

# Analisis Sentimen Wisatawan terhadap Taman Nasional Bunaken dan Top 10 Hotel Rekomendasi Tripadvisor Menggunakan Algoritma SVM dan DT berbasis CRISP-DM

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: 04/02/2023; Accepted: 20/02/2023; Published: 25/02/2023

**Abstrak**—Analisis sentimen wisatawan terhadap Taman Nasional Bunaken dan Top 10 Hotel rekomendasi Tripadvisor perlu dilakukan untuk mengidentifikasi kepuasan wisatawan terkait dengan atraksi, jasa akomodasi dan transportasi yang digunakan. Mempertimbangkan hal tersebut maka penelitian ini menggunakan kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM) dengan menguji performa algoritma Decision Tree (DT) dan Support Vector Machine (SVM). Terdapat enam tahapan dalam CRISP-DM sebagai berikut business understanding, data understanding, data preparation, modeling, evaluation dan deployment. Berdasarkan hasil pengolahan 398 data destinasi Taman Nasional Bunaken, dan 1793 data ulasan terhadap top 10 hotel yang direkomendasikan oleh Tripadvisor, algoritma SVM menunjukkan performa yang lebih baik. Dalam konteks data destinasi Taman Nasional Bunaken, performa algoritma SVM menggunakan operator SMOTE mampu menghasilkan nilai akurasi, presisi, recall, f-measure, AUC, dan t-Test sebanyak 100%. Selain itu, dalam mengolah dataset top 10 hotel yang direkomendasikan oleh Tripadvisor, algoritma SVM menggunakan operator SMOTE memberikan nilai akurasi 96,42%, nilai presisi 100%, nilai recall 92,83%, nilai f-measure 96,27%, nilai AUC 100%, dan nilai t-Test 96,4%. Adapun, hasil identifikasi lima kata yang paling sering muncul dalam ulasan wisatawan untuk destinasi wisata Bunaken Marine Park ialah 164 kata fish, 165 kata island, 193 kata dive, 230 diving, 280 kata bunaken. Hal ini menunjukkan bahwa faktor pendorong kunjungan wisata ke Bunaken Marine Park ialah keindahan alam bawah laut termasuk keanekaragaman jenis hewan laut, keindahan alam daerah kepulauan, dan aktivitas menyelam. Selanjutnya, hasil identifikasi lima kata yang paling sering muncul dalam ulasan wisatawan untuk top 10 hotel yang direkomendasikan Tripadvisor ialah 1170 kata great, 1222 kata bunaken, 1550 kata resort, 1613 kata diving, dan 1711 kata dive. Hal ini menunjukkan bahwa karakteristik wisatawan yang memiliki motif berwisata ke Taman Wisata Nasional Bunaken memiliki intensi untuk menggunakan akomodasi resort di sekitar Bunaken, serta memiliki mobilitas dan fasilitas pendukung aktivitas menyelam. Dengan demikian, luaran penelitian ini dapat digunakan sebagai rekomendasi bagi para pengelola jasa akomodasi agar menyiapkan fasilitas pendukung bagi wisatawan yang hendak mengunjungi Taman Nasional Bunaken.

**Kata Kunci:** Analisis Sentimen; Bunaken; Hotel; Decision Tree; Support Vector Machine

**Abstract**—It is necessary to analyze traveler sentiment towards Bunaken National Park and Tripadvisor's Top 10 Recommended Hotels to identify traveler satisfaction with the attractions, accommodation services, and transportation used. Considering this, this study uses the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework by testing the performance of the Decision Tree (DT) algorithm and the Support Vector Machine (SVM). CRISP-DM has six stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Based on the processing of 398 Bunaken National Park destination data and 1793 review data on the top 10 hotels recommended by Tripadvisor, the SVM algorithm performed better. In the context of Bunaken National Park destination data, the performance of the SVM algorithm using the SMOTE operator can produce 100% accuracy, precision, recall, f-measure, AUC, and t-Test values. In addition, in processing the top 10 hotel datasets recommended by Tripadvisor, the SVM algorithm using the SMOTE operator provides an accuracy value of 96.42%, a precision value of 100%, a recall value of 92.83%, an f-measure value of 96.27%, an AUC value of 100%, and a t-Test value of 96.4%. The results of identifying the five words that most often appear in tourist reviews for Bunaken Marine Park tourist destinations are 164 words fish, 165 words island, 193-word dive, 230 diving, and 280 words Bunaken. This indicates the driving factor for tourist visits to Bunaken Marine Park is the beauty of underwater nature, including the diversity of marine animal species, the natural beauty of the archipelago, and diving activities. In addition, the results of identifying the five words that most often appear in traveler reviews for Tripadvisor's top 10 recommended hotels are 1170 great words, 1222 bunaken words, 1550 resort words, 1613 diving words, and 1711 dive words. This indicates that the characteristics of tourists who have the motive of traveling to Bunaken National Tourism Park have the intention to use resort accommodations around Bunaken and have mobility and facilities to support diving activities. Thus, the output of this study can be used as a recommendation for accommodation service managers to prepare supporting facilities for tourists who want to visit Bunaken National Park.

**Keywords:** Sentiment Analysis; Bunaken; Hotel; Decision Tree; Support Vector Machine

## 1. PENDAHULUAN

Salah satu destinasi wisata bahari yang menarik wisatawan mancanegara ke Indonesia bagian Timur ialah Bunaken Marine Park. Priyatna et al. menunjukkan bahwa pengembangan destinasi wisata bahari Bunaken Marine Park bermanfaat bagi peningkatan kesejahteraan masyarakat lokal melalui aktivitas ekonomi pariwisata [1]. Selain itu, Tular et al. menunjukkan bahwa eksistensi mangrove menjadi salah satu sumberdaya pendukung aktivitas wisata bahari dan edukasi di sekitar Bunaken Marine Park [2]. Adapun, Lukman et al. menunjukkan bahwa kelompok masyarakat di daerah penyangga Taman Nasional Bunaken turut mendapatkan bantuan usaha ekonomi kreatif [3].

Hal ini berarti bahwa pengembangan destinasi wisata bahari di Taman Nasional Bunaken telah mempertimbangkan aspek keberlanjutan dari sisi sosial, ekonomi dan lingkungan. Dengan demikian, aktivitas pariwisata di Taman Nasional Bunaken menjadi salah satu sumberdaya yang perlu didukung dengan kajian yang berkontribusi pada optimalisasi manajemen destinasi.

