

Analisis Perilaku Wisatawan Berdasarkan Data Ulasan di Website Tripadvisor Menggunakan CRISP-DM: Wisata Minat Khusus Pendakian Gunung Rinjani dan Gunung Bromo

Yerik Afrianto Singgalen

Fakultas Ilmu Administrasi Bisnis dan Ilmu Komunikasi, Program Studi Pariwisata, Universitas Katolik Indonesia Atma Jaya, Jakarta, Indonesia

Email:yerik.afrianto@atmajaya.ac.id

Email Penulis Korespondensi: yerik.afrianto@atmajaya.ac.id

Submitted: 29/01/2023; Accepted: 22/02/2023; Published: 25/02/2023

Abstrak—Perilaku wisatawan perlu diidentifikasi dan dianalisis secara komprehensif untuk mengetahui perubahan dalam hal preferensi jenis wisata di Indonesia. Salah satu pendekatan yang relevan dengan upaya mengidentifikasi perilaku wisatawan ialah analisis sentimen melalui data ulasan di website tripadvisor melalui pendekatan teks mining. Penelitian ini bertujuan merekomendasikan model analisis sentimen yang bermanfaat bagi pengelola destinasi wisata minat khusus pendakian di gunung Rinjani dan gunung Bromo berdasarkan tipe kunjungan sendiri (solo), bersama pasangan (couple), bersama teman (friends) dan bersama keluarga (family). Metode penelitian yang digunakan ialah Cross Industry Standard Process for Data Mining (CRISP-DM) dengan algoritma yang disesuaikan dengan konteks pengelolaan destinasi wisata Gunung Rinjani dan Gunung Bromo yakni atraksi, jalan dan moda transportasi (aksesibilitas), dan akomodasi. Dalam upaya mengatasi permasalahan keseimbangan data dalam dataset, maka proses kalkulasi algoritma Decision Tree (DT) dan Support Vector Machine (SVM) dihubungkan dengan operator Synthetic Minority Over-sampling Technique (SMOTE) pada aplikasi Rapidminer. Hasil penelitian ini menunjukkan bahwa algoritma SVM menunjukkan performa yang lebih baik dengan nilai akurasi 97,67%, presisi 100%, recall 95,34 dan f-measure 97,61% dalam klasifikasi 1075 data teks Gunung Bromo dan 326 data ulasan Gunung Rinjani. Selain itu, dalam konteks gunung Rinjani, lima kata teratas yang sering muncul dalam data ulasan wisatawan di gunung Rinjani ialah sebagai berikut : summit (272), rinjani (259), trek (201), hike (170), mountain (159). Disisi lain, lima kata teratas yang sering muncul dalam data ulasan wisatawan di gunung Bromo ialah sebagai berikut : bromo (1864), sunrise (1124), view (854), crater (758), mount (577). Dengan demikian dapat diketahui bahwa wisatawan dengan tipe kunjungan sendiri (solo), bersama pasangan (couple), bersama teman (friends) dan bersama keluarga (family) memiliki preferensi terhadap jenis atraksi berupa puncak, kawah, keindahan alam pengunungan, jalur pendakian, jenis moda transportasi (aksesibilitas), serta akomodasi pendukung yang perlu disiapkan agar dapat mendukung keberlanjutan pariwisata.

Kata Kunci: Perilaku Wisatawan; Minat Khusus; Analisis Sentimen; Bromo; Rinjani

Abstract—Traveller behavior needs to be comprehensively identified and analyzed to determine changes in tourism preference in Indonesia. One relevant approach to identifying traveler behavior is sentiment analysis through review data on the Tripadvisor website by text mining approach. This study aims to recommend a sentiment analysis model that is useful for managers of particular interest climbing tourist destinations on Mount Rinjani and Mount Bromo based on the type of visit alone (solo), with couples (couple), with friends, and with family (family). The research method used is Cross Industry Standard Process for Data Mining (CRISP-DM), with an algorithm adapted to managing tourist destinations for Mount Rinjani and Mount Bromo, namely attractions, roads and modes of transportation (accessibility), and accommodation. To overcome the problem of data balance in datasets, the calculation process of the Decision Tree (DT) algorithm and the Support Vector Machine (SVM) is connected to the Synthetic Minority Over-sampling Technique (SMOTE) operator in the Rapidminer application. The results of this study showed that the SVM algorithm showed better performance with an accuracy value of 97.67%, precision of 100%, recall of 95.34, and f-measure of 97.61% in the classification of 1075 text data of Mount Bromo and 326 review data of Mount Rinjani. In addition, in the context of Mount Rinjani, the top five words that often appear in tourist review data on Mount Rinjani are as follows: summit (272), Rinjani (259), trek (201), hike (170), mountain (159). On the other hand, the top five words that often appear in tourist review data on Mount Bromo are as follows: Bromo (1864), sunrise (1124), view (854), crater (758), and mount (577). Thus, it can be seen that tourists with the type of visit alone (solo), with couples (couple), with friends (friends), and with family (family) have a preference for the types of attractions in the form of summits, craters, natural beauty of mountains, hiking trails, types of transportation modes as well as supporting accommodation that needs to be prepared to keep the sustainability of tourism.

Keywords: Tourist Behavior; Special Interest; Sentiment Analysis; Bromo; Rinjani

1. PENDAHULUAN

Wisata minat khusus untuk eksplorasi hingga pendakian Gunung Bromo dan Gunung Rinjani menjadi populer di kalangan wisatawan dengan tipe kunjungan solo, couple, friends, dan family. Hal tersebut dapat dilihat dari jumlah data ulasan di website Tripadvisor sebagai platform digital yang dapat digunakan untuk meninjau lokasi wisata di berbagai negara. Azzahra dan Wibowo berpendapat bahwa wisatawan seringkali mencari informasi tentang destinasi wisata di Website Tripadvisor sehingga emoticon, rating maupun data teks di kolom ulasan digunakan sebagai dataset untuk analisis preferensi wisatawan [1]. Disisi lain, beberapa studi terdahulu terkait dengan analisis sentimen wisatawan terhadap produk dan layanan di destinasi wisata, juga menggunakan data teks dari kolom ulasan pengguna website Tripadvisor [2]–[4]. Hal ini menunjukkan bahwa website Tripadvisor menyediakan data yang dapat digunakan untuk mengidentifikasi dan menganalisis perilaku wisatawan khususnya permintaan atau

preferensi terhadap produk dan layanan wisata, melalui klasifikasi data teks menjadi sentimen positif dan negatif. Dengan demikian, penelitian ini bertujuan untuk menganalisis perilaku wisatawan melalui data teks dalam kolom ulasan di destinasi wisata minat khusus Gunung Rinjani dan Gunung Bromo, pada website Tripadvisor.

Penelitian ini menggunakan metode Cross Industry Standar Process for Data Mining (CRISP-DM) yang terdiri dari enam tahapan yakni tahap business understanding, tahap data understanding, tahap data preparation, tahap modeling, tahap evaluation dan tahap deployment. Fattah et al. menunjukkan bahwa pada tahap business understanding, karakteristik institusi yang mengolah data perlu dianalisis secara komprehensif untuk memperoleh gambaran tentang proses bisnis, sedangkan data understanding merupakan proses eksplorasi dan analisis data berdasarkan tujuan dari pengolahan data untuk pengembangan organisasi maupun perusahaan [5]. Selain itu, Wulandari et al. menunjukkan bahwa pada tahap data preparation, semua data akan dirapikan dan dibersihkan sedangkan pada tahap modeling, dapat dilakukan kalkulasi menggunakan algoritma tertentu sesuai dengan tujuan pengolahan data [6]. Adapun, Christian dan Qi menjelaskan bahwa pada tahap evaluation, performa dari model atau algoritma yang digunakan akan ditinjau dan dianalisis, sedangkan pada tahap deployment, hasil identifikasi dan analisis model dapat direkomendasikan kepada pemangku kepentingan untuk mengoptimalkan kinerja dari organisasi atau perusahaan [7]. Hal ini menunjukkan bahwa pendekatan CRISP-DM menjadi relevan digunakan dalam penelitian ini, dimana hasil analisis perilaku wisatawan dapat dinarasikan dalam bentuk rekomendasi pengembangan destinasi wisata minat khusus pendakian. Dengan demikian, metode CRISP-DM dapat digunakan untuk menghasilkan rekomendasi pengembangan destinasi yang kontekstual sesuai dengan preferensi wisatawan terkait wisata minat khusus pendakian Gunung Rinjani dan Gunung Bromo di Indonesia.

Model analisis sentimen wisatawan yang dapat digunakan untuk klasifikasi sentimen berdasarkan data teks ialah Decision Tree (DT) dan Support Vector Machine (SVM). Munawwaroh dan Primandani menunjukkan bahwa analisis sentimen dan klasifikasi data menggunakan algoritma DT dalam kerangka kerja CRISP-DM memberikan gambaran yang jelas hingga tentang proses bisnis hingga tujuan pengolahan data di suatu perusahaan atau organisasi [8]. Selain itu, Cahyaningtias et al. menunjukkan bahwa implementasi analisis sentimen menggunakan algoritma DT dalam beberapa studi kasus menunjukkan performa yang baik ditinjau dari nilai akurasi, presisi, recall, dan Area Under Curve (AUC) [9]. Adapun, Rousyanti menunjukkan bahwa model pengolahan data menggunakan algoritma SVM juga dapat dibandingkan dengan DT agar dapat diketahui algoritma dengan performa yang terbaik, sehingga relevan digunakan dalam pengembangan produk maupun manajemen organisasi atau perusahaan [10]. Singgalen menunjukkan bahwa performa algoritma Decision Tree (DT), Support Vector Machine (SVM), k-Nearest Neighbour (k-NN), Naïve Bayes Classifier (NBC) [11]. Hal ini menunjukkan bahwa algoritma dengan performa terbaik dapat digunakan sebagai model untuk rekomendasi strategi peningkatan kualitas produk dan layanan destinasi wisata, maupun strategi mengoptimalkan sistem manajemen destinasi wisata minat khusus Gunung Bromo dan Gunung Rinjani. Dengan demikian, penelitian ini mengevaluasi algoritma Decision Tree (DT) dan Support Vector Machine (SVM) sehingga dapat diperoleh algoritma dengan performa terbaik dalam analisis sentimen wisatawan.

