

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/369691968>

Analisis Sentimen Wisatawan terhadap Kualitas Layanan Hotel dan Resort di Lombok Menggunakan SERVQUAL dan CRISP-DM

Article in Building of Informatics Technology and Science (BITS) · March 2023

DOI: 10.47065/bits.v4i4.3199

CITATIONS

20

READS

472

1 author:



Yerik Afrianto Singgalen

Atma Jaya Catholic University of Indonesia

237 PUBLICATIONS 1,151 CITATIONS

SEE PROFILE

Analisis Sentimen Wisatawan terhadap Kualitas Layanan Hotel dan Resort di Lombok Menggunakan SERVQUAL dan CRISP-DM

Yerik Afrianto Singgalen*

Faculty of Business Administration and Communication, Tourism Department, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia

Email: *yerik.afrianto@atmajaya.ac.id

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

Submitted: 27/02/2023; Accepted: 27/03/2023; Published: 31/03/2023

Abstrak—Era transformasi digital telah memantik inovasi dalam strategi pemasaran produk dan layanan di berbagai sektor, salah satunya ialah sektor pariwisata. Dalam konteks industri perhotelan, pemasaran produk menggunakan media digital berbasis website memungkinkan konsumen sebagai tamu hotel untuk memberikan ulasan terhadap produk dan layanan yang diterima. Website Tripadvisor menjadi salah satu platform pemasaran digital yang menyediakan fitur ulasan bagi pengguna aplikasi khususnya konsumen, untuk memberikan penilaian (rating) dan ulasan (review). Penelitian ini bertujuan menganalisis kualitas layanan hotel menggunakan kerangka kerja Service Quality (SERVQUAL) berdasarkan hasil klasifikasi data sentimen tamu hotel menggunakan algoritma Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM) sesuai dengan tahapan Cross-Industry Standard Process for Data Mining (CRISP-DM). Kerangka kerja CRISP-DM terdiri dari enam tahapan yakni : tahap business understanding, tahap data understanding, tahap data preparation, tahap modeling, tahap evaluation dan tahap deployment. Sedangkan, kerangka kerja SERVQUAL terdiri dari beberapa dimensi sebagai berikut : dimensi reliability; responsiveness; assurance; empathy; tangibles. Data ulasan yang akan diproses ialah data ulasan konsumen The Oberoi Beach Resort Lombok; Seraton Senggigi Beach Resort; Sudamala Resort Senggigi; Holiday Resort Lombok; Aston Sunset Beach Resort. Hasil penelitian ini menunjukkan bahwa algoritma SVM memiliki performa yang lebih baik dibandingkan NBC, dimana nilai akurasi berjumlah 98,57%, nilai presisi berjumlah 100%, nilai recall berjumlah 97,14%, dan nilai f-measure berjumlah 98,54%. Adapun nilai AUC ialah 100% dan nilai t-Test berjumlah 98,6%. Berbeda halnya dengan hasil evaluasi kinerja algoritma SVM tanpa menggunakan Operator SMOTE Upsampling, dimana nilai akurasi berjumlah 95,71%, nilai presisi berjumlah 95,71%, nilai recall berjumlah 100%, dan nilai f-measure berjumlah 97,81%. Selain itu, nilai AUC ialah 91,1% dan nilai t-Test berjumlah 95,7%.

Kata Kunci: Hotel; Resort; SERVQUAL; CRISP-DM; NBC; SVM; Lombok

Abstract— The era of digital transformation has sparked innovations in product and service marketing strategies in various sectors, one of which is the tourism sector. In the hospitality industry context, product marketing using website-based digital media allows consumers as hotel guests to review the products and services received. The Tripadvisor website is a digital marketing platform that provides review features for app users, especially consumers, to give ratings and reviews. This study aims to analyze the quality of hotel services using the Service Quality (SERVQUAL) framework based on the results of the classification of hotel guest sentiment data using the Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) algorithm by the stages of the Cross-Industry Standard Process for Data Mining (CRISP-DM). The CRISP-DM framework consists of six stages, namely: business understanding stage, data understanding stage, data preparation stage, modeling stage, evaluation stage, and deployment stage. The SERVQUAL framework consists of several dimensions: reliability dimension; responsiveness; assurance; empathy; tangibles. The review data that will be processed is the consumer review data of The Oberoi Beach Resort Lombok; Sheraton Senggigi Beach Resort; Sudamala Resort Senggigi; Holiday Resort Lombok; Aston Sunset Beach Resort. The results of this study show that the SVM algorithm performs better than NBC, where the accuracy value is 98.57%, the precision value is 100%, the recall value is 97.14%, and the f-measure value is 98.54%. The AUC value is 100%, and the t-Test value is 98.6%. Unlike the case with the results of SVM's algorithm performance evaluation without using the SMOTE Upgrading Operator, where the accuracy value is 95.71%, the precision value is 95.71%, the recall value is 100%, and the f-measure value is 97.81%. In addition, the AUC value is 91.1%, and the t-Test value is 95.7%.

