakan bagi para wisatawan dalam merencanakan perjalanan mereka. Ulasan wisatawan di platform ini

memberikan pandangan pribadi, pengalaman, dan penilaian terhadap destinasi wisata, memberikan petunjuk berharga bagi calon wisatawan yang ingin menjelajahi tempat baru[2]. Taman Nasional Tanjung Puting(TNTP), sebagai salah satu destinasi ekowisata, tak terkecuali menjadi fokus perhatian para pengguna TripAdvisor untuk memberikan ulasan setelah berkunjung. Destinasi TNTP adalah objek wisata Kalimantan Tengah, Indonesia yang menjadi nomor satu dan terfavorit di TripAdvisor.

Meningkatnya jumlah ulasan pengguna *online*, terdapat juga peningkatan kebutuhan akan pemrosesan otomatis sejumlah data yang besar ini. Sehingga digunakan teknik untuk memproses data tersebut seperti dengan analisis sentimen. Mengingat banyaknya data mengenai opini individu masyarakat, nalisis sentimen mempunyai kebutuhan untuk dikembangkan dan disempumakan. Alat-alat analisis sentimen ini tidak hanya berfungsi sebagai pemberi rekomendasi bagi individu mengenai cara mengoptimalkan pilihan layanan yang akan mereka gunakan, namun juga bagi pengambil keputusan dalam meningkatkan kualitas layanan mereka.[3].

Penelitian sebelumnya[3] menjelaskan dalam menganalisis sentimen ulasan wisatawan, sebagian besar terbatas pada klasifikasi positif/negatif. Selain itu, penelitian yang memasukkan klasifikasi peringkat dan prediksi berdasarkan ulasan wisatawan lebih sedikit. Sementara pada penelitian Ahmed dan Ghabayen disebutkan prediksi peringkat(*rating*) ulasan adalah prinsip dalam analisis sentimen[4]. Oleh karena itu, penelitian ini akan membahas tentang analisis sentimen terhadap prediksi peringkat (*rating*) ulasan wisatawan pada platform TripAdvisor dengan Taman Nasional Tanjung Puting sebagai studi kasus.

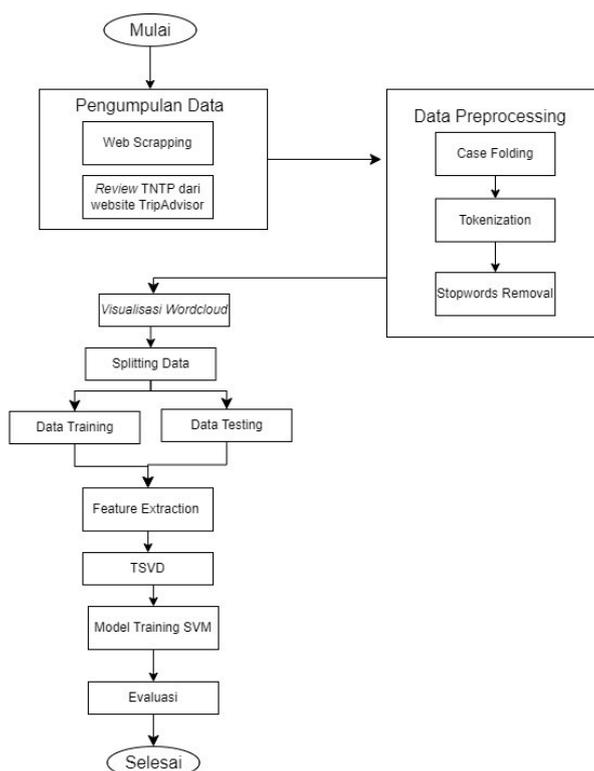
Penelitian mengenai analisis sentimen ulasan sebelumnya [5] untuk mengukur performa algoritma *Support vector machine*(SVM) dalam klasifikasi ulasan menjadi sentimen positif, negatif dan netral dari 2.000 data objek. Kelebihan penelitian ini adalah terdapat pengujian beberapa penggunaan kernel SVM, seperti *Radial basis function (RBF)*, *Polynomial*, *kernel Linear*, dan *Sigmoid*. Hasil pengujian *confusion matrix* yang menunjukkan tingkat akurasi 80,36%. Penelitian lainnya yang telah dilakukan oleh[6] melakukan klasifikasi data *review* berdasarkan kategori sentimen untuk mengekstraksi wawasan sejumlah besar data dari TripAdvisor dengan metode SVM. Hasil menunjukkan tiap kategori terdapat 305 *review* positif, 60 *review* netral dan sisanya 36 *review* negatif. Akurasi yang dilakukan model SVM adalah 100% yang telah diuji dengan 401 data pengujian dan dilatih dengan 247 data. SVM juga disebutkan dalam penelitian Puh K, dkk [3] memuskan dalam tugas sentimen pengaturan tertentu yaitu 80% dan *pra-processing* yang digunakan lebih sederhana dan konsumsi memori lebih sedikit. Metode *machine learning* dalam konteks analisis sentimen terkait pariwisata, metode yang paling banyak digunakan adalah SVM dan juga Naive Bayes seperti yang disebutkan pada penelitian Chen, dkk [7]. Dibandingkan dengan jaringan saraf(NN), SVM dan Naive Bayes memerlukan lebih sedikit anotasi kelas untuk melatih model. Sebagian besar penelitian tentang *online travel review text* menunjukkan analisis sentimen teks berbasis SVM memberikan hasil yang lebih unggul dibandingkan metode *machine learning* lainnya[7]. Penelitian ulasan situs TripAdvisor menggunakan metode linear SVM juga dilakukan oleh Imamah dkk[8]terkait pariwisata di kabupaten Bangkalan. Data *Review* yang sudah diberi label positif dan negatif kemudian diklasifikasi menggunakan metode Linear SVM mempunyai akurasi 70,22% untuk *review* berbahasa Indonesia.