Beberapa studi tentang destinasi wisata di Taman Nasional Bunaken tidak terbatas pada kajian kepariwisataan, melainkan juga kajian yang menghasilkan teknologi pendukung untuk aktivitas wisata bahari. Aryawan et al. membuat desain kapal sebagai moda transportasi yang cocok untuk para wisatawan ketika berkunjung ke Bunaken untuk melihat pemandangan alam bawah laut [4]. Selain itu, Namoua et al. mengkaji tentang serapan karbon pada lamun di kawasan pantai Tongkaina Kecamatan Bunaken [5]. Adapun, Pomantow et al. mengidentifikasi faktor-faktor yang menyebabkan gangguan dalam manajemen Taman Nasional Bunaken diantaranya : lemahnya penegakan hukum; minimnya Sumber Daya Manusia (SDM) terlatih; keterbatasan sarana dan prasarana penunjang; keterbatasan ekonomi penduduk lokal [6]. Adapun, Sahri et al. menunjukkan bahwa hasil analisis daya dukung menunjukkan bahwa daerah pesisir Desa Poopoh Taman Nasional Bunaken sesuai dan sangat sesuai untuk dikembangkan sebagai daya tarik ekowisata pantai dan bahari [7]. Hal ini menunjukkan bahwa penelitian tentang pariwisata di Taman Nasional Bunaken menekankan pada upaya untuk mempertahankan keberlanjutan sosio-ekonomi, sosio-kultural, dan sosio-ekologi. Dengan demikian diperlukan kajian yang secara komprehensif mengidentifikasi dan menganalisis konteks kepariwisataan di Bunaken Marine Park.

Studi tentang persepsi dan motivasi wisatawan di Taman Nasional Bunaken telah dilakukan sebelumnya, meskipun terbatas pada persepsi dan motivasi wisatawan nusantara [8]. Mempertimbangkan hal tersebut, maka penelitian ini akan menganalisis sentimen wisatawan mancanegara menggunakan dataset dari website Tripadvisor. Disisi lain, Bagindo et al. menunjukkan bahwa bisnis model Taman Nasional Bunaken memiliki segmentasi pasar mancanegara dan nusantara dimana arus kas (revenue stream) diperoleh dari tarif masuk, donasi, Anggaran Pendapatan dan Belanja Negara (APBN) dan Anggaran Pendapatan dan Belanja Daerah (APBD) [9]. Disisi lain, Santoso et al. menunjukkan bahwa terdapat 17 pemangku kepentingan yang berperan penting dalam tatakelola Taman Nasional Bunaken, yakni : pemerintah pusat dan pemerintah daerah; kelompok swasta; kelompok akademisi; kelompok masyarakat dan kelompok lainnya [10]. Hal ini menunjukkan bahwa optimalisasi manajemen Taman Nasional Bunaken membutuhkan koordinasi dan kerjasama untuk meminimalisir potensi konflik sektoral yang menghambat aktivitas pariwisata. Koordinasi intensif dalam implementasi program maupun sinkronisasi kegiatan antar sektor menjadi kunci keberhasilan manajemen Taman Nasional Bunaken. Berdasarkan hal tersebut, penelitian ini menawarkan gagasan untuk memberikan rekomendasi tatakelola Taman Nasional Bunaken dari perspektif wisatawan berdasarkan data ulasan di website Tripadvisor.

Penelitian ini menawarkan gagasan untuk mengidentifikasi dan menganalisis konteks kepariwisataan melalui pendekatan analisis sentimen menggunakan kerangka kerja Cross Industry Standard Process for Data Mining (CRISP-DM) agar dapat diketahui sentimen wisatawan terkait produk dan layanan yang diperoleh selama berwisata ke Bunaken. Singgalen menunjukkan bahwa hasil analisis sentimen berdasarkan kerangka kerja CRISP-DM merekomendasikan model yang relevan dengan konteks pengembangan destinasi wisata agar memenuhi ekspektasi wisatawan terkait kualitas produk dan layanan [11]. Disisi lain, Singgalen menunjukkan bahwa model yang direkomendasikan dapat disesuaikan dengan data ulasan wisatawan yang diperoleh dari website Tripadvisor [12]. Adapun, algoritma yang dijadikan model dapat diseleksi berdasarkan performa terbaik dari nilai akurasi, presisi, recall, f-measure, Area Under Curve (AUC), dan t-test [13]. Hal ini menunjukkan bahwa analisis sentimen diperlukan untuk mengidentifikasi dan menganalisis tingkat kepuasan wisatawan ketika menggunakan produk dan layanan selama berwisata. Dengan demikian, penelitian tentang analisis sentimen memiliki peluang untuk berkontribusi bagi optimalisasi manajemen destinasi wisata khususnya Taman Nasional Bunaken.

Beberapa penelitian terdahulu menunjukkan sejumlah keunggulan menggunakan kerangka kerja CRISP-DM dalam proses text mining untuk analisis sentimen. Christanto dan Singgalen menunjukkan bahwa implementasi analisis sentimen menggunakan algoritma Naïve Bayes Classifier, k-NN, dan Support Vector Machine (SVM) dapat merekomendasikan strategi optimalisasi manajemen hotel untuk mengoptimalkan kualitas produk dan layanan yang kontekstual [14]. Selain itu, Singgalen menunjukkan bahwa rekomendasi untuk meningkatkan kualitas produk dan jasa berperan penting dalam mempertahankan citra positif perusahaan agar meningkatkan daya beli konsumen [15]. Arifiyanti et al. menunjukkan bahwa analisis sentimen di bidang pariwisata menjadi kajian yang penting untuk dilakukan secara intensif guna mengidentifikasi kebutuhan atau preferensi wisatawan dari waktu ke waktu [16]. Syahid berpendapat bahwa kerangka kerja CRISP-DM dapat digunakan untuk memahami konteks bisnis, mengenal data yang digunakan, serta merekomendasikan model yang tepat sehingga bermanfaat bagi keberlanjutan perusahaan atau organisasi [17]. Hal ini menunjukkan bahwa CRISP-DM memberikan gambaran tentang manfaat dari analisis sentimen terhadap kepentingan organisasi atau perusahaan. Dengan demikian, luaran penelitian ini dapat menghasilkan model analisis sentimen berbasis data teks yang diperoleh dan dikelola dari website Tripadvisor dengan menyesuaikan konteks pengelolaan Taman Nasional Bunaken.