Keywords: Hotel; Resort; SERVQUAL; CRISP-DM; NBC; SVM; Lombok

1. PENDAHULUAN

Transformasi digital di Indonesia telah memantik inovasi bisnis di bidang pemasaran produk dan jasa. Firmansyah et al. menunjukkan bahwa transformasi digital didukung oleh kemampuan literasi digital dapat meningkatkan daya saing, oleh sebab itu kesiapan Sumber Daya Manusia (SDM) berperan penting dalam inovasi bisnis di era digital [1]. Disisi lain, Purnama et al. menunjukkan bahwa keterampilan baru, kompetensi inti, perilaku adaptif dan inovatif untuk menemukan dan mempraktikkan model bisnis baru menjadi faktor penentu kesuksesan di era industri 4.0 [2]. Adapun, Asrol et al. berpendapat bahwa percepatan transformasi digital sangat penting dalam keberlanjutan bisnis dan pemulihan Usaha Mikro Kecil Menengah (UMKM) [3]. Hal ini menunjukkan bahwa transformasi digital di Indonesia menjadi salah satu kunci keberhasilan ekonomi, meskipun tantangan transformasi digital ialah kesiapan SDM menanggapi perubahan model bisnis akibat perkembangan teknologi informasi. Dalam konteks penelitian ini, perkembangan teknologi memengaruhi model bisnis pariwisata khususnya di bidang pemasaran jasa akomodasi yang beralih dari proses konvensional ke digital berbasis website maupun mobile. Perubahan tersebut membentuk perilaku konsumen untuk mengakses informasi digital sebelum mengambil keputusan untuk menginap di hotel tertentu. Selanjutnya, konsumen yang telah menggunakan jasa akomodasi dapat mengulas produk dan layanan di website, sehingga bisa menjadi rujukan konsumen lainnya sebelum menggunakan produk dan layanan di hotel yang sama.

Mempertimbangkan hal tersebut, maka penelitian ini menawarkan gagasan untuk mengelaborasi kerangka kerja SERVQUAL dan CRISP-DM dengan memanfaatkan data ulasan konsumen terhadap produk dan layanan di hotel.

Kualitas layanan di industri perhotelan berperan penting dalam meningkatkan kepuasan tamu hotel, serta memengaruhi keputusan untuk menggunakan kembali layanan akomodasi hotel di masa mendatang. Ainun et al. menunjukkan bahwa indikator kualitas dapat klasifikasi menjadi lima dimensi sebagai berikut : dimensi *tangible*, yang berhubungan dengan fasilitas fisik atau peralatan; dimensi *reliability*, berhubungan dengan kemampuan untuk memberikan layanan secara akurat; dimensi *responsiveness*, merupakan kesediaan untuk membantu konsumen serta memberikan perhatian yang tepat; dimensi *guarantee/assurance*, merupakan keterampilan dan pengetahuan karyawan dalam meyakinkan konsumen; dimensi *empathy*, mencakup perhatian karyawan terhadap hal-hal yang dibutuhkan oleh konsumen [4]. Disisi lain, Prananda et al. menunjukkan bahwa penerapan metode SERVQUAL di industri perhotelan dapat meningkatkan kepuasan dan loyalitas tamu hotel untuk berkunjung kembali [5]. Selain itu, Sibua et al. menunjukkan bahwa SERVQUAL dapat dibandingkan dengan kerangka kerja SWOT untuk memberikan rekomendasi program peningkatan kualitas layanan di industri perhotelan [6]. Hal ini menunjukkan bahwa indikator penilaian dalam pendekatan SERVQUAL dapat dibandingkan dengan kerangka kerja lain seperti CRISP-DM untuk mengidentifikasi dan menganalisis model yang relevan dengan konteks dan kondisi eksisting hotel. Dengan demikian, penelitian ini akan menggunakan indikator SERVQUAL di tahap pemodelan metode CRISP-DM untuk merekomendasikan program peningkatan kapasitas Sumber Daya Manusia (SDM), fasilitas dan manajemen hotel yang tepat guna, sehingga memiliki daya saing tinggi dan berkelanjutan.