Berdasarkan penelitian-penelitian sebelumnya penulis menggunakan *Support Machine Learning*(SVM) sebagai metode pada penelitian ini.

Berdasarkan uraian sebelumnya penelitian ini bertujuan mengetahui performa model SVM dalam analisis sentimen terhadap prediksi *rating* berdasarkan ulasan wisatawan TripAdvisor pada objek wisata Taman Nasional Tanjung Puting serta menganalisis ulasan wisatawan dengan mengidentifikasi dan mengkategorikan opini secara umum berdasarkan peringkat 1 hingga 5. Penelitian ini diharapkan dapat menjadi dasar bagi penyedia layanan wisata untuk memahami preferensi dan kebutuhan wisatawan berdasarkan ulasan atau *review* dan *rating* di situs *online*, serta dapat memberikan pengetahuan untuk pengembangan strategi pemasaran yang lebih tepat sasaran dengan memahami sentimen wisatawan, selain itu penelitian ini dapat berkontribusi pada peningkatan kualitas layanan pemangku kepentingan dapat memfasilitasi pengambilan keputusan yang lebih informasional berdasarkan data.

II. METODE PENELITIAN

Penelitian ini menggunakan alur skema penelitian yang ditunjukkan pada gambar 1.



Gambar. 1. Diagram Alur Penelitian

A. Pengumpulan Data

Tahap ini melakukan teknik *crawling*, dataset *review* dalam bentuk file CSV yang di-*export* dengan *TripAdvisor Review Scraper* terhadap Taman Nasional Tanjung Puting (*Tanjung Puting National Park*). Berdasarkan pengumpulan data hingga 23 Januari 2024 didapatkan data sejumlah 390 baris data ulasan dengan *rating* bintang 1 hingga 5. Data ulasan merupakan kumpulan ulasan yang terdiri dari bahasa Inggris, Jepang, Belanda, Itali, Prancis, Norwegia, Spanyol, Portugal, Jerman dan Polandia.

B. Data Pre-processing

Setelah data diambil, selanjutnya dilakukan *preprocessing* yaitu tahapan membersihkan data ulasan atau kata-kata yang tidak di perlukan serta kata-kata yang tidak memiliki informasi. Proses tahapan ini dilakukan sesuai dengan isi data hasil proses pengambilan data atau *export* dari TripAdvisor[9]. Adapun tahapan yang digunakan pada *data pre-processing* ini adalah sebagai berikut.

- 1) *Case folding*: proses penyederhanaan karakter dalam teks dengan mengubah setiap huruf besar menjadi huruf kecil dan menghilangkan tanda baca dan karakter khusus dari dokumen [5]. Data huruf diubah menjadi huruf kecil dan karakter dianggap sebagai delimiter jika selain huruf 'a' sampai dengan 'z' [10].
- 2) *Tokenization*: yaitu tahap pemotongan kalimat menjadi kata [10] di mana pada *review* yang sudah menjadi huruf kecil, dan penandaannya hilang akan diberi token untuk dipecah menjadi beberapa bagian untuk dipisahkan menjadi kata-kata[11]
- 3) *Stopwords Removal*: menghilangkan kata-kata yang kurang penting menggunakan stoplist. Stopwords yaitu kata-kata yang sering muncul pada ulasan tetapi dianggap tidak bermakna informasi. Stopwords dalam bahasa Indonesia seperti "dan", "yang", "di", dll. Stopwords digunakan sehingga penelitian dapat fokus pada kata-kata yang penting[9]. Setiap kata yang terdaftar dalam stop list (kamus negatif) yang hentikan (disaring) setelah atau sebelum pemrosesan data teks bahasa alami. *Stop words* adalah kata umum yang biasanya dalam jumlah besar muncul namun dianggap tidak memiliki makna[12].