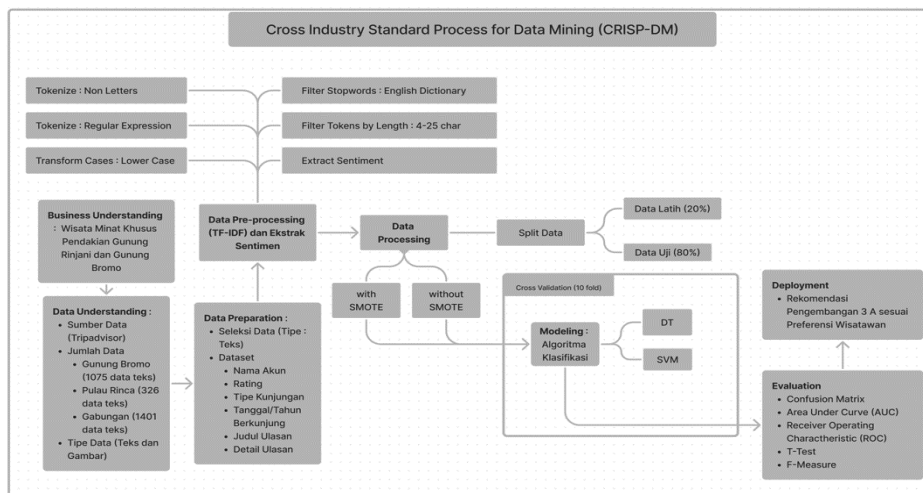
Pengetahuan, sikap dan perilaku wisatawan di destinasi wisata tidak terlepas dari kebutuhan akan suatu kondisi yang diharapkan oleh individu atau kelompok. Despica dan Zuriyani berpendapat bahwa kondisi yang diharapkan wisatawan berhubungan dengan sapta pesona yakni aman, tertib, bersih, sejuk, indah, ramah, dan kenangan [12]. Selanjutnya, Putri et al. menunjukkan bahwa wisatawan cenderung mencari informasi yang berhubungan dengan upaya memenuhi kebutuhan akan kondisi yang diharapkan, ketika merencanakan perjalanan wisata baik sendiri, bersama pasangan, teman, maupun keluarga [13]. Adapun, Sarudin dan Ismail menunjukkan bahwa website Tripadvisor merupakan salah satu platform digital yang diakses untuk memperoleh informasi tentang atraksi, akomodasi, hingga jenis moda transportasi yang dapat digunakan selama berwisata [14]. Hal ini menunjukkan adanya perilaku untuk mengakumulasi pengetahuan terkait 3 A (atraksi, aksesibilitas dan akomodasi) serta sikap membandingkan informasi 3A sebelum menentukan pilihan berwisata dengan menjadikan data ulasan di website Tripadvisor sebagai referensi. Dengan demikian, penelitian ini fokus menganalisis perilaku wisatawan berdasarkan data ulasan di website Tripadvisor.

Penelitian ini memiliki peluang berkontribusi di bidang pariwisata dan sistem informasi melalui model analisis perilaku wisatawan berbasis CRISP-DM. Beberapa kajian terdahulu tentang perilaku wisatawan menggunakan perspektif psiko-sosial, sosio-ekonomi, dan sosio-kultural. Seperti halnya, Abadi dan Herwin yang menganalisis perilaku wisatawan dari perspektif sosio-ekonomi untuk mengidentifikasi pengaruh kualitas produk dan layanan terhadap kepuasan wisatawan [15]. Disisi lain, Pomantow et al. berpendapat bahwa perilaku wisatawan dapat diamati secara spesifik dari latarbelakang budaya, keluarga, usia dan siklus hidup [16]. Adapun, Lubis et al. menunjukkan bahwa perilaku wisatawan sangat penting untuk diidentifikasi dan dianalisis secara kontekstual mengingat diversitas destinasi wisata dan faktor-faktor yang memengaruhi keputusan wisatawan berkunjung kembali ke suatu destinasi wisata [17]. Era digital menjadi stimulus perubahan perilaku pengguna teknologi informasi. Setiap orang dapat mengakses dan memanfaatkan informasi untuk memenuhi kebutuhannya. Hal ini menunjukkan bahwa analisis perilaku wisatawan dari perspektif sistem informasi juga penting untuk dikaji secara komprehensif, terutama dalam konteks pembahasan wisata minat khusus di Indonesia. Dengan demikian, penelitian ini dapat berkontribusi secara teoretis di bidang pariwisata dan sistem informasi.

2. METODOLOGI PENELITIAN

2.1 Cross-Industry Standard Process for Data Mining (CRISP-DM)

Metode yang digunakan dalam analisis perilaku wisatawan berdasarkan data ulasan website Tripadvisor untuk wisata minat khusus pendakian Gunung Rinjani dan Gunung Bromo ialah Cross Industry Standard Process for Data Mining (CRISP-DM) yang terbagi menjadi enam tahapan sebagai berikut business understanding, data understanding, data preparation, modeling, evaluation dan deployment. Pada tahap business understanding, dilakukan studi literatur tentang pengetahuan, sikap dan perilaku wisatawan dalam mengakses website Tripadvisor untuk merencanakan perjalanan wisata, khususnya wisata minat khusus pendakian Gunung Bromo dan Gunung Rinjani. Selanjutnya, pada tahap data understanding, dilakukan pengujian sistem validasi data ulasan oleh member website Tripadvisor, screening data ulasan berdasarkan nama akun, rating, tipe kunjungan, tanggal berkunjung, judul ulasan, dan detail ulasan. Adapun, proses scraping data teks dari website Tripadvisor menggunakan aplikasi webharvy. Pada tahap data preparation, dilakukan proses seleksi data yang akan digunakan dalam penelitian ini yakni data terkait dengan rating, tipe kunjungan, tahun berkunjung, dan ulasan. Data ulasan dalam bentuk teks juga dirapikan menggunakan sejumlah operator di aplikasi Rapidminer (tokenize, transform cases, filter stopwords, filter tokens by length); pada tahap data processing, dilakukan pembagian data latih (20%) dan data uji (80%) dari 1075 data teks Gunung Bromo dan 326 data teks Gunung Rinjani. Selanjutnya, untuk mengatasi masalah ketidakseimbangan data, digunakan operator SMOTE Upsampling. Agar dapat dianalisis secara komprehensif, maka pengujian juga dilakukan dengan dan tanpa operator SMOTE Upsampling sehingga dapat diketahui perbedaan nilai Area Under Curve (AUC) dari algoritma yang digunakan; pada tahap modeling, dilakukan klasifikasi dataset menggunakan operator cross validation pada algoritma Decision Tree (DT) dan Support Vector Machine (SVM); pada tahap evaluation, dilakukan analisis nilai confusion matrix, f-measure, Area Under Curve (AUC), Receiver Operating Characteristic (ROC), dan t-Test; pada tahap penerapan (deployment), hasil pengolahan data dapat direkomendasikan sebagai model analisis perilaku wisatawan sebagai rekomendasi untuk mengoptimalkan produk dan layanan wisata minat khusus pendakian di Gunung Bromo dan Gunung Rinjani, agar dapat mengoptimalkan atraksi, akomodasi dan aksesibilitas pendukung sehingga meningkatkan intensi wisatawan untuk berkunjung kembali. Adapun, alur dari keseluruhan tahapan dapat dilihat pada gambar 1 berikut ini.



Gambar 1. Tahapan Penelitian sesuai CRISP-DM Framework

Gambar 1 merupakan keseluruhan tahapan penelitian berdasarkan kerangka kerja CRISP-DM dalam menganalisis perilaku wisatawan sebagai rekomendasi untuk mengoptimalkan produk dan layanan wisata minat khusus pendakian di Gunung Bromo dan Gunung Rinjani, agar dapat mengoptimalkan atraksi, akomodasi dan aksesibilitas pendukung sehingga meningkatkan intensi wisatawan untuk berkunjung kembali. Adapun, algoritma yang digunakan ialah Decision Tree (DT) dan Support Vector Machine (SVM). Masing-masing algoritma memiliki keunggulan dan kelemahan sehingga perlu disesuaikan dengan karakteristik data yang akan dikelola untuk menghasilkan model yang sesuai. Hasil evaluasi algoritma dengan performa terbaik menunjukkan bahwa data tersebut dapat digunakan sebagai dasar untuk menghasilkan rekomendasi bagi pengelola destinasi wisata minat khusus pendakian di Gunung Bromo dan Gunung Rinjani, Indonesia.