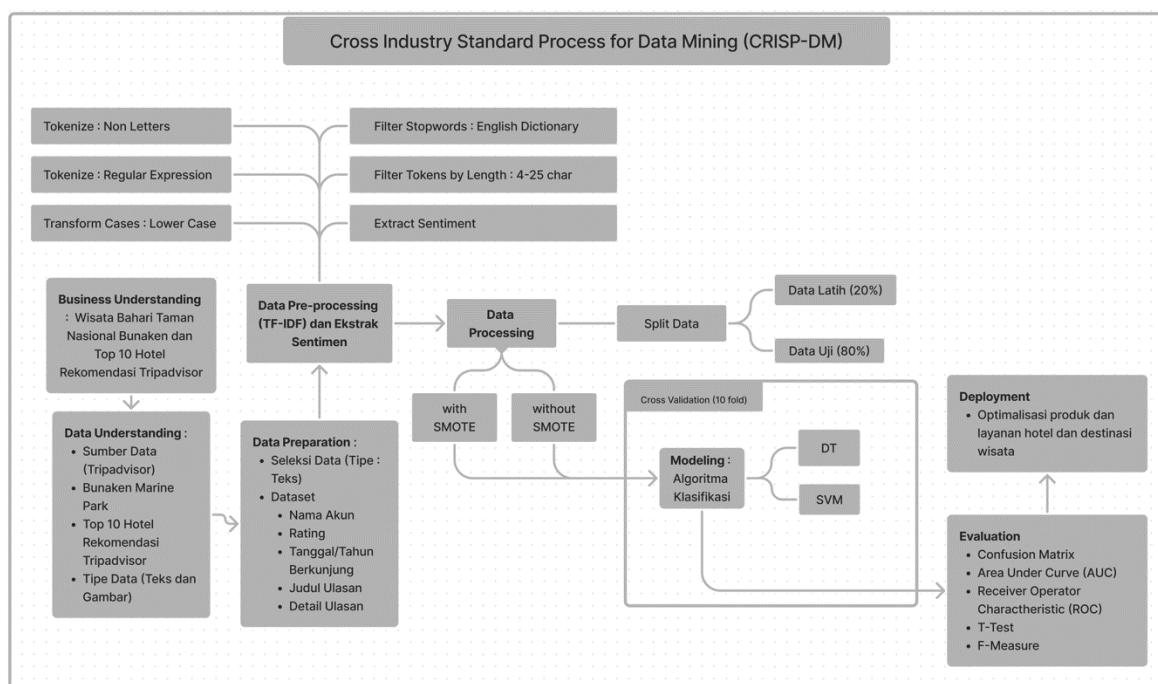
Pendekatan Cross Industry Standard Process for Data Mining (CRISP-DM) terbagi menjadi enam tahapan sebagai berikut : memahami konteks bisnis (business understanding); memahami data yang akan dikelola (data understanding); menyiapkan data (data preparation); membuat model (modeling); mengevaluasi model (evaluation); dan menerapkannya di organisasi atau perusahaan sebagai bentuk inovasi (deployment). Syahid dan

Mahdiana menunjukkan bahwa analisis sentimen menggunakan berbagai algoritma sebagai model dapat diperbandingkan untuk memperoleh petunjuk tentang performa algoritma yang paling optimal agar dapat digunakan dalam mengklasifikasi data dari berbagai platform digital termasuk media sosial [17]. Selain itu, Fadli et al. menunjukkan bahwa CRISP-DM untuk analisis sentimen di lingkungan pendidikan dapat meningkatkan kualitas tatakelola organisasi terutama dalam meningkatkan layanan di bidang kemahaisswaan [18]. Hal ini menunjukkan bahwa CRISP-DM juga dapat digunakan dalam mengidentifikasi dan menganalisis sentimen wisatawan untuk kepentingan manajemen destinasi wisata. Dengan demikian penelitian ini dapat menghasilkan rekomendasi dan model yang relevan dengan pengelolaan Taman Nasional Bunaken di Sulawesi Utara, Indonesia.

## 2. METODOLOGI PENELITIAN

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

Metode yang digunakan dalam menganalisis sentimen wisatawan terhadap wisata bahari Taman Nasional Bunaken ialah Cross Industry Standard Process for Data Mining (CRISP-DM). Terdapat enam tahapan dalam CRISP-DM sebagai berikut business understanding, data understanding, data preparation, data preparation, modeling, evaluation dan deployment. Pada tahap business understanding, dilakukan studi literatur tentang pemangku kepentingan yang terlibat dalam pengelolaan Taman Nasional Bunaken, business model serta faktor-faktor yang memengaruhi minat dan motivasi berkunjung. Selanjutnya, pada tahap data understanding, dilakukan pemetaan data ulasan yang perlu dikumpulkan berdasarkan hasil studi literatur, sehingga penelitian ini membutuhkan data ulasan wisatawan terhadap Bunaken Marine Park pada website Tripadvisor. Adapun, proses scraping data teks dari website Tripadvisor menggunakan aplikasi webharvy. Data ulasan yang dikumpulkan telah diseleksi berdasarkan nama akun, rating, tipe kunjungan, tanggal berkunjung, judul ulasan, dan detail ulasan. 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%). 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 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). 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 sentimen wisatawan terhadap pengelolaan Taman Nasional Bunaken hingga top 10 hotel yang direkomendasi Tripadvisor. 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.

## 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 :

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

$$W_{ij} = TF_{ij} \times \log\left(\frac{D}{DF_{ij}}\right) \quad (2)$$

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) [19]. 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. Barro et al. berpendapat bahwa 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 [20]. 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 [21]. 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 peubah 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 bahari Taman Nasional Bunaken 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 [22]. Decision Tree memiliki beberapa algoritma salah satunya Iterative Dychotomizer 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, pi 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, Si 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 [23]. 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 [24]. Karim menjelaskan bahwa konsep SVM menitikberatkan pada risk minimization, yaitu estimasi fungsi dengan cara meminimalisir batas dari generalization error, sehingga SVM mampu mengatasi overfitting [25]. 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 [26]. 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 [27]. 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

Analisis sentimen merupakan pendekatan yang dapat digunakan untuk membahas secara komprehensif kepuasan konsumen terhadap produk atau layanan yang dikonsumsi. Amsal et al. menunjukkan bahwa analisis sentimen merupakan bagian dari upaya untuk mengidentifikasi dan menganalisis kepuasan konsumen terhadap produk dan jasa, berdasarkan data teks yang diperoleh dari ulasan konsumen di berbagai platform digital [28]. Almuyasaar et al. berpendapat bahwa faktor-faktor yang memengaruhi kepuasan konsumen ketika melakukan transaksi di platform e-commerce ialah sebagai berikut : kinerja, ketersediaan, daya tahan, estetika, keandalan, pemeliharaan, logistik, dukungan, layanan pelanggan, pelatihan, pengiriman, penagihan, perbaikan, pemasaran, garansi, dan biaya siklus hidup [29]. Hal ini menunjukkan bahwa analisis sentimen merupakan bagian dari evaluasi strategi pemasaran produk dan jasa yang dapat digunakan sebagai rekomendasi untuk meningkatkan volume penjualan. Dengan demikian, kajian tentang analisis sentimen perlu ditingkatkan dari sisi kuantitas berdasarkan perspektif keilmuan dan bidang bisnis. Mempertimbangkan hal tersebut maka pembahasan dalam penelitian ini menitikberatkan pada analisis sentimen wisatawan terhadap produk dan jasa di Taman Nasional Bunaken, yang secara spesifik membahas tentang destinasi dan hotel.

Perkembangan studi tentang analisis sentimen menunjukkan bahwa data yang digunakan dapat berupa teks yang diperoleh dari kolom ulasan produk pada website e-commerce, maupun platform yang menyediakan transaksi daring. Yumarlin et al. menunjukkan bahwa pengguna platform digital untuk aktivitas pembelian atau transaksi daring, memiliki kewenangan untuk mengulas produk dan layanan yang dapat dijadikan sebagai rekomendasi untuk meningkatkan kualitas produk hingga layanan penjualan [30]. Pamungkas dan Februriyanti menunjukkan bahwa proses analisis sentimen konsumen terhadap produk e-commerce terdiri dari tahap pengumpulan (scraping), pengolahan data (pre-processing and modeling) menggunakan algoritma tertentu, tahap evaluasi dan pemanfaatan data (evaluation and deployment) [31]. Hal ini menunjukkan bahwa analisis sentimen menggunakan data ulasan dengan format data teks untuk diklasifikasikan menggunakan algoritma tertentu, hingga menghasilkan model yang relevan dengan kepentingan bisnis. Dalam pembahasan penelitian ini, kerangka kerja CRISP-DM digunakan untuk menganalisis konteks bisnis dan data yang digunakan dalam proses analisis sentimen.