C. Visualisasi Data dengan WordCloud

Tahap visualisasi dilakukan untuk kelas sentimen untuk setiap klasifikasi. Visualisasi digunakan dengan tujuan menampilkan informasi mengenai emosi positif dan negatif serta menggali informasi dalam bentuk topik yang paling sering dibicarakan[13]. *Wordcloud* adalah sebuah gambar elektronik yang menyajikan kata-kata dalam series teks. Besarnya ukuran kata-kata ditentukan oleh banyaknya frekuensi kata tersebut.[14].

D. Split data

Tahap Ini adalah proses membagi kumpulan data menjadi data pelatihan dan data pengujian. Data pelatihan dimaksudkan untuk melatih algoritma, dan data pengujian dimaksudkan untuk mengetahui kinerja algoritma [13].

E. Feature Extraction

Pada tahap Selanjutnya, memisahkan data pelatihan menjadi data numerik dan tekstual dan melihat apakah algoritma bekerja lebih baik jika datanya numerik atau tekstual [15]. Mengubah data teks menjadi numerik penulis menggunakan model *bag of word* (BoW) agar pengklasifikasi dapat memahaminya dengan memecah kalimat ulasan menjadi kata-kata (token) dengan *Countvectorizer*. Tahapan ini melibatkan konversi kata-kata dalam ulasan dari teks menjadi bilangan bulat sehingga memiliki representasi numerik yang dapat digunakan dalam model pembelajaran mesin [16]. Website resmi *scikit-learn* menyebutkan metode *CountVectorizer* berfungsi mengubah teks menjadi *vector* dengan *stop word* standar Bahasa Inggris [15]. Model *bag-of-words* membangun kosakata dari semua kata dalam dokumen. Kemudian menghitung berapa kali sebuah kata dalam kosakata muncul dan memodelkan setiap dokumen. Mengingat kumpulan data berisi ulasan dalam jumlah besar, sehingga menghasilkan kosakata yang banyak, penulis dapat membatasi ukuran vektor fitur dengan memilih ukuran kosakata maksimum dari 5000 kata yang paling sering muncul. Fitur *bag-of-words* dibuat untuk set pelatihan dan pengujian. [14]

F. Singular-value decomposition (SVD)

Singular Value Decomposition (SVD) adalah teknik reduksi dimensi yang berguna untuk mereduksi nilai kompleksitas ketika berhadapan dengan matriks *term-dokumen* [17]. (SVD) adalah suatu metode yang memisahkan suatu matriks menjadi tiga bagian yaitu *left eigenvectors*, *singular values*, dan *right eigenvectors*. Ini dapat digunakan untuk menguraikan data seperti gambar dan teks [18]. Tahap ini mengurangi waktu dan ruang penyimpanan dengan mengubah dimensi yang lebih tinggi menjadi dimensi yang lebih rendah. Dengan menghilangkan multi-kolinearitas, ini meningkatkan kinerja algoritma pembelajaran mesin. *Truncate Singular-value decomposition* TSVD lebih disukai daripada teknik reduksi dimensi lain karena kesederhanaannya dan biaya komputasi yang relatif rendah [19]. Penggunaan SVD pada penelitian Andana dkk bukan bertujuan untuk meningkatkan akurasi tetapi untuk mengimbangi waktu pemrosesan yang besar dengan tetap menjaga kinerja algoritma yang baik. SVD tidak mampu mempengaruhi kinerja algoritma menjadi lebih baik seperti penelitian lain yang sudah ada. Fitur yang dikurangi dengan SVD menyebabkan hilangnya nilai uniknya yang merupakan representasi data dalam suatu dokumen. Sehingga semakin mengurangi keakuratan algoritma yang digunakan. SVD didukung untuk membuat ekstraksi informasi lebih efisien namun tidak selalu bekerja dengan baik. Selain mengukur performa algoritma dari presisi dan *recall*, juga diukur dari waktu pemrosesan [20]

G. Model Training SVM

Support Vector Machine (SVM) adalah algoritma *machine learning* yang menggunakan *hyperplane* untuk memisahkan kelas data yang berbeda. *Hyperplane* adalah subruang yang selalu berukuran satu dimensi lebih kecil dari dimensi induknya. Misalnya, jika kita berada di ruang dua dimensi maka *hyperplane* akan berupa garis. Tujuan utama dari algoritma ini adalah untuk mencari *hyperplane* yang mempunyai jarak (*margin*) terbesar antara *hyperplane* dengan data terdekat yang disebut *support vector*. Data baru sedang diklasifikasikan berdasarkan sisi *hyperplane* mana mereka berada. Selain itu, semakin besar marginnya, semakin besar keyakinan kita dalam menentukan kelas data [3]. SVM adalah sebuah teknik *supervised learning* untuk mengklasifikasikan data. Kernel linier dipilih pada penelitian ini karena memiliki beberapa keunggulan seperti akurasi yang tinggi, menghindari overkonfigurasi, pelatihan lebih cepat, dan efisiensi penggunaan big data. Untuk menggunakan SVM, teks diubah menjadi data vektor. [21]. Tahap klasifikasi dengan menggunakan metode SVM, mencari batas keputusan (*hyperline*) terbaik yang memisahkan setiap review kedalam dua kelas yaitu Review negatif dan review positif. Data tersebut selanjutnya akan digunakan pada prediksi data testing [13]. *Support Vector Machine (SVM)* adalah sistem pembelajaran yang klasifikasinya menggunakan ruang hipotesis berupa fungsi linier dalam ruang fitur berdimensi tinggi, yang dilatih dengan algoritma pembelajaran berdasarkan teori optimasi, menerapkan bias pembelajaran yang berasal dari teori pembelajaran statistik [22].