2.2 Term Frequency-Inverse Document Frequency (TF-IDF)

Pada tahap data understanding dan data preparation, proses seleksi diperlukan untuk membersihkan dan merapikan data ulasan yang telah dikumpulkan. Selanjutnya, pembobotan kata diperlukan untuk memperoleh informasi tentang jumlah kata yang paling sering muncul dalam data ulasan. Pembobotan kata merupakan proses pemberian

nilai pada setiap kata yang telah melewati tahap Pre-Processing. Penelitian ini menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) dengan persamaan (1) dan (2) berikut :

$$\text{IDF}(w) = \log\left(\frac{N}{\text{DF}(w)}\right) \quad (1)$$

$$W_{ij} = \text{TF}_{ij} \times \log\left(\frac{D}{\text{DF}_j}\right) \quad (2)$$

Kosasih dan Alberto menunjukkan bahwa proses pembobotan kata dapat dilakukan dengan beberapa tahapan sebagai berikut : pertama, menghitung jumlah Term Frequency (TF) tiap kata, dimana kalimat yang telah dipisah menjadi kata akan diberi nilai dan setiap kata yang muncul akan diberi nilai 1; kedua, menghitung jumlah Document Frequency (DF) tiap kata dengan cara menjumlahkan nilai TF pada tiap kata; ketiga, menghitung jumlah Inverse Document Frequency (IDF) yang ditunjukkan pada persamaan (1); keempat, menghitung bobot (Weight) pada tiap kata yang diperoleh dari hasil perkalian nilai TF dengan IDF sebagaimana persamaan (2) [18]. Selanjutnya, hasil pembobotan kata akan digunakan sebagai input untuk algoritma Decision Tree (DT) dan Support Vector Machine (SVM).

2.3 Synthetic Minority Oversampling Technique (SMOTE)

Penelitian ini menggunakan operator SMOTE Upsampling di aplikasi Rapidminer untuk mengatasi masalah ketidakseimbangan data, terutama pada dataset Gunung Rinjani. Ketidakseimbangan data akan terjadi apabila jumlah objek di suatu kelas data memiliki kuantitas yang lebih tinggi dibandingkan dengan kelas lain, dimana kelas data yang objeknya lebih banyak disebut kelas mayor sedangkan yang lain disebut minor. Disisi lain, Kurniawati menekankan bahwa pengolahan algoritma yang tidak mempertimbangkan ketidakseimbangan data cenderung menitikberatkan kelas mayor dan bukan kelas minor, oleh sebab itu diperlukan teknik SMOTE yang menggunakan metode oversampling untuk memperbanyak pengamatan secara acak dengan menambah jumlah data kelas minor (data buatan) agar setara dengan kelas mayor [19]. Adapun, data buatan atau sintesis tersebut dibuat berdasarkan k-tetangga terdekat (k-Nearest Neighbor). Pembangkit data buatan yang berskala numerik diukur jarak kedekatannya dengan jarak euclidean sedangkan data kategorik berdasarkan kelas minor yang peubahnya berskala kategorik, dilakukan dengan rumus Value Difference Metric (VDM) yaitu :

$$\Delta(x, y) = w_x w_y \sum_{i=1}^N \delta(x_i y_i)^r \quad (3)$$

Persamaan (3) merupakan proses untuk membangkitkan data numerik. Dimana $\Delta(x, y)$ adalah jarak antara amatan x dengan y , sementara $w_x w_y$ merupakan bobot amatan (dapat diabaikan), N merupakan banyaknya pebuah penjelas, r bernilai 1 (jarak manhattan) atau 2 (jarak euclidean), serta $\delta(x_i y_i)^r$ jarak antar kategori. Adapun, proses pembangkit data buatan (sintesis) untuk data numerik dilakukan dengan menghitung perbedaan antar vektor utama dengan k-tetangga terdekatnya, kalikan perbedaan dengan angka yang diacak diantara 0 dan 1, kemudian tambahkan perbedaan tersebut ke dalam nilai utama pada vektor utama asal sehingga diperoleh vektor utama yang baru. Selanjutnya, pembangkit data kategorik dapat dilakukan melalui persamaan (4) sebagai berikut.

$$\delta(V_1 V_2) = \sum_{i=1}^n \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right|^k \quad (4)$$

Dimana, $\delta(V_1 V_2)$ merupakan jarak antara nilai V_1 dan V_2 sedangkan C_{1i} merupakan banyaknya V_1 yang termasuk kelas I, dan C_{2i} merupakan banyaknya V_2 yang termasuk kelas I. Sementara itu, i merupakan banyaknya kelas, C_1 banyaknya nilai 1 terjadi, C_2 banyaknya nilai 2 terjadi, n merupakan banyaknya kategori, dan k merupakan konstansi. Proses pembangkitan data buatan (sintesis) untuk data kategori dilakukan dengan memilih mayoritas antara vektor utama yang dipertimbangkan dengan k-tetangga terdekatnya untuk nilai nominal, jika nilai sama maka akan dipilih secara acak. Selanjutnya, nilai tersebut dijadikan data contoh kelas buatan yang baru. Dengan operator SMOTE, evaluasi algoritma DT dan SVM akan menunjukkan performa yang lebih optimal.

2.4 Algoritma Decision Tree

Pengujian pertama dalam memperoleh model yang tepat untuk menghasilkan rekomendasi bagi pengelola destinasi wisata minat khusus pendakian ke Gunung Rinjani dan Gunung Bromo ialah Decision Tree (DT). Algoritma Decision Tree (DT) merupakan salah satu metode klasifikasi populer yang sering digunakan secara praktis, model prediksi yang menggunakan struktur pohon untuk mencari dan membuat keputusan, serta memecahkan masalah dengan mempertimbangkan berbagai faktor di dalam lingkup masalah tersebut [20]. Decision Tree memiliki beberapa algoritma salah satunya Iterative Dichotomizer version (ID3), yaitu model klasifikasi yang berupa pohon keputusan secara top-down dengan cara kerja mengevaluasi semua atribut menggunakan suatu ukuran statistik berupa information gain untuk mengukur efektifitas suatu atribut dalam mengklasifikasi sample data. Dalam algoritma ini, dibutuhkan nilai entropy dan gain, dimana entropy merupakan parameter untuk mengukur jumlah keberagaman atau keberadaan dalam sebuah himpunan data, sedangkan gain merupakan perolehan informasi sebagai ukuran efektifitas suatu atribut. Berikut adalah persamaan untuk mendapatkan nilai entropy dan gain.

$$\text{Entropy (S)} = \sum_{i=1}^n -p \times \log_2 p_i \quad (5)$$

$$\text{Gain (S, A)} = S - \sum_{i=1}^n \frac{|S_i|}{|S|} \times S_i \quad (6)$$

Dimana S merupakan nilai Entropy, p_i jumlah yang memiliki nilai positif atau negatif pada kumpulan data untuk sifat tertentu. Disisi lain, Gain (S,A) adalah hasil informasi yang berasal dari luaran data yang dikelompokkan sesuai dengan atribut A. Selanjutnya, S_i adalah subset dari nilai entropy yang mempunyai nilai i . Adapun, S adalah subset dari nilai Entropy. Puspita & Widodo berpendapat bahwa algoritma Decision Tree memiliki keunggulan dan kekurangan, dimana keunggulannya ialah konsep yang jelas dan mudah dipahami serta diimplementasikan menggunakan algoritma rekursif, sedangkan kekurangannya ialah tidak dapat diaplikasikan pada himpunan data yang sangat besar dan mudah mengalami overfit karena proses pelatihan greedy [21]. Dengan demikian perlu dilakukan pengujian algoritma SVM untuk membandingkan performa algoritma yang lebih baik.

2.5 Algoritma Support Vector Machine

Pengujian kedua dalam penelitian ini ialah dengan mengklasifikasikan dataset menggunakan algoritma Support Vector Machine (SVM). Algoritma Support Vector Machine (SVM) merupakan salah satu algoritma yang dapat digunakan untuk klasifikasi data menggunakan hyperplane [22]. Karim menjelaskan bahwa konsep SVM menitikberatkan pada risk minimization, yaitu estimasi fungsi dengan cara meminimalisir batas dari generalization error, sehingga SVM mampu mengatasi overfitting [23]. Adapun, fungsi regresi dari metode SVM adalah sebagai berikut.

$$f(x) = w^T \varphi(x) + b \quad (7)$$

Dimana w merupakan vector pembobot, $\varphi(x)$ merupakan sebuah fungsi yang memetakan x ke dalam suatu dimensi, dan b merupakan faktor bias. Selanjutnya, Nida et al., menunjukkan bahwa SVM memiliki kelebihan dalam generalisasi data yang tinggi serta mampu menghasilkan model klasifikasi yang baik meskipun dilatih dengan data yang relatif sedikit. Meskipun demikian, sangat sulit diaplikasikan untuk himpunan data dengan sampel dan dimensi yang besar [24]. Hal ini menunjukkan bahwa SVM mampu menghasilkan performa yang baik meskipun dengan jumlah data yang relatif sedikit.

2.6 Evaluasi Klasifikasi (Confusion Matrix)

Pada tahap evaluasi, algoritma DT dan SVM dievaluasi berdasarkan nilai akurasi, presisi, recall, dan nilai f-measure. Kasanah et al. menunjukkan bahwa Evaluasi klasifikasi didasarkan pada pengujian pada objek yang benar dan objek yang salah. Validasi digunakan untuk menentukan jenis model yang terbaik melalui confusion matrix sebagai informasi mengenai hasil klasifikasi actual yang dapat diprediksi oleh suatu sistem melalui nilai akurasi, presisi, dan recall, melalui persamaan berikut.

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

$$\text{Presisi/Specificity} = \frac{TP}{TP+FP} \quad (9)$$

$$\text{Recall/Sensitivity} = \frac{TP}{TP+FN} \quad (10)$$

$$f\text{-measure} = \frac{2 \times (\text{Presisi} \times \text{recall})}{\text{presisi} + \text{recall}} \quad (11)$$

Hairani et al. memberikan penjelasan tentang confusion matrix dari akurasi, presisi, recall dan f-measure : akurasi adalah ketepatan sistem dalam melakukan proses klasifikasi dengan benar; presisi atau sensitivity adalah rasio jumlah dokumen yang relevan dengan total jumlah dokumen yang ditemukan pada sistem klasifikasi; recall atau specificity adalah rasio jumlah dokumen yang ditemukan kembali oleh sistem klasifikasi dengan total jumlah dokumen yang relevan; f-measure adalah metrik evaluasi yang populer untuk menangani masalah imbalance class dengan mengombinasikan recall/sensivitas dan presisi sehingga menghasilkan metrik yang efektif untuk mencari kembali informasi dalam himpunan yang tidak seimbang. Dengan demikian, evaluasi performa algoritma terbaik dapat direkomendasikan sebagai model yang relevan dengan dataset untuk memperoleh luaran analisis sentimen yang tergolong fit atau sesuai.