Perkembangan bisnis di era digital menunjukkan bahwa teknologi menjadi faktor penting dalam mempertahankan dan meningkatkan keuntungan perusahaan. Intensitas penggunaan platform digital sebagai media komunikasi telah meningkatkan jumlah data digital (teks, audio, gambar, video) yang dapat dimanfaatkan sebagai aset digital di era *big data*. Singgalen berpendapat bahwa salah satu kerangka kerja yang dapat digunakan dalam pendekatan *data mining* untuk optimalisasi manajemen organisasi atau perusahaan ialah *Cross Industry Standard Process for Data Mining* (CRISP-DM) [7]–[9]. Fadli et al. menunjukkan bahwa pendekatan CRISP-DM lebih relevan digunakan dalam optimalisasi manajemen organisasi atau perusahaan, karena menghubungkan model algoritma yang digunakan berdasarkan karakteristik data dan proses bisnis dari suatu organisasi atau perusahaan [10]. Selain itu, Burhanuddin et al. berpendapat bahwa pendekatan CRISP-DM dapat digunakan untuk mengidentifikasi, mengolah dan menganalisis data ulasan yang mencerminkan sentimen konsumen terhadap produk dan jasa menggunakan model yang relevan dengan karakteristik bisnis masing-masing [11]. Hal ini menunjukkan bahwa CRISP-DM dapat digunakan sebagai kerangka kerja dimana sumber data yang digunakan serta karakteristik data ulasan yang diolah akan disesuaikan dengan karakteristik bisnis industri perhotelan. Dengan demikian penelitian ini mengelaborasi indikator SERVQUAL dalam kerangka kerja CRISP-DM untuk menghasilkan rekomendasi terkait dengan kualitas layanan di industri perhotelan sebagai pendukung sektor pariwisata Indonesia.

Beberapa studi tentang analisis sentimen wisatawan terhadap produk dan layanan di sektor pariwisata menekankan pada evaluasi performa algoritma dalam klasifikasi sentimen menjadi kelas negatif dan kelas positif berdasarkan data teks yang diperoleh dari website Tripadvisor [12] maupun media sosial Twitter [13]. Christanto dan Singgalen menunjukkan bahwa pengguna website Tripadvisor merupakan representasi dari wisatawan yang telah memiliki pengalaman perjalanan wisata kemudian mengulas pengalaman pribadi dan kolektif (bersama teman, pasangan, dan keluarga) ketika mengonsumsi produk dan layanan di suatu destinasi, hotel, maupun restoran [14]. Disisi lain, Abadi dan Herwin berpendapat bahwa wisatawan yang memperoleh layanan terbaik cenderung mengulas hal-hal yang positif dalam pengalaman menginap, demikian juga sebaliknya, ketidakpuasan wisatawan terhadap layanan di hotel berdampak pada tindakan untuk memberikan ulasan negatif [15]. Adapun, Dharmawan et al. mengadopsi pendekatan *data mining* melalui algoritma *Support Vector Machine* (SVM) untuk memprediksi kepuasan pengunjung di suatu destinasi wisata [16]. Hal ini menunjukkan bahwa kajian analisis sentimen tidak terbatas pada uraian tentang jumlah data yang telah diklasifikasi berdasarkan sentimen positif dan negatif, melainkan uraian yang menekankan pada aspek manajerial berdasarkan karakteristik bisnis. Mempertimbangkan hal tersebut maka penelitian ini menggunakan indikator SERVQUAL dan kerangka kerja CRISP-DM berdasarkan konteks bisnis pariwisata khususnya industri perhotelan, sebagai penyedia jasa akomodasi bagi wisatawan. Dengan demikian, pembahasan dapat dilakukan secara komprehensif dengan menyeimbangkan analisis sesuai konteks bisnis dan tahapan dalam proses analisis sentimen.