H. Evaluasi

Tahap evaluasi ditampilkan dengan *confusion matrix* untuk mengevaluasi performa model yang dibentuk oleh algoritma klasifikasi. Dalam melakukan evaluasi model, untuk mendapatkan nilai akurasi terbaik dapat dilakukan lima kali percobaan dataset [5] namun pada penelitian ini hanya menggunakan satu kali percobaan dataset.

Penggunaan *confusion matrix* dalam evaluasi agar mengetahui perkiraan yang salah dan benar dari hasil klasifikasi[13] Kinerja algoritma diukur dengan empat metrik yaitu presisi, *recall*, akurasi dan *f-measure*[13]

III. HASIL PEMBAHASAN

A. Pengumpulan Data

Dataset pada penelitian ini adalah hasil export dari *TripAdvisor Review Scrapper* dengan 390 baris data ulasan terdiri dari 11 variabel sebagai berikut.

- *user name*
- *user profile*,
- *user avatar*,
- *user location*,
- *rating*,
- *rating text*,
- *Review Title*,
- *Review Text*,
- *photos*,
- *published date*,
- *disclaimer*.

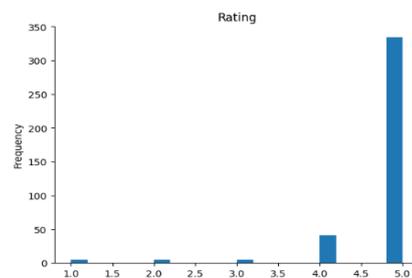
Karena penelitian difokuskan untuk mempelajari sentimen dari ulasan wisatawan dan bagaimana hal tersebut memprediksi rating, variabel yang relevan dipilih untuk analisis adalah “*Review Text*,” “*Rating*”. Sample dataset yang menjadi objek dalam penelitian ini adalah seperti pada Gambar 2. Perbandingan jumlah ulasan berdasarkan *rating* ditunjukkan pada Tabel 1 dan Gambar 3 .