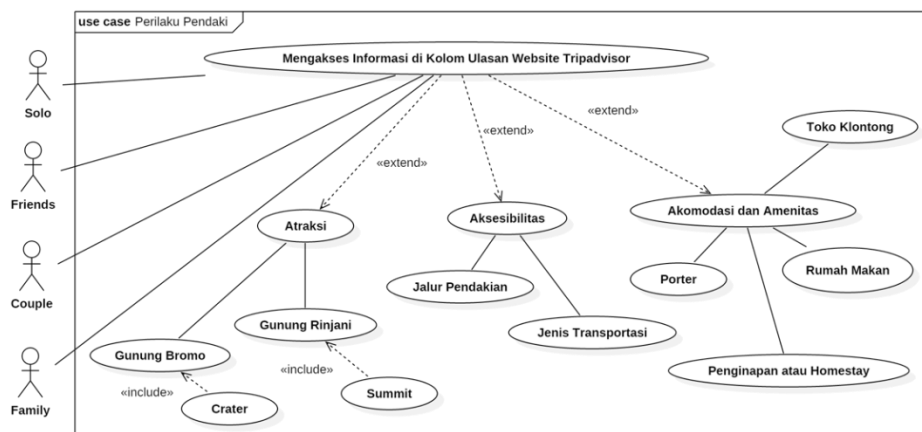
3. HASIL DAN PEMBAHASAN

Pembahasan dalam penelitian ini terbagi menjadi dua bagian yaitu interpretasi perilaku wisatawan berdasarkan tipe kunjungan dan limat kata teratas dari ulasan pengunjung Gunung Rinjani dan Gunung Bromo di website Tripadvisor. Selanjutnya, analisis dan evaluasi performa algoritma Decision Tree (DT) dan Support Vector Machine (SVM).

3.1 Interpretasi Perilaku Wisatawan berdasarkan Tipe Kunjungan dan Lima Kata Teratas dari Ulasan Pengunjung Gunung Rinjani dan Gunung Bromo di Website Tripadvisor

Beberapa studi tentang perilaku wisatawan menunjukkan adanya proses pencarian informasi di beberapa platform digital sebelum memutuskan untuk berkunjung ke suatu destinasi wisata. Oleh sebab itu, pengelola website informatif menyediakan informasi dengan klasifikasi data yang sesuai dengan kebutuhan pengguna. Beberapa informasi yang dibutuhkan ialah atraksi wisata, jenis akomodasi dan moda transportasi yang dapat digunakan mulai dari daerah asal ke daerah tujuan wisata. Adapun, persepsi calon wisatawan juga dipengaruhi oleh popularitas destinasi wisata yang ramai diulas oleh pengguna sistem. Semakin banyak ulasan tentang atraksi wisata, jenis akomodasi dan moda transportasi semakin meningkatkan intensi untuk berkunjung destinasi tersebut menggunakan produk dan layanan serupa. Hal ini menunjukkan bahwa media pemasaran digital berbasis website untuk destinasi pariwisata, perlu menampilkan media komunikasi antar pengguna dalam bentuk ulasan atau storytelling tentang kondisi eksisting dari 3A (atraksi, akomodasi, dan aksesibilitas) sehingga menjadi rujukan atau rekomendasi sebelum mengambil keputusan berwisata. Dengan demikian, ulasan pengguna dalam bentuk data teks dapat dihimpun, diidentifikasi dan dianalisis secara komprehensif untuk meningkatkan produk dan layanan suatu destinasi wisata sehingga meningkatkan tingkat kunjungan wisatawan dan mendukung keberlanjutan destinasi wisata tersebut.

Penelitian ini berupaya menghubungkan perilaku wisatawan dengan sistem informasi melalui kajian analisis sentimen berbasis CRISP-DM pada kasus wisata minat khusus pendakian Gunung Rinjani dan Gunung Bromo. Pembahasan tentang wisata minat khusus pendakian di Indonesia masih sangat terbatas dari perspektif sistem informasi khususnya data mining. Sapardi et al. menunjukkan bahwa wisata minat khusus pendakian memberikan keuntungan ekonomi bagi masyarakat di wilayah perdesaan yang menjadi jalur pendakian, dimana pendaki akan dilayani untuk mendaftarkan diri sebelum mulai pendakian ke puncak [25]. Masyarakat desa dapat menyediakan bahan makanan hingga jasa porter bagi pendaki untuk mendapatkan keuntungan finansial, bahkan aktivitas ekonomi pariwisata akan berkelanjutan apabila ramai dikunjungi oleh para pendaki [26]. Adapun, Wardana et al. menunjukkan tingginya minat terhadap wisata minat khusus pendakian memantik akademisi untuk mengembangkan aplikasi jalur pendakian berbasis mobile. Hal ini menunjukkan bahwa pendaki memiliki kecenderungan untuk mencari informasi dari platform digital berbasis website maupun mobile, mengumpulkan informasi, memanfaatkan informasi sebagai rujukan untuk menentukan destinasi wisata minat khusus pendakian, sebagai rujukan untuk menggunakan jasa transportasi menuju lokasi pendakian, hingga jasa akomodasi selama pendakian. Adapun gambaran tentang perilaku pendaki dalam menetapkan destinasi wisata pendakian dapat divisualisasikan sebagai gambar 2 berikut.



Gambar 2. Perilaku Wisatawan dalam Proses Mencari Informasi untuk Wisata Minat Khusus Pendakian

Gambar 2 merupakan gambaran tentang perilaku wisatawan yang memanfaatkan informasi dari website Tripadvisor untuk menentukan destinasi wisata minat khusus pendakian Gunung Bromo dan Gunung Rinjani berdasarkan tipe kunjungan solo, friends, couple, dan family. Secara umum, informasi terkait dengan atraksi, aksesibilitas, akomodasi dan amenitas yang telah diulas oleh para pengunjung sebelumnya, dijadikan rujukan dalam merencanakan pendakian. Disisi lain, pendakian yang dilakukan bersama keluarga dan pasangan (family/couple) didorong oleh intensi yang beragam, seperti halnya edukasi anak dan keluarga tentang lingkungan alam, maupun kepentingan untuk dokumentasi persiapan acara pernikahan pasangan dengan latarbelakang alam pegunungan. Selanjutnya, pendakian yang dilakukan bersama teman atau sendiri (friends/solo) didorong oleh intensi untuk menambah pengalaman, mencari suasana baru dan menantang, maupun untuk keakraban komunitas. Masing-masing pendaki memiliki cerita atas pengalaman yang beragam, dan hal tersebut diakomodir oleh pengelola website Tripadvisor sehingga dapat dibagikan melalui fitur kolom ulasan pengguna. Semakin baik, rating dan ulasan yang diberikan oleh pendaki, semakin meningkatkan citra/brand dari destinasi wisata minat khusus pendakian tersebut. Dalam konteks Gunung Bromo, atraksi yang menarik wisatawan ialah kawah (crater)

sedangkan dalam konteks Gunung Rinjani, atraksi yang menarik wisatawan ialah puncak (summit). Hal ini menunjukkan bahwa data teks website Tripadvisor dapat digunakan sebagai dataset dalam proses analisis perilaku wisatawan berdasarkan konteks wisata minat khusus pendakian. Mempertimbangkan hal tersebut maka, penelitian ini mengumpulkan data teks dari hasil ulasan wisatawan sebagai pendaki di Gunung Rinjani dan Gunung Bromo pada website Tripadvisor, sebagaimana Tabel 1 berikut ini.

Tabel 1. Jumlah Data Teks Ulasan Wisatawan Gunung Rinjani dan Gunung Bromo

Dataset	Jumlah Data (Sebelum Filter)	Jumlah Data (Setelah Filter)	Link Tripadvisor
Gunung Rinjani	636	326	https://www.tripadvisor.com/Attraction_Review-g297733-d386938-Reviews-Mount_Rinjani-Lombok_West_Nusa_Tenggara.html
Gunung Bromo	2532	1075	https://www.tripadvisor.com/Attraction_Review-g6643743-d446941-Reviews-Mount_Bromo-Bromo_Tengger_Semeru_National_Park_East_Java_Java.html

Tabel 1 merupakan hasil pengumpulan data teks dari kolom ulasan Website Tripadvisor untuk destinasi wisata minat khusus pendakian Gunung Rinjani dan Gunung Bromo. Dalam proses pengumpulan data ulasan, digunakan aplikasi webharvy dengna konfigurasi nama akun, rating, tipe kunjungan, tanggal berkunjung, judul ulasan, dan detail ulasan. Berdasarkan hasil pengumpulan dan seleksi data, diperoleh 326 data dari 636 data ulasan Gunung Rinani, serta 1075 data dari 2532 data ulasan Gunung Bromo. Adapun, data ulasan yang digunakan telah diseleksi khusus data ulasan dalam bahasa inggris. Selanjutnya data teks dibersihkan dan dirapikan Kembali menggunakan aplikasi Rapidminer melalui operator process documents from data yang didalamnya terdapat operator untuk membersihkan dan menyiapkan data teks seperti tokenize, transform cases, filter tokens by length, dan filter stopwords. berdasarkan konfigurasi sebagai berikut : pertama, tokenize 1 (non-letters), tokenize 2 (regular expression [-!"#\$%&'()*+,-./:;<=>?@[\\]_`{|}~]); kedua, transform cases (lower case); ketiga, filter tokens by length (min char =4 & max chars=25); keempat, filter stopwords (dictionary english). Selain itu, terdapat operator remove duplicate, untuk memastikan tidak ada data ulasan yang sama atau redundant, kemudian operator sentiment extract digunakan untuk pembobotan kata sehingga mendapatkan string score sebelum diklasifikasi menggunakan algoritma Decision Tree (DT) dan Support Vector Machine (SVM). Adapun, hasil pembobotan kata dapat dilihat pada tabel berikut ini.