Intensitas penggunaan teknologi informasi dalam berbagai layanan transaksi untuk kepentingan bisnis di Indonesia, telah meningkatkan partisipasi pengguna aktif (user). Rammadhan et al. menunjukkan bahwa peningkatan pengguna aktif teknologi di platform e-commerce telah memantik pendekatan baru dalam identifikasi dan analisis sentimen melalui data ulasan konsumen yang dikenal dengan Knowledge Discovery in Database (KDD) yang menarik pemasar untuk menggunakan kerangka serupa seperti CRISP-DM [32]. Amara et al. menunjukkan bahwa proses analisis sentimen menghasilkan luaran yang sama dengan implementasi kerangka Service Quality (SQ) dan Importance Performance Analysis (IPA) dimana keduanya menganalisis ulasan konsumen terkait dengan kepuasan terhadap produk dan layanan yang dipasarkan melalui berbagai macam platform digital [33]. Hal ini menunjukkan bahwa proses identifikasi, klasifikasi, dan analisis kepuasan konsumen dapat menggunakan pendekatan analisis sentimen berbasis CRISP-DM untuk merekomendasikan model pemasaran yang sesuai dengan karakteristik bisnis. Dengan demikian, pembahasan dalam penelitian ini menggunakan data ulasan wisatawan terhadap destinasi Bunaken Marine Park dan top 10 hotel yang direkomendasikan oleh website Tripadvisor, sebagaimana Tabel 1 berikut ini.

**Tabel 1.** Jumlah Data Teks Ulasan Wisatawan Bunaken Marine Park dan Top 10 Hotel Rekomendasi Tripadvisor

Data	Dataset	Jumlah Data (Sebelum Filter)	Jumlah Data (Setelah Filter)	Link Tripadvisor
Destinasi	Bunaken National Marine Park	871	398	<a href="https://www.tripadvisor.com/Attraction_Review-g1071666-d583306-Reviews-Bunaken_National_Marine_Park-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html">https://www.tripadvisor.com/Attraction_Review-g1071666-d583306-Reviews-Bunaken_National_Marine_Park-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html</a>
	Raja Laut Dive Resort	250	149	<a href="https://www.tripadvisor.com/Hotel_Review-g1071666-d1744754-Reviews-Raja_Laut_Dive_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html">https://www.tripadvisor.com/Hotel_Review-g1071666-d1744754-Reviews-Raja_Laut_Dive_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html</a>
Hotel	Bastianos Bunaken Dive Resort	329	166	<a href="https://www.tripadvisor.com/Hotel_Review-g1071666-d615408-Reviews-Bastianos_Bunaken_Dive_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html">https://www.tripadvisor.com/Hotel_Review-g1071666-d615408-Reviews-Bastianos_Bunaken_Dive_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html</a>
	Bunaken Oasis Dive Resort and Spa	123	105	<a href="https://www.tripadvisor.com/Hotel_Review-g1071666-d10307120-Reviews-Bunaken_Oasis_Dive_Resort_and_Spa-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html">https://www.tripadvisor.com/Hotel_Review-g1071666-d10307120-Reviews-Bunaken_Oasis_Dive_Resort_and_Spa-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html</a>

Data	Dataset	Jumlah Data (Sebelum Filter)	Jumlah Data (Setelah Filter)	Link Tripadvisor
The Village Bunaken		287	161	<a href="https://www.tripadvisor.com/Hotel_Review-g1071666-d1438211-Reviews-The_Village_Bunaken-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html">https://www.tripadvisor.com/Hotel_Review-g1071666-d1438211-Reviews-The_Village_Bunaken-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html</a>
Two Fish Dive Bunaken		464	345	<a href="https://www.tripadvisor.com/Hotel_Review-g1071666-d1315549-Reviews-Two_Fish_Divers_Bunaken-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html">https://www.tripadvisor.com/Hotel_Review-g1071666-d1315549-Reviews-Two_Fish_Divers_Bunaken-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html</a>
Happy Gecko Dive Resort		147	81	<a href="https://www.tripadvisor.com/Hotel_Review-g1071666-d1742764-Reviews-Happy_Gecko_Dive_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html">https://www.tripadvisor.com/Hotel_Review-g1071666-d1742764-Reviews-Happy_Gecko_Dive_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html</a>
Raja Laut Coral View Bunaken		47	29	<a href="https://www.tripadvisor.com/Hotel_Review-g1071666-d10681526-Reviews-Raja_Laut_Coral_View-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html">https://www.tripadvisor.com/Hotel_Review-g1071666-d10681526-Reviews-Raja_Laut_Coral_View-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html</a>
Cha Cha Nature Resort		206	150	<a href="https://www.tripadvisor.com/Hotel_Review-g1071666-d665593-Reviews-Bunaken_Cha_Cha_Nature_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html">https://www.tripadvisor.com/Hotel_Review-g1071666-d665593-Reviews-Bunaken_Cha_Cha_Nature_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html</a>
Panorama Diving Resort Bunaken		218	99	<a href="https://www.tripadvisor.com/Hotel_Review-g1071666-d3313907-Reviews-Panorama_Diving_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html">https://www.tripadvisor.com/Hotel_Review-g1071666-d3313907-Reviews-Panorama_Diving_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html</a>
Sea Garden Resort		173	110	<a href="https://www.tripadvisor.com/Hotel_Review-g1071666-d1573422-Reviews-Bunaken_SeaGarden_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html">https://www.tripadvisor.com/Hotel_Review-g1071666-d1573422-Reviews-Bunaken_SeaGarden_Resort-Bunaken_Island_Manado_North_Sulawesi_Sulawesi.html</a>

Tabel 1 merupakan jumlah data ulasan dan link destinasi, top 10 hotel yang direkomendasikan oleh Website Tripadvisor. Dalam proses pengumpulan data ulasan, digunakan aplikasi webharvy dengan konfigurasi nama akun, rating, tipe kunjungan, tanggal berkunjung, judul ulasan, dan detail ulasan. Berdasarkan hasil pengumpulan dan seleksi data, diperoleh 398 data destinasi yang siap diproses ke tahap pemodelan, dari 871 data ulasan di website Tripadvisor. Sementara itu, data teks untuk top 10 hotel digabung berdasarkan jumlah data yang direkomendasikan Tripadvisor. Terdapat 1793 data ulasan dari 2244 data ulasan (10 hotel) yang akan diproses ke tahap pembersihan. Selanjutnya data teks dibersihkan dan dirapikan 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 terhadap Bunaken Marine Park