Perkembangan teknologi digital telah memengaruhi proses bisnis di sektor pariwisata, dimana konsep *digital tourism* menjadi populer di kalangan wisatawan milenial. Aryawati berpendapat bahwa *digital tourism* menjadi populer di kalangan wisatawan milenial karena perilaku yang menunjukkan ketertarikan terhadap pengalaman baru, unik, otentik dan personal [17]. Kurniansah dan Purnama menunjukkan bahwa Pemerintah Indonesia tengah mengembangkan infrastruktur pariwisata di Nusa Tenggara Barat (NTB) yakni Lombok sebagai salah satu destinasi super prioritas nasional [18]. Selanjutnya, Tumbade et al. menunjukkan bahwa pemerintah tengah mengembangkan infrastruktur teknologi di Lombok berdasarkan konsep *smart city* [19]. Disisi lain, Haster dan Hartomo menunjukkan bahwa berdasarkan tingkat kematangan *smart city* Kabupaten Lombok Utara, terdapat tiga aspek yang perlu dikembangkan yakni *smart governance*, *smart living*, dan *smart society* [20]. Hal ini menunjukkan bahwa pengembangan wilayah Lombok berdasarkan konsep *smart city* memungkinkan inovasi bisnis pariwisata dalam optimalisasi kualitas produk dan layanan berbasis teknologi informasi. Ketersediaan infrastruktur penunjang serta kemudahan akses terhadap jaringan komunikasi dan internet, memungkinkan industri perhotelan untuk memanfaatkan

media pemasaran digital berbasis website yakni Tripadvisor dalam mempromosikan produk dan layanan secara virtual. Mempertimbangkan hal tersebut maka penelitian ini menawarkan gagasan untuk menganalisis sentimen wisatawan sebagai tamu hotel dan resort di wilayah Lombok yang secara spesifik pada *The Oberoi Beach Resort Lombok*, *Seraton Senggigi Beach Resort*, *Sudamala Resort Senggigi*, *Holiday Resort Lombok*, dan *Aston Sunset Beach Resort* yang diperoleh dari website Tripadvisor. Dengan demikian, dapat diperoleh gambaran secara komprehensif, sentimen wisatawan terhadap kualitas layanan (SERVQUAL) di hotel dan resort, maupun algoritma yang tepat digunakan sebagai model untuk pengolahan data teks sesuai kerangka kerja CRISP-DM.

Penelitian ini menggunakan *Support Vector Machine* (SVM) sebagai algoritma dalam tahap pemodelan untuk klasifikasi data sebagaimana CRISP-DM. Sukiakhy dan Jummi menunjukkan bahwa algoritma SVM menunjukkan performa yang baik dalam klasifikasi data ulasan wisatawan di media sosial (*twitter*) berdasarkan hasil evaluasi *confusion matrix* pada nilai akurasi, presisi, recall, dan f-measure [21]. Adapun, Muthia menunjukkan bahwa algoritma SVM dapat dibandingkan dengan algoritma *Naïve Bayes Classifier* (NBC) dalam mengukur performa klasifikasi data [22]. Beberapa penelitian terdahulu menunjukkan sejumlah algoritma populer yang dapat digunakan sebagai model dalam mengklasifikasi data teks, seperti halnya algoritma *Decision Tree* (DT) [23], algoritma *k-Nearest Neighbor* (k-NN) [24], dan algoritma *k-Means* [25]. Hal ini menunjukkan bahwa analisis sentimen merupakan suatu pendekatan yang perlu disesuaikan dengan latarbelakang bisnis, tujuan pengolahan data, karakteristik data, dan algoritma yang digunakan. Dengan demikian, penelitian ini fokus menggunakan algoritma SVM sebagai model klasifikasi data berdasarkan kerangka kerja CRISP-DM.