Tabel 2. Proses Ekstrak Sentimen dari Ulasan Wisatawan

Review					
This mountain is such a sad, depressing place. The incredible beauty of nature brutally destroyed by us humans. Rubbish wherever you go, sleep or eat. And most locals couldn't care less. The path, the lunch spots, the spring, the camps - one horrible collection of food waste, litter, toilet paper, thrown away clothes and human waste. Another issue are the "graffiti". Local tourists are smearing their names on everything. Rocks, trees... The worst place of the whole trek is the lake. From far away its stunning and you really wanna swim in it. If you come closer, you don't even wanna get near the water anymore. Everything is covered in flies. Rubbish fires, rubbish piles, swimming plastic bags, dead fish. And rats. They came to check out our food, only 30 centimetres away from my foot. Shocking. And as soon as their tourist group has left the spot, most porters are starting to burn the rubbish. Only very few companies are taking it down. Like Rudy Trekker. So if you really really wanna go, please choose a decent company. For the mountains sake and for your own. The good companies provide way better food, English speaking guides, better material (mattress, sleeping bag, etc) and: they take their rubbish. We went with Rudy Trekker and our guide even picked up rubbish on the way down. We helped him. The porters not so much. Maybe because their colleagues who were at the resting spots with other companies laughed at them. Depressing.					
String Scoring					
sad (-0.54) depressing (-0.41) beauty (0.72) brutally (-0.77) destroyed (-0.56) care (0.56) horrible (-0.64) waste (-0.46) waste (-0.46) worst (-0.79) stunning (0.41) dead (-0.85) shocking (-0.44) like (0.38) please (0.33) good (0.49) better (0.49) better (0.49) laughed (0.51) depressing (-0.41)					
Sentiment	Score	Negativity	Positivity	Uncovered Tokens	Tokens
Negatif	-1,9	6,3	4,4	285	305

Review
Rinjani Climb is so far the most challenging climb i have done. Though the trek up to the summit is tough, i am glad that my friends and i made it. Certain physical fitness is still required as the terrains are not as easy as it seem. But once you reached the summit, the view you see are simply amazing! Its worth all the hard work. Gald to have a good guide and our strong porters who never fails to get us a good campsite spot. I was definitely

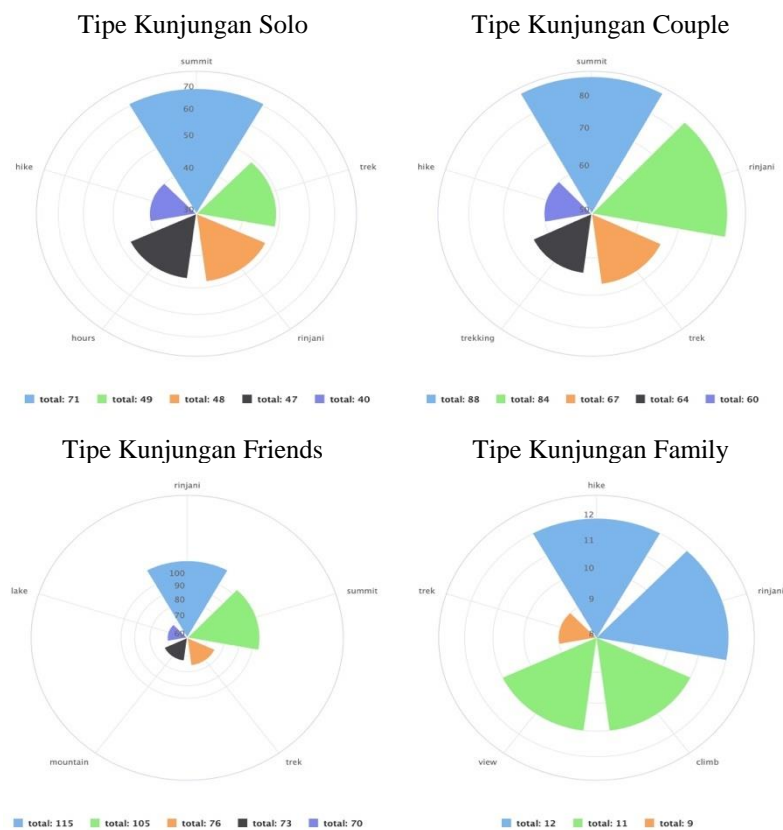
amazed by the porters strength and speed. The service provided by them is just wonderful. We are very well fed and taken care of.

String Scoring

challenging (0.15) tough (-0.13) glad (0.51) friends (0.54) certain (0.28) fitness (0.28) easy (0.49) reached (0.10) amazing (0.72) worth (0.23) hard (-0.10) good (0.49) strong (0.59) fails (-0.46) good (0.49) definitely (0.44) amazed (0.56) strength (0.56) wonderful (0.69) well (0.28) care (0.56)

Sentiment	Score	Negativity	Positivity	Uncovered Tokens	Tokens
Positif	7,3	0,7	8	103	124

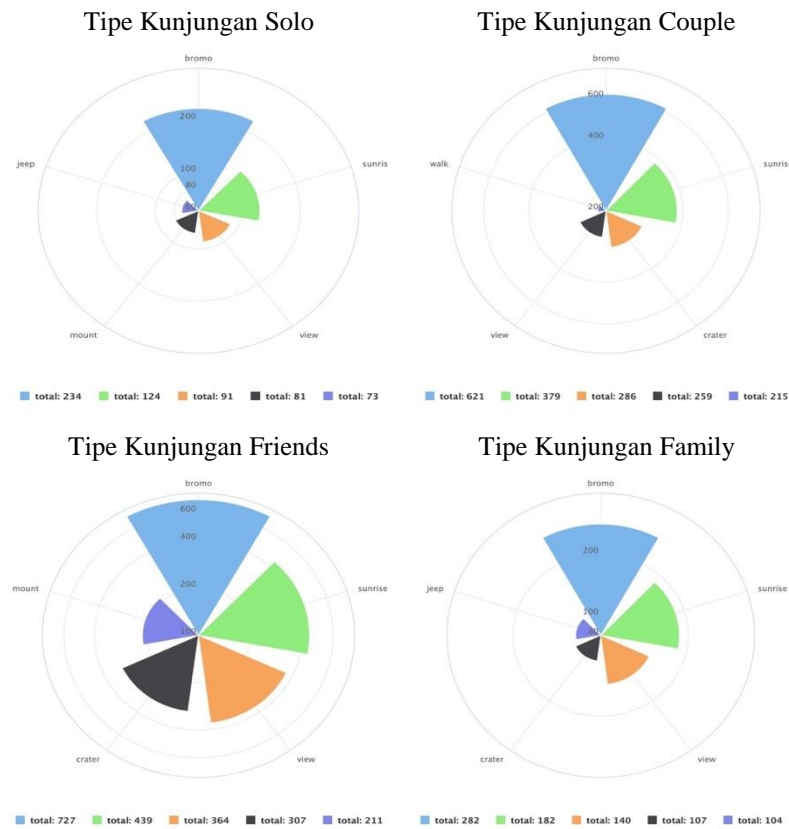
Tabel 2 merupakan hasil ekstrak sentimen dari ulasan wisatawan sebagai pendaki dimana operator extract sentiment di aplikasi Rapidminer akan mengidentifikasi dan melakukan seleksi kata, memberikan bobot, kemudian mengakumulasi bobot tersebut sehingga dapat diklasifikasikan berdasarkan sentimen positif dan sentimen negatif. Score dengan angka negatif menunjukkan banyaknya kosa kata yang mengandung makna negatif dalam data ulasan, sementara score dengan angka positif menunjukkan banyaknya kosa kata yang mengandung makna positif dalam ulasan. Salah satu kekurangan dari operator extract sentiment ialah pembobotan (string scoring) hanya diberikan berdasarkan diksi atau kosa kata yang mengandung makna positif atau negatif, yang telah dikenali dan diberikan bobot oleh sistem. Sementara itu, analisis sentimen merupakan proses identifikasi dan klasifikasi berdasarkan makna teks secara keseluruhan, bukan parsial. Hal ini menjadi tantangan bagi pengembangan sistem aplikasi dalam konteks data mining agar dapat mempertimbangkan makna keseluruhan dari kalimat yang diulas oleh wisatawan tentang destinasi wisata minat khusus Gunung Rinjani dan Gunung Bromo, untuk ekspresi seperti kekaguman, sarkasme, sindiran, protes, komplain dari pendaki Meskipun demikian, hasil TF-IDF dan proses pembobotan kata menggunakan operator wordlist dan sortir data dapat divisualisasikan dalam bentuk polar chart lima kata teratas yang paling sering muncul dalam ulasan wisatawan berdasarkan tipe kunjungan ke Gunung Rinjani dan Gunung Bromo sebagaimana gambar 3 berikut.