Review
<p>June 18, my family and I traveled to bunaken via the marina plaza dock and rented a catamaran boat in a small office in the marina plaza dock We plan to snorkel in Bunaken and siladen island We rented a catamaran at the marina dock From the initial price of idr 1.4 million for 15 people to Bunaken, we must add idr 300k for journey to siladen island At the boat charter office i made it clear that we will snorkel in siladen too When we arrived in Bunaken, we pay the taxes to the area bunaken IDR 5k rupiah per person Then we rented snorkeling equipment in bunaken for 150k rupiah each set.</p> <p>I began to suspect because the boat motorist insisted that snorkeling equipment should be hired a complete set, while internet references mentioned may be rented as needed Then I also have to pay two boat mtororist that becomes a snorkeling tour guide in bunaken by hiring additional 2 sets of snorkeling equipment, in total I hired 7 sets of snorkeling equipment the weirdness grew to be because after 1 hour of water play and snorkeling in bunaken, the motorist said we had to leave immediately to siladen to avoid low tide. we agreed because in the hope we will go snorkeling in siladen it turns out the snorkeling equipment we rent 150rb only used 1 hour because we must return it to the owner because they said it is not allowed to be brought to siladen until here I still follow the dirty game by the motorist we finally went to siladen and arrived around 11.30 finally in siladen we can only swim around the dock and walk on the island of siladen when we were enjoyinging swimming,</p>

motorist said we had to go home for fear of rain but it's still 1 pm so just 1 hour we played in the water and the motorist already lifted the ladder access to the boat we finally went back to manado to the marina plaza dock and arrived at 2 PM noon a very unpleasant and unsatisfactory experience not in accordance with the cost of IDR 3 million more we spend I DO NOT RECOMMEND TO RENTING ANY BOAT IN MARINA PLAZA BECAUSE YOU CAN BE FOOLED BY THEM NEVER RENT BOAT AT MARIZA PLAZA DOCK IT IS useless

**String Scoring**

clear (0.41) pay (-0.10) suspect (-0.31) pay (-0.10) weirdness (-0.23) play (0.36) leave (-0.05) avoid (-0.31) low (-0.28) agreed (0.28) hope (0.49) dirty (-0.49) fear (-0.56) played (0.36) unpleasant (-0.54) recommend (0.38) fooled (-0.41) useless (-0.46)

Sentiment	Score	Negativity	Positivity	Uncovered Tokens	Tokens
<b>Negatif</b>	-1,6	3,8	2,3	386	404

**Review**

We stayed at Bunaken for 6 days. The beauty of the underwater world here is simple incredible! Each day, we delighted ourselves snorkelling for hours at various sites around the islands and found that we liked best the Bunaken Timur reef which stretches the whole length of Pantai Pangalisang. Many resorts are located along the entire beach front. It is amazing to be able to access the best corals in the world within a few minutes swim from the beach such that you don't really need to hire a boat out to enjoy this reef if you stayed at any of these resorts. The beach is quite a long one though. The sheer volume of beautiful corals in good condition and the terrific diversity leaves one nothing short of astounded. We marvelled at the great myriad of species of soft corals and unique marine life like common lionfish, spotfin lionfish, various types of boxfish, anemone fish, unicorn fish, pipefish, butterfly fish, angelfish, parrotfish, pipefish, needlefish, damsels, wrasse, turtles and countless other species.. We loved the colourful orange and blue starfish lying lazily in abundance all over the seagrass bed and corals. Your resort may also offer dolphin tours to the north. We were there in August and enjoyed taking a boat to the northern swimming grounds of the tuna to see the dolphins. A unique experience having groups or dolphins gracefully leap out of water as they race around the waters. We stopped by Siladen Island to snorkel after the trip and that was also a good place. It is easy to get to Bunaken. Fly in to Manado from various places like Jakarta, Surabaya, Makassar, Singapore etc. The harbour is 30 min to 1 hour from the Sam Ratulangi airport depending on traffic. You can stay the night at hotels near the harbour like Celebes Hotel or arrange for a boat on same day if you arrive early enough. It only takes 40 min boat ride from Manado to get there. Hotels in Bunaken may arrange your boat transfer but this is usually more pricey than if you simply go to the harbour and find a boat yourself. There are many available boatmen waiting at the harbour. I negotiated a one way trip at IDR400,000 and arranged for the boatman to pick me up for the return trip 6 days later at the same price. (The boatman's opening price was IDR1,200,000 but when I told him I know the going rate, he reduced it to IDR400,000!). I read reports of how rubbish would float over from the mainland to the snorkelling sites. I encountered no rubbish at all at the long stretch of Bunaken Timur Reef. Perhaps it's the angle of this area which is a little away from the mainland. But I noticed the stretches of the island that faces the mainland tend to have a little bit of stuff in the water but it was not really a big problem. We did not pay any national park fee as our resort refused to collect the same, maintaining the money was not well spent. Although it was a long way to get here with connecting flights for us, the beauty of underwater world was just so amazing and worth the time and effort and money making the trip Highly recommended!

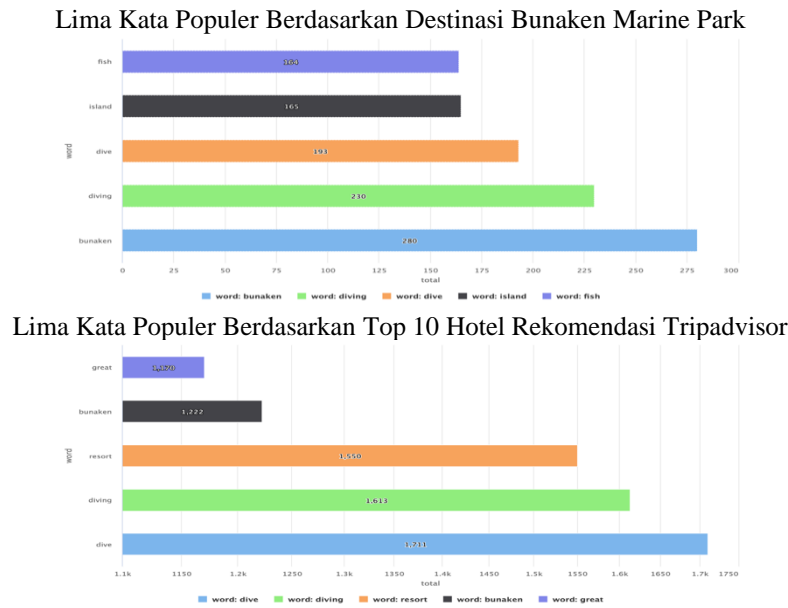
**String Scoring**

beauty (0.72) delighted (0.59) liked (0.46) best (0.82) amazing (0.72) best (0.82) enjoy (0.56) beautiful (0.74) good (0.49) terrific (0.54) astounded (0.46) great (0.79) like (0.38) loved (0.74) lying (-0.62) enjoyed (0.59) gracefully (0.62) stopped (-0.23) good (0.49) easy (0.49) like (0.38) like (0.38) no (-0.31) problem (-0.44) pay (-0.10) refused (-0.31) well (0.28) beauty (0.72) amazing (0.72) worth (0.23) recommended (0.21)

Sentiment	Score	Negativity	Positivity	Uncovered Tokens	Tokens
<b>Positif</b>	11,9	2	13,9	585	616

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 Taman Nasional Bunaken, 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 sebagaimana gambar 3 berikut.