index	Rating	Review Text
385	5	Via een tour een boot tocht van meerdere dagen door dit gebied gemaakt. Prachtig om alle natuur en de wilde dieren te zien maar ook treuring om te horen hoe de lokale bevolking alles platbrand voor de palmolie.
386	5	Visitar el Parque Nacional Tanjung Puting fue una experiencia unica- />La jungla es realmente increible y el contacto con la vida salvaje y con los orangutanes es algo extraordinario- />Recomiendo sacar un minimo de 3 noches a bordo de un barco y de esa manera poder explorar con mucho detenimiento la selva- />Para ver animales hay que ir despacio y prestando atencion, la gente que va a las apuradas se pierden todo.- />Nosotros pasamos 4 dias y en ese tiempo vimos orangutanes a montones, monos narizones, macacos cangrejeros, macacos cola de mono, gibones, varanos de agua, jabalies, tupayas, ardillas, dragones voladores, gaviales, cocodrilos de agua salada y hasta una piton de 5 metros atacando y deborando un jabali.- />La unica pega es que la visita te deja un sabor agri dulce ya que al estar alli puedes ver el gran problema que existe producto de la produccion de aceite de palma, que esta destruyendo todo el ambiente.- />Cuando estuvimos nosotros los incendios quemaron 70.000 hectareas de selva. Actualmente solo el 40% del parque conserva la selva.- />El turismo es la ultima esperanza para la supervivencia de este magico lugar
387	4	Taman Nasional Tanjung Puting Bagi saya adalah Sebagai Gerbang Masuk Wisata Ecotourism dan merupakan Salah satu objek wisata unggulan di Kalimantan Tengah. sebagai wahana pelestarian orang utan yang terkenal di negara kita dan mancanegara. Di musim liburan selain pengunjung lokal dan wisatawan domestik. banyak turis asing berwisata berkunjung ke sana. Bila anda pergi dan melakukan aktifitas wisata anda maka disana akan disuguhkan berbagai macam aktifitas wisata pilihan tersedia paket-paket tour yang ditawarkan travel agent atau biro perjalanan dengan dipandu guide profesional aktifitas wisata yang bisa dilakukan adalah seperti susur sungai, trekking menjelajahi hutan, camping, penelitian, melihat orang hutan dan lain-lain. Akses menuju kesana dilakukan dengan aktifitas perjalanan yang bervariasi ketika tiba di kumai registrasi ke petugas, kita diberikan arahan mengenai pengenalan Taman Nasional Tanjung Puting. Setelahnya kita menyusuri teluk kumai dan masuk sungai sekonyer dengan menggunakan kapal atau speed boat pemandangan lanscape alamnya sangatlah indah ketika menyusuri sungai sekonyer ada beberapa tempat dermaga persinggahan hingga persinggahan terakhir di camp leakey.saran saya selama aktifitas wisata di tanjung puting janganlah mengotori sungai, jangan membuang sampah sembarangan di lokasi manapun, taati aturan dan ketentuan yang berlaku selama berwisata di lokasi Taman Nasional Tanjung Puting, anda Tertarik? Datang dan rasakan pengalaman seru di Taman Nasional Tanjung Puting
388	4	Akhirnya saya bisa berkunjung ke Camp Leakey Taman Nasional Tanjung Puting meskipun dalam rangka tugas. Perjalanan kurang lebih 1,5 jam dengan speed boat menyusuri sungai sekonyer terbayar saat melihat orang utan begitu bebasnya berjalan-jalan disekitar dermaga camp leakey, namun tidak perlu khawatir akan diserang oleh orang utan karena selalu dipantau oleh para polhut yang penting kita mendengarkan arahan para Polhut tersebut. Wisatawan mancanegara biasanya menyewa kapal, mereka makan dan bermalam diatas kapal di dalam kawasan Taman Nasional Tanjung Puting demi bisa melihat orang utan dan satwa lainnya.
389	5	Realmente yo era bastante reacia a pasar dos dias en medio de la jungla durmiendo en la cubierta de un barquito remontando un río en Borneo! Pero la experiencia lejos de decepcionarme me maravilló. Si bien la primera noche tuve un momento de angustia cuando la lluvia torrencial parecía que iba a traspasar la lona y la negrura de la noche sólo se interrumpía por la luz de los relámpagos, reconozco que para la segunda ya estaba habituado a los sonidos y a dormir a la intemperie. La comida en el barco fue de lo mejor que hemos comido en Indonesia. Una pena que el guía hablara tan poco inglés ya que cuánto más fluido, obviamente más enriquecedor! El contacto con la naturaleza y los orangutanes hace que esta sea una experiencia imperdible!!

Gambar. 2. Sample Dataset Review Text dan Rating Wisatawan

TABEL I JUMLAH ULASAN BERDASARKAN RATING

Rating	Jumlah ulasan
5	334
4	41
3	5
2	5
1	5
Total ulasan	390



Gambar. 3. Grafik Perbandingan Jumlah Review

B. Pengolahan Data (*Data Pre-Processing*)

Analisis Sentimen:

Prediksi Rating Terhadap Reviews Wisatawan Tanjung Puting Pada Tripadvisor Menggunakan Support Vector Machine

Proses pengolahan data terdiri dari proses *case folding*, *remove punctuation* dan *stopwords removal*.

1) *Case folding*

Proses penyederhanaan huruf dalam teks dengan mengubah semua huruf besar menjadi huruf kecil dan menghapus tanda baca dan karakter khusus yang terdapat pada ulasan wisatawan seperti #@:;”. Gambar 4 menunjukkan hasil data yang telah menjadi huruf kecil dan Gambar 5 menunjukkan data yang telah dihapus tanda baca atau simbol-simbol atau disebut *remove punctuation*.



Gambar. 4. Data Sebelum Dan Sesudah Case Folding



Gambar. 5. Data Hasil Punctuation Remove

2) *Tokenization*

Data *review* yang sudah melalui *case folding* diberi token untuk dipecah dan dipisahkan menjadi kata-kata. Sample hasil *tokenization* ditunjukkan pada Gambar 6.



Gambar. 6. Hasil Tokenization

3) *Stopwords Removal*

Stopwords adalah setiap kata yang muncul dalam jumlah besar namun tidak memiliki makna akan didaftarkan pada stoplist. Pada data *review text* wisatawan diimplementasikan dengan menyaring kata-kata yang dianggap tidak memiliki makna terhadap sentimen dan prediksi *rating* seperti pada kata-kata yang didaftarkan pada Tabel II.

TABEL II DAFTAR KATA STOPWORDS REMOVAL
Daftar Kata yang Diabaikan dalam Analisa Data
en, u, e, de, a, de, la, lo, br, un, es, en, et, el, y, por, muy, day, con, wasn't, back, at, pero, et, los, change, nos, et, next, del, una, otro, bit, di, las, us, dan, yang, one, que, saw, lot, many, will, se, para, really, que.