Gambar 3. Lima Kata Populer dalam Data Ulasan Pengunjung Gunung Rinjani

Gambar 3 merupakan hasil identifikasi lima kata populer berdasarkan tipe kunjungan ke Gunung Rinjani, sebagai berikut : pertama, tipe kunjungan sendiri (Solo) menunjukkan bahwa puncak (summit), jalur pendakian (trek) dan waktu pendakian (hours) menjadi sorotan utama; kedua, tipe kunjungan bersama pasangan (couple) paling banyak menyoroti puncak (summit) dan jalur pendakian (trekking dan trek); ketiga, tipe kunjungan bersama teman (friends) paling banyak menyoroti puncak (summit) dan jalur pendakian (trek) dan danau (lake); keempat, tipe kunjungan bersama keluarga paling banyak menyoroti aktivitas pendakian (climb dan hike), jalur pendakian (trek), dan pemandangan (view). Hal ini menunjukkan bahwa tipe kunjungan wisatawan menentukan preferensi

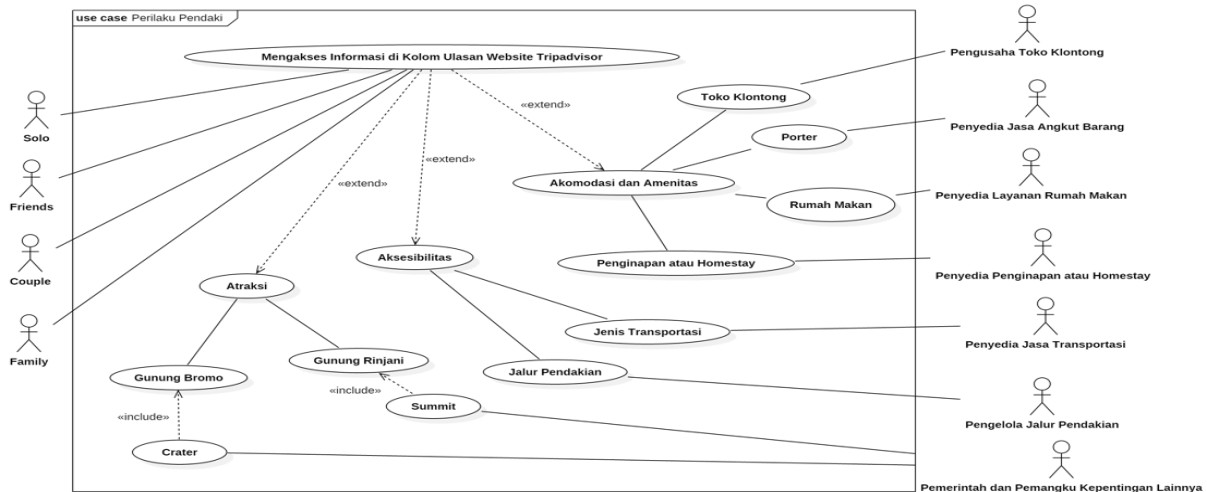
terhadap kebutuhan selama berwisata. Apabila diamati dari jumlah kata yang disoroti tipe pengunjung dengan keluarga (family), dapat diketahui bahwa puncak gunung bukan utama melainkan pemandangan dan aktivitas pendakian. Berbeda halnya dengan tipe kunjungan sendiri (Solo), bersama pasangan maupun dengan teman (friends) yang menyoroti puncak gunung (summit) sebagai orientasi utama. Perbedaan preferensi wisatawan terhadap produk dan layanan wisata minat khusus pendakian juga dapat dilihat berdasarkan lima kata populer yang muncul dalam ulasan Gunung Bromo berdasarkan tipe kunjungan pada gambar 4 berikut.



Gambar 4. Lima Kata Populer dalam Data Ulasan Pengunjung Gunung Bromo

Gambar 4 merupakan hasil identifikasi lima kata populer berdasarkan tipe kunjungan ke Gunung Bromo, sebagai berikut : pertama, tipe kunjungan sendiri (Solo) menunjukkan bahwa pemandangan saat matahari terbit (sunrise), pemandangan sekitar (view), dan transportasi lokal (jeep) menjadi sorotan utama; kedua, tipe kunjungan bersama pasangan (couple) paling banyak menyoroti kawah (crater), pemandangan saat matahari terbit (sunrise), dan aktivitas jalan-jalan di area sekitar gunung (walk); ketiga, tipe kunjungan bersama teman (friends) paling banyak menyoroti pemandangan saat matahari terbit (sunrise), pemandangan sekitar (view), dan kawah (crater); keempat, tipe kunjungan bersama keluarga (family) paling banyak menyoroti pemandangan saat matahari terbit (sunrise), pemandangan sekitar (view), kawah (crater) dan transportasi lokal (jeep). Hal ini menunjukkan bahwa masing-masing pengunjung memiliki preferensi terhadap jenis transportasi yang digunakan dalam berwisata ke Gunung Bromo untuk melihat pemandangan saat matahari terbit, pemandangan sekitar dan kawah. Hal ini menunjukkan bahwa wisatawan menggunakan jasa layanan transportasi (aksesibilitas) dan akomodasi (penginapan) untuk menikmati pemandangan pagi hari ketika matahari terbit (sunrise) maupun pemandangan sekitar kawah. Dengan demikian, dapat diketahui bahwa tipe kunjungan wisatawan memberikan penilaian yang berbeda-beda terkait dengan atraksi, akomodasi dan aksesibilitas di bidang pariwisata.

Berdasarkan hasil analisis sentimen, dapat diketahui bahwa wisatawan yang berkunjung ke Gunung Rinjani dan Gunung Bromo memberikan penilaian yang positif terhadap produk dan layanan. Meskipun demikian, secara spesifik, dapat diketahui bahwa jumlah kata yang paling banyak diulas memberikan sejumlah data terkait dengan kritik atas persoalan sampah di lokasi pendakian, keramahan masyarakat lokal yang mengelola jalur pendakian, serta pemangku kepentingan lainnya yang bertanggungjawab atas keselamatan para pendaki. Hal ini menunjukkan bahwa dampak dari pengembangan wisata minat khusus pendakian, berkontribusi bagi perkembangan ekonomi pihak-pihak yang terlibat dalam pengembangan destinasi, sebagaimana gambar berikut.



Gambar 5. Penerima Manfaat Wisata Minat Khusus Pendakian Gunung Rinjani dan Gunung Bromo

Gambar 5 merupakan gambaran tentang pihak-pihak yang terlibat dalam pengembangan destinasi wisata minat khusus pendakian di Gunung Bromo dan Gunung Rinjani. Dalam upaya mempertahankan dan meningkatkan kunjungan ke Gunung Bromo dan Gunung Rinjani, diperlukan kerjasama antar berbagai pemangku kepentingan untuk meningkatkan kualitas produk dan layanan yang relevan dengan tipe kunjungan wisatawan. Secara umum, dapat diketahui bahwa lima kata teratas yang sering muncul dalam data ulasan wisatawan di gunung Rinjani ialah summit (272), rinjani (259), trek (201), hike (170), mountain (159). Sementara itu, lima kata teratas yang sering muncul dalam data ulasan wisatawan di gunung Bromo ialah bromo (1864), sunrise (1124), view (854), crater (758), mount (577). Hal ini menunjukkan bahwa kata puncak (summit), kawah (crater), pemandangan (view) pemandangan saat matahari terbit (sunrise), alat transportasi (jeep) dan jalur pendakian (trek) menjadi faktor penarik wisatawan ke Gunung Bromo dan Gunung Rinjani. Masing-masing pemangku kepentingan dapat menyediakan produk atau layanan yang berhubungan dengan kebutuhan wisatawan selama berkunjung, sehingga memberikan manfaat ekonomi. Demikian halnya, pemerintah bersama masyarakat lokal dapat mengelola jalur pendakian atau akses menuju puncak agar lingkungan tetap bersih dan asri, serta tidak dirusak oleh oknum-oknum yang tidak bertanggungjawab. Dengan demikian, pengembangan wisata minat khusus pendakian menjadi populer, menarik wisatawan mancanegara, menguntungkan masyarakat lokal, serta menjaga keberlanjutan lingkungan.