**Gambar 2.** Lima Kata Populer dalam Data Ulasan Wisatawan terhadap Destinasi dan Hotel

Gambar 2 merupakan lima kata populer yang dapat divisualisasikan dalam bentuk diagram bar horizontal untuk destinasi wisata Bunaken Marine Park dan top 10 hotel yang direkomendasi Tripadvisor berdasarkan wilayah di Bunaken. Berdasarkan hasil identifikasi, lima kata yang paling sering muncul dalam ulasan wisatawan untuk destinasi wisata Bunaken Marine Park, terdapat 164 kata fish, 165 kata island, 193 kata dive, 230 diving, 280 kata bunaken. Hal ini menunjukkan bahwa faktor pendorong kunjungan wisata ke Bunaken Marine Park ialah keindahan alam bawah laut termasuk keanekaragaman jenis hewan laut, keindahan alam daerah kepulauan, dan aktivitas menyelam. Dengan demikian para pemangku kepentingan yang berperan penting dalam pengelolaan Taman Nasional Bunaken, perlu memperhatikan keberlanjutan ekosistem agar tetap berkelanjutan, serta tidak mengalami degradasi atau kerusakan akibat aktivitas wisata. Disisi lain, hasil identifikasi limat kata yang paling sering muncul dalam ulasan wisatawan untuk top 10 hotel yang direkomendasikan Tripadvisor ialah, 1170 kata great, 1222 kata bunaken, 1550 kata restort, 1613 kata diving, dan 1711 kata dive. Hal ini menunjukkan bahwa karakteristik wisatawan yang memiliki motif berwisata ke Taman Wisata Nasional Bunaken memiliki intensi untuk menggunakan akomodasi resort di sekitar Bunaken, serta memiliki mobilitas dan fasilitas pendukung aktivitas menyelam. Hal ini dapat menjadi rekomendasi bagi para pengelola jasa akomodasi agar menyiapkan fasilitas pendukung bagi wisatawan yang hendak mengunjungi Taman Nasional Bunaken.

Pengembangan destinasi wisata memiliki hubungan dengan pengelolaan akomodasi serta amenitas. Frederick dan Kurniawati menunjukkan bahwa upaya untuk meningkatkan kepuasan wisatawan ketika berwisata di Taman Nasional Bunaken ialah dengan menyediakan fasilitas pariwisata yang sesuai dengan kebutuhan [34]. Friliyantini et al. menunjukkan bahwa salah satu upaya untuk meningkatkan partisipasi masyarakat dalam pembangunan pariwisata di Bunaken ialah dengan mengembangkan Usaha Kecil dan Menengah masyarakat lokal yang berhubungan dengan jasa akomodasi dan transportasi sebagai pendukung pariwisata [35]. Hal ini menunjukkan bahwa pengembangan Taman Nasional Bunaken perlu memperhatikan aspek sosial dan lingkungan, selain ekonomi. Dengan demikian, upaya untuk meningkatkan kepuasan wisatawan ketika berkunjung ke Bunaken dapat didukung oleh berbagai kondisi positif di bidang sosial-budaya, ekonomi, dan lingkungan. Selain itu, hasil pengolahan 398 data destinasi Taman Nasional Bunaken, dan 1793 data ulasan terhadap Top 10 Hotel yang direkomendasikan oleh Tripadvisor, menunjukkan hasil yang berbeda-beda sebagaimana confusion matrix dapat dilihat pada tabel 3 dan 4 berikut.

**Tabel 3.** Confusion Matrix Data Ulasan Bunaken Marine Park

Algoritma	Bunaken Marine Park										
	TP	FN	FP	TN	Accuracy	Precision	Recall	F-Measure	AUC	t-Test	Keterangan
DT	29	6	7	30	97,87%	97,74%	98,04%	97,85%	97,7%	97,7%	with SMOTE

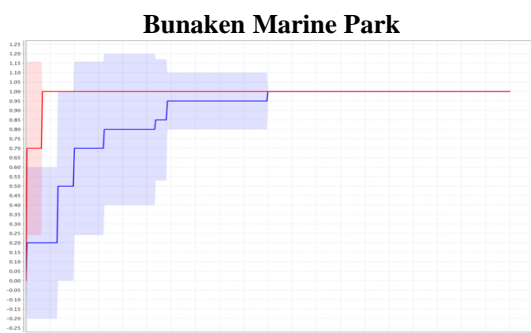
Bunaken Marine Park											
Algoritma	TP	FN	FP	TN	Accurac y	Precisio n	Recall	F- Measur e	AUC	t-Test	Keteranga n
DT	30							unknow n			
	1	11	5	0	94,97%	0,00%	0,00%		50%	95%	without SMOTE
SVM	30			30			100%	100%			with SMOTE
	6	0	0	6	100%	100%			100%	100%	without SMOTE
SVM	0	0	8	6	89,31%	89,31%	100%	94,35%	80,7 %	96,5%	without SMOTE

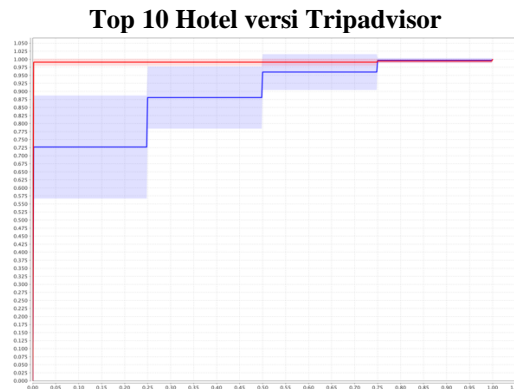
Tabel 3 menunjukkan bahwa hasil evaluasi performa algoritma Decision Tree (DT) menggunakan operator SMOTE memberikan nilai akurasi 97,87%, nilai presisi 97,74%, nilai recall 98,04%, f-measure 97,85%, nilai AUC 97,7%, nilai t-Test 97,7%. Sementara itu, performa algoritma Decision Tree (DT) tanpa menggunakan operator SMOTE memberikan nilai akurasi 94,97%, namun untuk nilai presisi nilai recall ialah 0,00%, dan unknown untuk f-measure. Adapun nilai AUC sebesar 50%, dan t-Test 95%. Hal ini menunjukkan bahwa hasil pengolahan dataset destinasi wisata Taman Nasional Bunaken harus menggunakan operator SMOTE untuk mendapatkan hasil yang lebih optimal. Disisi lain, hasil evaluasi performa algoritma Support Vector Machine (SVM) menggunakan operator SMOTE memberikan nilai akurasi, presisi, recall, f-measure, AUC, dan t-Test 100%. Berbeda halnya dengan performa algoritma SVM tanpa SMOTE dimana nilai akurasi 89,31%, nilai presisi 89,31%, nilai recall 100%, nilai f-measure 94,35%, nilai AUC 80,7% dan nilai t-Test 96,5%. Apabila dibandingkan performa terbaik antara DT dan SVM (menggunakan SMOTE dan tanpa menggunakan SMOTE) maka model terbaik diperoleh dari hasil klasifikasi SVM dengan menggunakan operator SMOTE. Selain itu dapat diketahui performa algoritma DT dan SVM (menggunakan SMOTE dan tanpa menggunakan SMOTE) dalam klasifikasi sentimen untuk dataset top 10 hotel yang direkomendasikan Tripadvisor, pada tabel berikut.