C. Visualisasi Kata-kata Ulasan Konsumen menggunakan *WordCloud*

Penelitian ini mempelajari sentimen dan sifat kata-kata yang digunakan dalam teks ulasan, ulasan dibagi menjadi 5 subkelompok berdasarkan peringkat (*rating*) yang diberikan yaitu *excellent*= bintang 5, *very good*= bintang 4, *good*= bintang 3, *average*= bintang 2, dan *bad*= bintang 1. Penggunaan *Word Cloud* untuk memvisualisasikan kata-kata yang paling sering digunakan dalam ulasan berdasarkan *rating* 1 hingga 5. Visualiasi data *Wordcloud* ditampilkan pada Gambar 7,8,9,10,11,dan 12.



Gambar. 7. TOP 10 Kata Seluruh Ulasan Wisatawan

Gambar 8 menunjukkan 10 kata yang paling sering muncul pada keseluruhan ulasan wisatawan Taman Nasional Tanjung Putting (TNTP) yaitu “klotok” atau kapal yang digunakan di kegiatan wisata, “orangutan”, “tour”, “boat”,

“see”, “tanjung putting”, “guide”, “jungle”, “trip”, dan “river”. Sehingga wisatawan lebih banyak memberikan ulasan yang berkaitan dengan kata-kata tersebut. Kata-kata yang paling banyak muncul pada ulasan adalah berbahasa Inggris.



Gambar. 8. Top 10 Kata Rating Bintang 5 (Excellent)



Gambar. 9. Top 10 Kata Rating Bintang 4 (Very Good)



Gambar. 10. Top 10 Kata Rating Bintang 3 (Good)



Gambar. 11. Top 10 Kata Rating Bintang 2 (Average)



Gambar. 12. Top 10 Kata Rating Bintang 1 (Bad)

Gambar 8,9,10,11 dan 12 menunjukkan top 10 kata-kata dari rating 1 hingga 5, di mana telah ditemukan bahwa rating 1 & 2 didominasi oleh Bahasa selain Inggris yaitu Spanyol, Italia, Perancis dan Portugis. Sedangkan rating 3 hingga 5 didominasi oleh Bahasa Inggris. Ini dapat disimpulkan bahwa wisatawan yang menggunakan bahasa selain Bahasa Inggris adalah wisatawan yang sering memberikan sentimen negatif atau rating 1 atau 2 kepada Taman Nasional Tanjung Puting.

D. Analisis Sentimen Data

Data yang telah melewati tahap preprocessing adalah data yang telah siap dianalisa ke tahap selanjutnya. Perbandingan data ditunjukkan pada Gambar 13.

review	Rating	Cleaned_Review
We chose Jennie Subaru to help set up a 3D2N h...	5	we chose jennie subaru to help set up a 3d2n h...
especially since by choosing him you also con...	5	especially since by choosing him you also cont...
Jenie is making a huge effort to invest in bu...	5	jenie is making a huge effort to invest in buy...
We booked our trip through Aria Prasetio from...	5	we booked our trip through aria prasetio from ...
Anwar was our guide	5	anwar was our guide

Gambar. 13. Perbandingan Review Text Setelah Data Pre-Processing

1) *Splitting* Data Testing dan data Training

Pemisahan data training dan data testing yang digunakan pada penelitian ini adalah 30:70 dari jumlah dataset. 30% atau 901 data training dan 70% atau 2.105 data testing.

2) *Feature Extraction*

Data training dan data testing selanjutnya dipisah menjadi data teks dan numerik untuk mengukur performa algoritma apakah akan lebih baik jika data berupa teks atau numerik. Adapun data training dan data testing berupa teks sebelumnya dilakukan tokenisasi dengan metode *CountVectorizer*. Gambar 14 merupakan data training dan testing berupa teks, sedangkan Tabel III merupakan data training dan data testing berupa numerik.

```
3001 namun tidak perlu khawatir akan diserang oleh ...
3002 mereka makan dan bermalam diatas kapal di dala...
3003 realmente yo era bastante reacia a pasar dos d...
3004 reconozco que para la segunda ya estaba habitu...
3005 obviamente más enriquecedor el contacto con la...
```

Gambar. 14. Data Teks

TABEL III DATA NUMERIK

Kata	Numerik	Kata	Numerik
orangutan	2161	Viele	3272
klotok	1626	Hasta	1353
guide	1302	traditional	3125
jungle	1578	Trek	3146
nature	2015	experiencing	1096
amazing	175	Drinks	892
pero	2312	Snacks	2822

3) *Truncate singular-value decomposition* (TSVD)

Penggunaan TSVD berfungsi mengurangi waktu dan ruang penyimpanan dengan mengubah dimensi yang lebih tinggi menjadi lebih rendah. Analisis sentimen yang hemat waktu dan biaya melalui pengurangan dimensi. Pada penelitian ini penggunaan TSVD pada Gambar 15.

```
from sklearn.decomposition import TruncatedSVD

tsvd = TruncatedSVD(n_components = 20)
x_train = tsvd.fit_transform(X_train)
x_test = tsvd.fit_transform(X_test)
```