3.2 Analisis dan Evaluasi Performa Algoritma Decision Tree (DT) dan Support Vector Machine (SVM)

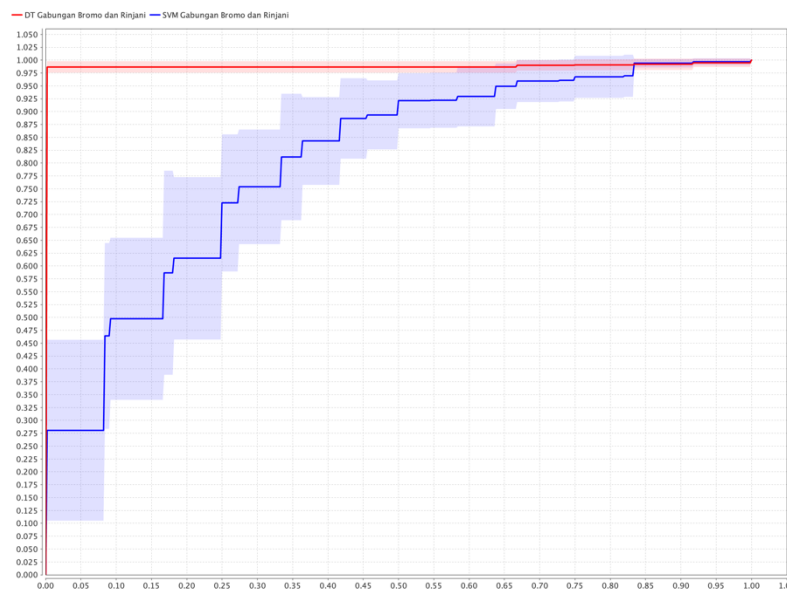
Berdasarkan hasil klasifikasi 326 data ulasan pengunjung Gunung Rinjani dan 1075 data ulasan pengunjung Gunung Bromo, menggunakan algoritma DT dan SVM dapat diketahui bahwa algoritma SVM memiliki performa yang lebih baik dibandingkan dengan DT. Hal tersebut dapat dilihat dari hasil kalkulasi algoritma DT menggunakan operator SMOTE yang menunjukkan nilai akurasi 93,05%, presisi 89,21%, recall 98,07%, f-measure 93,40%, AUC 93,3%, t-test 93,1%. Adapun, algoritma DT tanpa menggunakan operator SMOTE menghasilkan nilai akurasi 89,86%, presisi 90,85%, recall 98,58%, f-measure 94,55%, AUC 57,7%, t-test 89,9%. Berbeda halnya dengan algoritma SVM yang menggunakan operator SMOTE mampu menghasilkan nilai akurasi 97,67%, presisi 100%, recall 95,34%, f-measure 97,61%, AUC 100%, t-test 97,7%. Demikian juga hasil kalkulasi algoritma SVM tanpa menggunakan operator SMOTE yang menunjukkan nilai akurasi 89,31%, presisi 89,31%, recall 100%, f-measure 94,35%, AUC 80,7%, t-test 89,3%. Hal ini menunjukkan bahwa SVM dapat digunakan sebagai model dalam menganalisis sentimen dataset Gunung Rinjani dan Gunung Bromo untuk memberikan rekomendasi bagi pengelola maupun pemangku kepentingan lainnya agar dapat mengoptimalkan produk dan layanan bagi para aktor sebagai penerima manfaat dari bisnis wisata minat khusus pendakian. Dengan demikian strategi pemasaran destinasi dapat dioptimalkan berdasarkan tipe kunjungan dan preferensi wisatawan yang diperoleh dari hasil pengolahan data ulasan pengunjung Gunung Rinjani dan Gunung Bromo di website Tripadvisor. Pada tabel 3 dapat dilihat hasil evaluasi algoritma dalam bentuk confusion matrix yang diklasifikasikan berdasarkan penggunaan operator SMOTE dan tanpa operator SMOTE, dimana nilai akurasi, presisi, recall, f-measure, dan AUC mencerminkan performa dari algoritma DT dan SVM.

Tabel 3. Hasil Klasifikasi Algoritma untuk Data Ulasan Gunung Rinjani dan Gunung Bromo

Gunung Rinjani												
Algoritma	TP	FN	FP	TN	Accuracy	Precision	Recall	F-Measure	AUC	t-Test	Keterangan	
DT	86	11	97	8	19	8	6	93,03%	89,21%	98,07%	93,3%	with SMOTE
	8	19	8	6	93,03%	89,21%		93,40%		93,1%		

Gunung Rinjani											
Algoritma	TP	F N	FP	T N	Accura cy	Precisio n	Recall	F- Measu re	AUC	t-Test	Keterang an
DT	20	14	98	2	89,86%	90,85%	98,58 %	94,55%	57,7 %	89,9%	without SMOTE
SVM	98			94			95,34 %		100		with SMOTE
	6	46	0	0	97,67%	100%		97,61%	%	97,7%	without SMOTE
SVM			11	98					80,7		without SMOTE
	0	0	8	6	89,31%	89,31%	100%	94,35%	%	89,3%	with SMOTE

Tabel 3 merupakan hasil klasifikasi data ulasan wisatawan terhadap destinasi wisata minat khusus pendakian Gunung Rinjani dan Gunung Bromo dapat diketahui bahwa nilai f-measure algoritma SVM menggunakan SMOTE lebih tinggi dibandingkan dengan yang lain, demikian juga nilai AUC yang mencapai 100%. Hal ini menunjukkan bahwa algoritma SVM dengan operator SMOTE menunjukkan performa yang baik dalam proses klasifikasi ulasan wisatawan ke dalam kelas positif dan kelas negatif. Berdasarkan kerangka CRISP-DM, algoritma SVM dengan operator SMOTE dapat digunakan dalam proses pengolahan dataset sehingga dapat memberikan rekomendasi manajemen destinasi wisata minat khusus pendakian Gunung Rinjani dan Gunung Bromo. Munir et al. berpendapat bahwa dalam klasifikasi dataset untuk analisis sentimen di bidang pariwisata khususnya ulasan tamu hotel, algoritma SVM mampu menunjukkan performa yang baik [27]. Disisi lain, Idris et al. menunjukkan bahwa algoritma SVM mampu mengklasifikasi data ulasan dengan nilai akurasi yang tinggi [28]. Hal ini menunjukkan bahwa dalam berbagai proses klasifikasi sentimen, algoritma SVM telah menunjukkan performa yang baik. Selain SVM, algoritma DT juga mampu menunjukkan kinerja yang tinggi, sebagaimana nilai ROC pada gambar berikut.



Gambar 6. ROC Algoritma DT dan SVM untuk Dataset Gunung Rinjani dan Gunung Bromo

Gambar 6 menunjukkan bahwa selain algoritma SVM, algoritma DT juga memiliki performa yang baik. Beberapa studi terdahulu menunjukkan bahwa penggunaan algoritma DT dalam klasifikasi sentimen mampu menghasilkan nilai akurasi yang tinggi. Zy menunjukkan bahwa algoritma DT dapat digunakan dalam klasifikasi sentimen menggunakan dataset dari media sosial [29]. Adapun, Syahid dan Mahdiana menunjukkan bahwa algoritma DT juga dapat dijadikan model dalam kerangka kerja CRISP-DM dengan menggunakan dataset dari media sosial, salah satunya ialah twitter [30]. Ginantra juga menunjukkan bahwa operator SMOTE dapat digunakan untuk meningkatkan performa algoritma DT dalam menganalisis sentimen tamu hotel [31]. Hal ini menunjukkan bahwa analisis sentimen tidak terbatas pada evaluasi klasifikasi berbasis data confusion matrix melainkan suatu rangkaian proses identifikasi, seleksi, dan analisis dalam kerangka CRISP DM. Dengan demikian, setiap tahapan dalam kerangka kerja CRISP-DM dimulai dari tahap business understanding, data understanding, data preparation, modeling, evaluation and deployment memiliki tujuan untuk memperjelas tujuan dari analisis sentimen serta manfaat dari pengolahan data untuk kepentingan manajerial organisasi/perusahaan atau destinasi wisata. Dengan demikian dapat diketahui bahwa proses analisis sentimen berdasarkan kerangka CRISP-DM dapat digunakan sebagai model untuk generate rekomendasi bagi pengelola destinasi wisata minat khusus pendakian Gunung Rinjani dan Gunung Bromo, serta pemangku kepentingan lainnya dalam memasarkan produk dan layanan usaha pendukung pariwisata maupun dalam menetapkan target pasar berdasarkan tipe kunjungan (solo, friends,

couple, family), dengan menyiapkan sarana dan prasarana yang berhubungan dengan kebutuhan atau preferensi konsumen terkait akomodasi, amenitas, jenis transportasi, hingga atraksi wisata.

4. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa dalam konteks gunung Rinjani, lima kata teratas yang paling sering muncul dalam ulasan wisatawan berdasarkan tipe kunjungan ialah sebagai berikut : tipe kunjungan sendiri (solo), summit (71), trek (49), rinjani (48), hours (47), hike (40); tipe kunjungan dengan pasangan (couple), summit (88), rinjani (84), trek (67), trekking (64), hike (60); tipe kunjungan bersama teman (friends), rinjani (115), summit (105), trek (76), mountain (73), lake (70); tipe kunjungan dengan keluarga (family), hike (12), rinjani (12), climb (11), view (11), trek (9). Secara keseluruhan, lima kata teratas yang sering muncul dalam data ulasan wisatawan di gunung Rinjani ialah sebagai berikut : summit (272), rinjani (259), trek (201), hike (170), mountain (159). Disisi lain, dalam konteks gunung Bromo, lima kata teratas yang paling sering muncul dalam ulasan wisatawan berdasarkan tipe kunjungan ialah sebagai berikut : tipe kunjungan sendiri (solo), bromo (234), sunrise (124), view (91), mount (81), jeep (73); tipe kunjungan dengan pasangan (couple), bromo (621), sunrise (379), crater (286), view (259), walk (215); tipe kunjungan bersama teman (friends), bromo (727), sunrise (439), view (364), crater (307), mount (211); tipe kunjungan dengan keluarga (family), bromo (282), sunrise(182), view (140), crater (107), jeep (104). Adapun, lima kata teratas yang sering muncul dalam data ulasan wisatawan di gunung Bromo ialah sebagai berikut : bromo (1864), sunrise (1124), view (854), crater (758), mount (577). Berdasarkan hasil pengolahan 1075 data teks gunung Bromo, dan 326 data ulasan gunung Rinjani, menggunakan algoritma Decision Tree (DT) dan Support Vector Machine (SVM) dapat diketahui bahwa algoritma DT menggunakan operator SMOTE menghasilkan nilai akurasi 93,05%, presisi 89,21%, recall 98,07%, f-measure 93,40%, AUC 93,3%, t-test 93,1%. Algoritma DT tanpa menggunakan operator SMOTE menghasilkan nilai akurasi 89,86%, presisi 90,85%, recall 98,58%, f-measure 94,55%, AUC 57,7%, t-test 89,9%. Algoritma SVM menggunakan operator SMOTE menghasilkan nilai akurasi 97,67%, presisi 100%, recall 95,34%, f-measure 97,61%, AUC 100%, t-test 97,7%. Algoritma SVM tanpa menggunakan operator SMOTE menghasilkan nilai akurasi 89,31%, presisi 89,31%, recall 100%, f-measure 94,35%, AUC 80,7%, t-test 89,3%. Hal ini menunjukkan bahwa algoritma SVM menunjukkan performa yang lebih baik dalam melakukan klasifikasi sentimen berdasarkan kelas positif dan negatif. Dengan demikian, model untuk analisis sentimen wisatawan dalam konteks wisata minat khusus pendakian Gunung Rinjani dan Gunung Bromo, dapat menggunakan algoritma SVM. Selanjutnya, atraksi, aksesibilitas dan akomodasi perlu dikembangkan sesuai dengan tipe kunjungan wisatawan (solo, couple, friends, and family) dengan demikian dapat meningkatkan intensi untuk berkunjung kembali dan mendukung keberlanjutan destinasi wisata Gunung Rinjani dan Gunung Bromo.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada seluruh lembaga yang turut berkontribusi memberikan dukungan hingga penelitian ini dapat selesai. Terima kasih kepada Lembaga Penelitian dan Pengabdian Masyarakat (LPPM), Fakultas Ilmu Administrasi Bisnis dan Ilmu Komunikasi (FIABIKOM), Program Studi Pariwisata, Fakultas Teknik (FT), Program Studi Sistem Informasi, Universitas Katolik Indonesia Atma Jaya.