Tabel 4. Confusion Matrix Data Ulasan Top 10 Hotel Rekomendasi Tripadvisor

Data Gabungan Top 10 Hotel Rekomendasi Tripadvisor											
Algoritma	TP	FN	FP	TN	Accurac y	Precisio n	Recall	F- Measure	AUC	t- Test	Keterangan
DT	107	20					91,22			90,4	
	1	2	4	873	90,42%	99,54%	%	89,42%	50%	%	with SMOTE
DT				106			98,97		54,5	95,8	without SMOTE
	4	11	36	4	95,78%	96,73%	%	97,84%	%	%	
SVM	107						92,83		100	96,4	with SMOTE
	5	77	0	998	96,42%	100%	%	96,27%	%	%	
SVM				107					89,8	96,4	without SMOTE
	0	0	40	5	96,41%	96,41%	100%	98,17%	%	%	

Tabel 4 menunjukkan bahwa hasil evaluasi performa algoritma Decision Tree (DT) menggunakan operator SMOTE memberikan nilai akurasi 90,42%, nilai presisi 99,54%, nilai recall 91,22%, nilai f-measure 89,42%, nilai AUC 50%, nilai t-Test 90,7%. Sementara itu, performa algoritma Decision Tree (DT) tanpa menggunakan operator SMOTE memberikan nilai akurasi 95,78%, nilai presisi 96,73%, nilai recall 98,97%, nilai f-measure 97,84%, nilai AUC 54,5 dan nilai t-Test 95,8%. Disisi lain, hasil evaluasi performa algoritma Support Vector Machine (SVM) menggunakan operator SMOTE memberikan nilai akurasi 96,42%, nilai presisi 100%, nilai recall 92,83%, nilai f-measure 96,27%, nilai AUC 100%, dan nilai t-Test 96,4%. Berbeda halnya dengan performa algoritma SVM tanpa SMOTE dimana nilai akurasi 96,41%, nilai presisi 96,41%, nilai recall 100%, nilai f-measure 98,17%, nilai AUC 89,8% dan nilai t-Test 96,4%. Apabila dibandingkan performa terbaik antara DT dan SVM (menggunakan SMOTE dan tanpa menggunakan SMOTE) maka model terbaik diperoleh dari hasil klasifikasi SVM dengan menggunakan operator SMOTE. Hal ini menunjukkan bahwa algoritma SVM menggunakan operator SMOTE mampu meningkatkan performa algoritma. Meskipun demikian, algoritma DT juga menunjukkan performa yang baik sebagaimana nilai ROC dalam gambar berikut.





**Gambar 3.** ROC Bunaken Marine Park dan Top 10 Hotel rekomendasi Tripadvisor

Gambar 3 merupakan nilai ROC dari pengolahan dataset destinasi Bunaken Marine Park, dan dataset Top 10 Hotel yang direkomendasikan Tripadvisor yang menunjukkan bahwa algoritma DT dan SVM dapat digunakan untuk klasifikasi data berdasarkan kelas positif dan negatif. Meskipun demikian, jumlah data yang dikelola perlu diseimbangkan dengan menggunakan operator SMOTE agar meningkatkan kinerja algoritma. Beberapa studi terdahulu menunjukkan bahwa nilai ROC dan AUC dapat dijadikan sebagai pertimbangan untuk memilih model yang tepat dalam pengelolaan data. Pravina et al. menunjukkan bahwa algoritma SVM mampu menunjukkan performa yang baik dalam klasifikasi data berdasarkan kelas sentimen [36]. Fitriana dan Sibaroni menunjukkan bahwa SVM juga dapat digunakan dalam analisis sentimen pengguna jasa aplikasi yang diunggah melalui platform digital media sosial [24]. Hal ini menunjukkan bahwa SVM dalam beberapa studi kasus analisis sentimen selalu menunjukkan performa yang optimal. Dengan demikian penelitian ini juga menunjukkan bahwa SVM dengan operator SMOTE dapat digunakan sebagai model sebagaimana kerangka kerja CRISP-DM. Dengan demikian, rekomendasi pengembangan destinasi wisata Taman Nasional Bunaken dan jasa akomodasi perlu memperhatikan kebutuhan wisatawan terkait dengan fasilitas, serta sarana dan prasarana pendukung aktivitas wisata bahari.

#### 4. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa SVM dengan operator SMOTE dapat digunakan sebagai model sebagaimana kerangka kerja CRISP-DM. Berdasarkan hasil pengolahan 398 data destinasi Taman Nasional Bunaken, dan 1793 data ulasan terhadap top 10 hotel yang direkomendasikan oleh Tripadvisor, algoritma SVM menunjukkan performa yang lebih baik. Dalam konteks data destinasi Taman Nasional Bunaken, performa algoritma Support Vector Machine (SVM) menggunakan operator SMOTE mampu menghasilkan nilai akurasi, presisi, recall, f-measure, AUC, dan t-Test sebanyak 100%. Dalam konteks, dataset top 10 hotel yang direkomendasikan oleh Tripadvisor, algoritma Support Vector Machine (SVM) menggunakan operator SMOTE memberikan nilai akurasi 96,42%, nilai presisi 100%, nilai recall 92,83%, nilai f-measure 96,27%, nilai AUC 100%, dan nilai t-Test 96,4%. Selain itu, berdasarkan hasil identifikasi lima kata yang paling sering muncul dalam ulasan wisatawan untuk destinasi wisata Bunaken Marine Park, terdapat 164 kata fish, 165 kata island, 193 kata dive, 230 diving, 280 kata bunaken. Selanjutnya, hasil identifikasi lima kata yang paling sering muncul dalam ulasan wisatawan untuk top 10 hotel yang direkomendasikan Tripadvisor ialah, 1170 kata great, 1222 kata bunaken, 1550 kata restort, 1613 kata diving, dan 1711 kata dive. Hal ini dapat menjadi rekomendasi bagi para pengelola jasa akomodasi agar menyiapkan fasilitas pendukung bagi wisatawan yang hendak mengunjungi Taman Nasional Bunaken.