Gambar. 15. Implementasi TSVD

Gambar 15 terdapat *n_components* yang berfungsi menentukan jumlah komponen yang ingin digunakan dalam pengurangan dimensi matriks BoW (Bag of Words) yang telah dilakukan pada tahap sebelumnya *Feature Extraction*.

```
x_train.shape
(901, 20)

x_test.shape
(2105, 20)
```

Gambar. 16. Matriks TSVD

Gambar 16 menunjukkan bahwa matriks TSVD di mana data training memiliki 901 baris dan 20 kolom serta data test 2.105 baris dan 20 kolom. Artinya, setelah proses TSVD, jumlah fitur yang digunakan dalam klasifikasi/analisis terkait teks dikurangi menjadi 20 fitur saja.

E. *Training and Evaluation*

Pada metode SVM, kernel yang digunakan yaitu *Linear*. Dalam pengujian ini akan menggunakan perbandingan dataset 30:70. 30% atau 901 data training dan 70% atau 2.105 data testing. Selanjutnya nilai *rating* 1 hingga 5 dibuat lima kategori kelas untuk memberlakukan prediksi *rating* sebagai *multi-class classification*. Model SVM menunjukkan akurasi sebesar 80%.

Confusion matrix adalah metrik penting yang digunakan untuk mengukur efisiensi pengklasifikasi. Metrik ini mewakili jumlah sampel yang diklasifikasikan dengan benar dan digunakan untuk menentukan *precision*, *recall*, dan *F1 Score*. *Confusion matrix* ditunjukkan pada Tabel IV.

TABEL IV CONFUSION METRIX

Model	Accuracy	Precision	Recall	F1 Score
SVM	80%	81%	99%	89%

Tabel IV menunjukkan tingkat akurasi SVM yang dihasilkan adalah 80%, lebih besar dibandingkan dengan penelitian sebelumnya oleh Imamah[8] yang menggunakan SVM dengan data ulasan TripAdvisor pada Pariwisata Kabupaten Bangkalan dengan tingkat akurasi 70.22%, namun lebih kecil dari hasil penelitian oleh Widodo dkk [6] dengan data ulasan TripAdvisor Borobudur yaitu 100%.

IV. KESIMPULAN

Berdasarkan evaluasi, penerapan kombinasi *Feature extraction*, *Truncate singular-value decomposition* (TSVD) dan *Support Vector Machine* (SVM) dengan kernel *Linear* dapat memberikan tingkat akurasi dengan rata-rata 80%, *precision* = 81%, *recall* = 99%, dan *F1-score* = 89%. Visualisasi data dengan *Wordcloud* top 10 ulasan masing-masing *rating* mampu memberikan pengetahuan terhadap sentimen dari ulasan wisatawan.

Dengan demikian kontribusi penelitian ini terhadap perkembangan ilmu pengetahuan menekankan pada intensitas kajian analisis sentimen dalam industri pariwisata. Data ulasan pengguna situs TripAdvisor merupakan informasi yang tidak dapat diabaikan, sebagai respons terhadap era digital yang disruptif. Data *review* wisatawan objek wisata dapat meningkatkan strategi pemasaran, termasuk upaya optimalisasi layanan dan produk makanan dan minuman di objek wisata. Penelitian ini dapat menguraikan persepsi wisatawan yang merupakan bagian penting dari industri pariwisata Indonesia khususnya Kabupaten Kotawaringin Barat, Provinsi Kalimantan Tengah.