REFERENCES

- [1] S. A. Azzahra and A. Wibowo, "Analisis Sentimen Multi-Aspek Berbasis Konversi Ikon Emosi dengan Algoritme Naïve Bayes untuk Ulasan Wisata Kuliner Pada Web Tripadvisor," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 4, pp. 737–744, 2020, doi: 10.25126/jtiik.2020731907.
- [2] H. J. Christanto and Y. A. Singgalen, "Sentiment Analysis on Customer Perception towards Products and Services of Restaurant in Labuan Bajo," *J. Inf. Syst. Informatics*, vol. 4, no. 3, pp. 511–523, 2022, doi: 10.51519/journalisi.v4i3.276.
- [3] Y. A. Singgalen, "Analisis Performa Algoritma NBC , DT , SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1634–1646, 2022, doi: 10.47065/bits.v4i3.2766.
- [4] Y. A. Singgalen, "Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, p. 1343–1352, 2022, doi: 10.47065/bits.v4i3.2486.
- [5] A. M. M. Fattah, A. Voutama, N. Heryana, and N. Sulistiyowati, "Pengembangan Model Machine Learning Regresi sebagai Web Service untuk Prediksi Harga Pembelian Mobil dengan Metode CRISP-DM," *JURIKOM J. Ris. Komput.*, vol. 9, no. 5, pp. 1669–1678, 2022, doi: 10.30865/jurikom.v9i5.5021.
- [6] C. Wulandari, Y. Ansori, and K. Fahmi, "CRISP-DM Method On Indonesian Micro Industries (UMKM) Using K-Means Clustering Algorithm," *MATICS J. Ilmu Komput. dan Teknol. Inf. (Journal Comput. Sci. Inf. Technol.)*, vol. 14, no. 2, pp. 35–40, 2022, doi: 10.18860/mat.v14i2.13760.
- [7] Y. Christian and K. O. Y. R. Qi, "Penerapan K-Means pada Segmentasi Pasar untuk Riset Pemasaran pada Startup Early Stage dengan Menggunakan CRISP-DM," *JURIKOM J. Ris. Komput.*, vol. 9, no. 4, pp. 966–973, 2022, doi: 10.30865/jurikom.v9i4.4486.
- [8] D. A. Munawwaroh and A. H. Primandari, "Implementasi Crisp-Dm Model Menggunakan Metode Decision Tree Dengan

- Algoritma Cart Untuk Prediksi Lila Ibu Hamil Berpotensi Gizi Kurang,” *J. Ilm. Pendidik. Mat.*, vol. 10, no. 2, pp. 367–380, 2022, [Online]. Available: <http://dx.doi.org/10.31941/delta.v10i2.2172>
- [9] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiyanti, “Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE,” *Aiti*, vol. 18, no. 2, pp. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.
- [10] Rousyati, W. Gata, D. Pratmanto, and N. K. Warchani, “Analisis Sentimen Financial Technology Peer to Peer Lending Pada Aplikasi Koinworks,” *J. Teknol. Infor*, vol. 9, no. 6, pp. 1167–1176, 2022, doi: 10.25126/jtiik.202294409.
- [11] Y. A. Singgalen, “Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Systematic Literature Review,” *J. Inf. Syst. Informatics*, vol. 3, no. 2, pp. 278–302, 2021.
- [12] R. Despica and E. Zuriyani, “Tourism Awareness and Sapta Pesona Training for the Community of the Purus Beach Tourism Area , Padang City,” *Asian J. Community Serv.*, vol. 1, no. 6, pp. 305–314, 2022.
- [13] L. K. M. Putri, P. A. A. Damayanti, and N. L. P. E. Diarthini, “Pengetahuan, Sikap dan Perilaku Terhadap Demam Berdarah Dengue Pada Wisatawan di Kecamatan Ubud, Gianyar Bali,” *Hurnal Med. Udayana*, vol. 11, no. 3, pp. 10–17, 2022.
- [14] R. Sarudin and A. Ismail, “Analisis Online Review Tripadvisor.com Terhadap Minat Pembelian Produk Jasa Akomodasi Di Hotel Manhattan,” *J. Hosp. dan Pariwisata*, vol. 7, no. 4, pp. 33–43, 2021.
- [15] F. Abadi and Herwin, “Pengaruh Kualitas Pelayanan Terhadap Kepuasan Wisatawan Berdampak Kepada Perilaku Masa Depan Wisatawan Domestik,” *Manag. Account. Expo.*, vol. 3, no. 2, pp. 134–142, 2020, doi: 10.36441/mae.v3i2.210.
- [16] C. Pomantow, F. M. Langi, and C. N. Waworuntu, “Analisis Perilaku Wisatawan Dalam Memilih Objek Wisata di Kota Manado,” *Humanlight J. Psychol.*, vol. 3, no. 2, pp. 102–113, 2022.
- [17] R. Lubis, A. Munang, and H. Q. Karima, “Analisis Faktor yang Mempengaruhi Keputusan Berkunjung Pada Waterpark Top 100 Batu Aji,” *JIEOM*, vol. 5, no. 2, pp. 178–189, 2022.
- [18] R. Kosasih and A. Alberto, “Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor,” *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 6, no. 1, pp. 134–139, 2021.
- [19] Y. E. Kurniawati, “Class Imbalanced Learning Menggunakan Algoritma Synthetic Minority Over-sampling Technique – Nominal (SMOTE-N) pada Dataset Tuberculosis Anak,” *J. Buana Inform.*, vol. 10, no. 2, pp. 134–143, 2019, doi: 10.24002/jbi.v10i2.2441.
- [20] M. F. Asshiddiqi and K. M. Lhaksana, “Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI,” in *e-Proceeding of Engineering*, 2020, vol. 7, no. 3, pp. 9936–9948.
- [21] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, pp. 646–654, 2021, doi: 10.32493/informatika.v5i4.7622.
- [22] D. N. Fitriana and Y. Sibaroni, “Sentiment Analysis on KAI Twitter Post Using Multiclass Support Vector Machine (SVM),” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 846–853, 2020, doi: 10.29207/resti.v4i5.2231.
- [23] A. Karim, “Perbandingan Prediksi Kemiskinan di Indonesia Menggunakan Support Vector Machine (SVM) dengan Regresi Linear,” *J. Sains Mat. dan Stat.*, vol. 6, no. 1, pp. 107–112, 2020, doi: 10.24014/jsms.v6i1.9259.
- [24] E. A. Nida, “Analisis Kinerja Algoritma Support Vector Machine (SVM) Guna Pengambilan Keputusan Beli/Jual Pada Saham PT Elnusa Tbk. (ELSA),” *J. Transform.*, vol. 17, no. 2, pp. 160–170, 2020, doi: 10.26623/transformatika.v17i2.1649.
- [25] Sapardi, E. Martius, and Erwin, “Analisis Dampak Pengelolaan Wisata Minat Khusus Terhadap Ekonomi dan Sosial Budaya Masyarakat Nagari Air Batumbuk Kabupaten Solok (Studi Kasus Objek Pendakian Gunung Talang Melalui Jalur Air Batumbuk),” *J. Ilm. MEA (Manajemen, Ekon. dan Akuntansi)*, vol. 5, no. 1, pp. 484–501, 2021.
- [26] P. N. Sadikin, S. Mulatsih, B. Pramudya, and H. S. Arifin, “Analisis Willingness-To-Pay Pada Ekowisata Taman Nasional Gunung Rinjani,” *J. Anal. Kebijak. Kehutan.*, vol. 14, no. 1, pp. 31–46, 2017, doi: 10.20886/jakk.2017.14.1.31-46.
- [27] A. Munir, E. P. Atika, and A. D. Indraswari, “Analisis Sentimen pada review hotel menggunakan metode pembobotan dan klasifikasi,” *J. Jnanaloka*, vol. 3, no. 1, pp. 33–38, 2022, doi: 10.36802/jnanaloka.2022.v3-no1-33-38.
- [28] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023.
- [29] A. T. Zy, “Comparison Algorithm Classification Naive Bayes, Decision Tree, and Neural Network for Analysis Sentiment,” *J. Pelita Teknol.*, vol. 12, no. 1, pp. 1–14, 2017.
- [30] J. A. Syahid and D. Mahdiana, “Perbandingan algoritma untuk klasifikasi analisis sentimen terhadap Genose pada media sosial Twitter,” *semanTIK*, vol. 7, no. 1, pp. 9–16, 2021, doi: 10.5281/zenodo.5034916.
- [31] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvandana, and I. K. A. G. Wiguna, “Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan k-NN,” *Janapati*, vol. 11, no. 3, pp. 205–216, 2022.