#### REFERENCES

- [1] R. D. Priyatna, F. Kerebungu, and Y. D. A. Santie, "Dampak Ekowisata Taman Laut Bunaken Berdasarkan Perspektif Masyarakat," *Indones. J. Soc. Sci. Educ.*, vol. 2, no. 2, pp. 58–68, 2022.
- [2] F. G. Tular, H. J. Lawalata, and M. L. S. Taulu, "Study of Mangrove Vegetation Community at Bunaken National Park Area, Teling Village," *Indones. Biodivers. J.*, vol. 3, no. 3, pp. 20–29, 2022.
- [3] E. I. Lukman, A. T. Utomo, and E. W. Handoyo, "Fasilitasi Bantuan Usaha Ekonomi Kreatif Kelompok Masyarakat Binaan Balai Taman Nasional Bunaken di Pulau Nain," *Dharma J. Pengabd. Masy.*, vol. 3, no. 1, pp. 39–59, 2022.
- [4] W. D. Aryawan, M. H. N. Aliffrananda, D. Utama, Y. A. Hermawan, and Hasanudin, "Design of a Semi-Submersible Tourism Ship for Bunaken Underwater Recreation in Manado, Indonesia," *Kapal J. Ilmu Pengetah. dan Teknol. Kelaut.*, vol. 19, no. 3, pp. 151–164, 2022.
- [5] D. J. Namoua, A. S. Wantasen, K. I. F. Kondoy, R. C. Kepel, F. S. I. Menajang, and W. Pelle, "Carbon Absorption in Seagrasses in Tongkaina Coastal Waters, Bunaken District, Manado City, North Sulawesi," *J. Ilm. PLATAX*, vol. 10, no. 2, pp. 433–440, 2022, doi: 10.35800/jip.v10i2.43485.
- [6] E. S. Pomantow, M. A. Langi, and F. B. Saroinsong, "Strategi Penanggulangan Gangguan Kawasan Konservasi Di Taman Nasional Bunaken ( Studi Kasus Di Pulau Mantehage )," *Agri-Sosioekonomi*, vol. 18, no. 3, pp. 775–784, 2022.
- [7] M. H. Sahri, P. A. Pangemanan, and F. B. Saroinsong, "Analisis Kesesuaian Ekowisata Pantai Dan Bahari Di Pesisir

- Perairan Tanjung Kelapa Desa Poopoh Taman Nasional Bunaken,” *Agri-Sosioekonomi*, vol. 18, no. 3, pp. 751–758, 2022, [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/jisep/article/view/44550>
- [8] E. W. Handoyo, H. D. Walangitan, and P. A. Pangemanan, “Persepsi dan Motivasi Pengunjung Nusantara Taman Nasional Bunaken,” *Agri-Sosioekonomi*, vol. 18, no. 2, pp. 449–460, 2022.
- [9] M. P. Bagindo, B. Sanim, and T. Saptono, “Model Bisnis Ekowisata di Taman Nasional Laut Bunaken dengan Pendekatan Business Model Canvas,” *Manaj. IKM*, vol. 11, no. 1, pp. 80–88, 2016, doi: 10.29244/mikm.11.1.80-88.
- [10] H. Santoso, H. Kartodihardjo, and R. Soekmadi, “Peran dan Kebutuhan Pemangku Kepentingan dalam Tata Kelola Pariwisata di Taman Nasional Bunaken, Sulawesi Utara,” *J. Penelit. Sos. dan Ekon. Kehutan.*, vol. 12, no. 3, pp. 197–211, 2015.
- [11] Y. A. Singgalen, “Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM,” *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 614–625, 2023, doi: 10.47065/josh.v4i2.2999.
- [12] 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, pp. 1343–1352, 2022, doi: 10.47065/bits.v4i3.2486.
- [13] 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.
- [14] 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.
- [15] Y. A. Singgalen, “Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia,” *J. Inf. Syst. Informatics*, vol. 4, no. 1, pp. 459–470, 2021, [Online]. Available: <http://journal-isi.org/index.php/isi/article/view/171>
- [16] A. A. Arifiyanti, M. F. Pandji, and B. Utomo, “Analisis Sentimen Ulasan Pengunjung Objek Wisata Gunung Bromo pada Situs Tripadvisor,” *Explor. J. Sist. Inf. dan Telemat.*, vol. 13, no. 1, p. 32, 2022, doi: 10.36448/jsit.v13i1.2539.
- [17] 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.
- [18] A. Fadli, M. I. Zulfa, and Y. Ramadhani, “Performance Comparison of Data Mining Classification Algorithms for Early Warning System of Students Graduation Timeliness,” *J. Teknol. dan Sist. Komput.*, vol. 6, no. 4, pp. 158–163, 2018, doi: 10.14710/jtsiskom.6.4.2018.158-163.
- [19] 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.
- [20] R. A. Barro, I. D. Sulvianti, and M. Afendi, “Penerapan Synthetic Minority Oversampling Technique (Smote) Terhadap Data Tidak Seimbang Pada Pembuatan Model Komposisi Jamu,” *Xplore J. Stat.*, vol. 1, no. 1, pp. 1–6, 2013.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] Hairani, N. A. Setiawan, and T. B. Adji, “Metode Klasifikasi Data Mining dan Teknik Sampling SMOTE Menangani Class Imbalance untuk Segmentasi Customer pada Industri Perbankan,” in *Seminar Nasional Sains dan Teknologi*, 2016, pp. 168–172.
- [28] A. A. Amsal, B. B. Albar, and Y. H. Yeni, “Pendekatan Text Analysis Terhadap Ulasan Pengunjung Hotel: Apa yang Mempengaruhi Kepuasan Pelanggan?,” *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 57–66, 2021, doi: 10.35889/jutisi.v10i1.580.
- [29] M. H. Almuyassar, S. P. Fidi, and R. Akbar, “Peranan Prinsip Kepuasan Pelanggan Guna Meningkatkan Jumlah Penjualan pada Ecommerce,” *Integr. (Information Technol. Vocat. Educ.)*, vol. 2, no. 1, pp. 25–28, 2020.
- [30] M. Yumarlin, J. E. Bororing, S. Rahayu, and J. F. Andika, “Analisis Sentimen Terhadap Layanan Tokopedia Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM),” *Smart Comp*, vol. 12, no. 1, pp. 154–163, 2023, doi: 10.30865/mib.v4i3.2181.
- [31] M. D. Pamungkas and H. Februariyanti, “Penerapan Algoritma k-Means Clustering untuk Mengelompokkan Data Review Barang pada e-Commerce Lazada,” *semanTIK*, vol. 8, no. 2, pp. 99–104, 2022, doi: 10.55679/semanik.v8i2.29058.
- [32] B. Z. Ramadhan, I. Riza, and I. Maulana, “Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes,” *J. Appl. Informatics Comput.*, vol. 6, no. 2, pp. 220–225, 2022.
- [33] S. R. Amara, N. Mutiah, and F. Febriyanto, “Pengaruh Kualitas Layanan e-Commerce terhadap Kepuasan Pengguna Menggunakan Metode Mobile Service Quality dan Importance Performance Analysis (IPA) (Studi Kasus : Pengguna e-Commerce Shopee di Kota Pontianak),” *Coding J. Komput. dan Apl.*, vol. 10, no. 2, pp. 274–287, 2022.
- [34] J. T. Frederick and H. A. Kurniawati, “Desain Floating Club House untuk Kawasan Wisata Taman Nasional Bunaken,

- Sulawesi Utara,” *J. Tek. ITS*, vol. 10, no. 2, pp. 122–129, 2021, doi: 10.12962/j23373539.v10i2.80682.
- [35] T. Friliyantini, A. V. S. Hubies, and A. Munandar, “Strategi Pengembangan Usaha Mikro dan Kecil Sektor Wisata Bahari di Pulau Kecil (Studi Kasus Pulau Bunaken, Kota Manado, Sulawesi Utara),” *Manaj. IKM*, vol. 6, no. 1, pp. 55–63, 2011, [Online]. Available: <https://journal.ipb.ac.id>
- [36] A. M. Pravina, I. Cholissodin, and P. P. Adikara, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4793>