DAFTAR PUSTAKA

- [1] S. W. Lee, G. Jiang, H. Y. Kong, and C. Liu, "A difference of multimedia consumer's rating and review through sentiment analysis," *Multimed Tools Appl*, vol. 80, no. 26–27, pp. 34625–34642, Nov. 2021, doi: 10.1007/s11042-020-08820-x.
- [2] M. M. Agüero-Torales, M. J. Cobo, E. Herrera-Viedma, and A. G. López-Herrera, "A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 392–399. doi: 10.1016/j.procs.2019.12.002.
- [3] K. Puh and M. Bağıc Babac, "Predicting sentiment and rating of tourist reviews using machine learning," *Journal of Hospitality and Tourism Insights*, vol. 6, no. 3, pp. 1188–1204, Jun. 2023, doi: 10.1108/JHTI-02-2022-0078.
- [4] B. H. Ahmed and A. S. Ghabayen, "Review rating prediction framework using deep learning," *J Ambient Intell Humaniz Comput*, vol. 13, no. 7, pp. 3423–3432, Jul. 2022, doi: 10.1007/s12652-020-01807-4.
- [5] "Sentiment Analysis of OYO App Reviews Using the Support Vector Machine Algorithm Analisis Sentimen terhadap Ulasan Aplikasi OYO menggunakan Algoritma Support Vector Machine Zaenal, Ika Ratna Indra Astutik."
- [6] E. Widodo, G. D. Nursyafitri, O. N. Putri, and Y. Nurmalarari, "Sentiment Analysis of Visitors in Borobudur Temple With Support Vector Machine Based on TripAdvisor Review," 2020.
- [7] W. Chen, Z. Xu, X. Zheng, Q. Yu, and Y. Luo, "Research on sentiment classification of online travel review text," *Applied Sciences (Switzerland)*, vol. 10, no. 15, Aug. 2020, doi: 10.3390/AP10155275.
- [8] Imamah *et al.*, "Text Mining and Support Vector Machine for Sentiment Analysis of Tourist Reviews in Bangkalan Regency," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, 2020. doi: 10.1088/1742-6596/1477/2/022023.
- [9] A. A. Ningtyas, A. Solichin, and R. Pradana, "ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG PREDIKSI RESESI EKONOMI TAHUN 2023 MENGGUNAKAN ALGORITME NAÏVE BAYES," 2023.
- [10] A. Nugraha, Y. H. Chrisnanto, and R. Yuniarti, *Seminar Nasional Sains & Teknologi Informasi (SENSASI) Prediksi Sentimen Pada Sosial Media Twitter Mengenal Produk Smartphone Menggunakan Algoritma K-NN Classification*. [Online]. Available: <http://prosiding.seminar-id.com/index.php/sensasi/issue/archivePage|251>
- [11] S. A. Amira and M. I. Irawan, "Opinion Analysis of Traveler Based on Tourism Site Review Using Sentiment Analysis," *IPTEK The Journal for Technology and Science*, vol. 31, no. 2, p. 223, May 2020, doi: 10.12962/j20882033.v31i2.6338.
- [12] R. A. E. Virgana, T. Sapanji, D. Hamdani, and P. Harahap, "Sentiment Analysis of the Top 5 E-commerce Platforms in Indonesia using Text Mining and Natural Language Processing (NLP)," 2023. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [13] R. Wahyudi *et al.*, "Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *JURNAL INFORMATIKA*, vol. 8, no. 2, 2021, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [14] A. I. KABIR, K. AHMED, and R. KARIM, "Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language," *Informatica Economica*, vol. 24, no. 4/2020, pp. 55–71, Dec. 2020, doi: 10.24818/issn14531305/24.4.2020.05.
- [15] T. R. Sanusi, F. Andreas, B. N. Sari, U. Singaperbangsa, and K. Abstract, "Implementasi Algoritma Support Vector Classifier (SVC) dengan Data Training Numerik dan Teks untuk Mengklasifikasi SMS Spam," *Jurnal Ilmiah Wahana Pendidikan*, vol. 2022, no. 14, pp. 346–354, doi: 10.5281/zenodo.6994895.
- [16] J. Barry, "Sentiment Analysis of Online Reviews Using Bag-of-Words and LSTM Approaches." [Online]. Available: <https://www.yelp.com/dataset/challenge>
- [17] G. Setiawan, H. Novianus Palit, and E. Setyati, "Aspect Based Sentiment Analysis pada Layanan Umpan Balik Universitas dengan Menggunakan Metode Naïve Bayes dan Latent Semantic Analysis." [Online]. Available: www.kateglo.com
- [18] J. Khairnar and M. Kinikar, "Sentiment Analysis Based Mining and Summarizing Using SVM-MapReduce." [Online]. Available: <http://www.cs.comell.edu/people/pabo/movie-review->
- [19] H. Duong Thi, K. Hoang Manh, V. Trinh Anh, T. Pham Thi Quynh, and T. Nguyen Viet, "Dimensionality Reduction with Truncated Singular Value Decomposition and K-Nearest Neighbors Regression for Indoor Localization." [Online]. Available: www.ijacsa.thesai.org
- [20] E. K. Andana, M. Othman, and R. Ibrahim, "Comparative analysis of text classification using naive bayes and support vector machine in detecting negative content in Indonesian twitter," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 1.3 S1, pp. 356–362, 2019, doi: 10.30534/ijatcse/2019/6481.32019.
- [21] F. Nufairi, N. Pratiwi, and F. Herlando, "JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) Journal homepage: <https://jurnal.stkipgritulungagung.ac.id/index.php/jipi> ANALISIS SENTIMEN PADA ULASAN APLIKASI THREADS DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE," vol. 9, no. 1, pp. 339–348, 2024, doi: 10.29100/jipi.v9i1.4929.

- [22] M. P. Dwi Cahyo, Widodo, and B. Prasetya Adhi, “Kinerja Algoritma Support Vector Machine dalam Menentukan Kebenaran Informasi Banjir di Twitter,” *PINTER : Jurnal Pendidikan Teknik Informatika dan Komputer*, vol. 3, no. 2, pp. 116–121, Dec. 2019, doi: 10.21009/pinter.3.2.5.