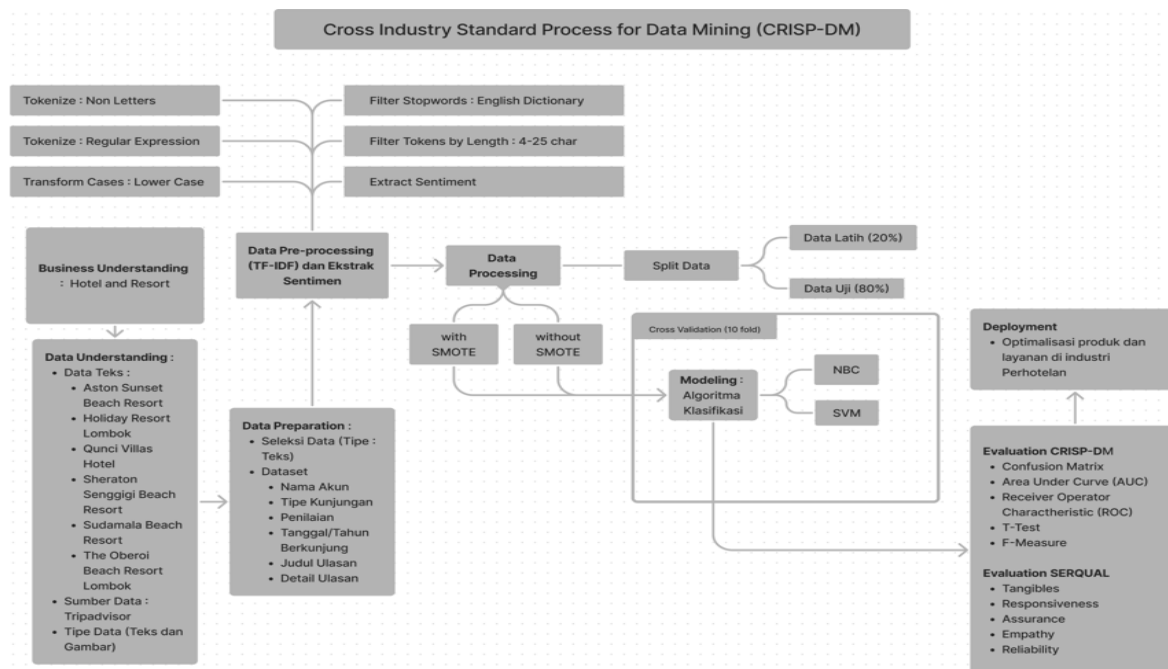
Penelitian tentang analisis sentimen wisatawan terhadap kualitas pelayanan hotel di Lombok sangat penting untuk dilakukan. Dengan menganalisis sentimen konsumen, para pengelola hotel dapat mengetahui apa yang dianggap penting oleh pelanggan mereka, dan dengan demikian dapat meningkatkan kepuasan pelanggan dan meningkatkan loyalitas pelanggan. Selain itu, dengan mengetahui kelemahan dan kekuatan layanan hotel, para pengelola hotel dapat memperbaiki kekurangan dan meningkatkan kualitas layanan yang mereka tawarkan. Hasil penelitian ini dapat digunakan sebagai dasar untuk pengambilan keputusan dan perencanaan strategis, sehingga dapat membantu pengelola hotel untuk meningkatkan daya saing mereka di pasar. Selain itu, penelitian ini dapat membantu meningkatkan reputasi pariwisata di Lombok dan menghasilkan pendapatan yang lebih besar bagi para pemangku kepentingan di sektor pariwisata.

2. METODOLOGI PENELITIAN

2.1 Cross-Industry Standard Process for Data Mining (CRISP-DM) and Service Quality (SERVQUAL)

Metode yang digunakan dalam analisis sentimen wisatawan terhadap kualitas layanan hotel dan resort di Lombok ialah *Cross Industry Standard Process for Data Mining* (CRISP-DM) yang dihubungkan dengan metode *Service Quality* (SERVQUAL). Metode CRISP-DM meliputi beberapa tahapan sebagai berikut : tahap memahami proses bisnis (*business understanding*), tahap memahami data (*data understanding*), tahap menyiapkan data (*data preparation*), tahap pemodelan (*modeling*), tahap evaluasi (*evaluation*) dan tahap implementasi (*deployment*). Sementara itu, metode SERVQUAL menekankan pada dimensi bukti fisik (*tangible*), dimensi daya tanggap (*responsiveness*), dimensi kehandalan (*reliability*), dimensi jaminan (*assurance*), dimensi empati (*empathy*). Latarbelakang penggunaan metode CRISP-DM ialah untuk mengidentifikasi sentimen, mengklasifikasi data sentimen, dan menganalisis data sentimen wisatawan terhadap kualitas produk dan layanan di hotel. Sementara itu, metode SERVQUAL digunakan untuk mengklasifikasi data sentimen serta menganalisis kepuasan pelanggan.

Gambaran tentang implementasi metode CRISP-DM dan SERVQUAL ialah sebagai berikut : tahap *business understanding*, dilakukan studi literatur tentang manajemen pelayanan hotel, psikologi pelayanan, perilaku wisatawan, dan analisis sentimen tamu hotel untuk memahami segmentasi pasar di industri perhotelan sebagaimana konteks Indonesia; tahap *data understanding*, dilakukan seleksi dan validasi data teks dalam bentuk ulasan member website Tripadvisor berdasarkan kategori nama akun, penilaian, judul, ulasan, tipe kunjungan (*solo*, *couple*, *with friends*, *with family*, *for business*), serta tanggal menginap; tahap *data preparation*, dilakukan proses seleksi data yang akan dibersihkan (*tokenize*, *transform cases*, *filter stopwords*, *filter tokens by length*); tahap *data processing*, dilakukan pembagian data latih (20%) dan data uji (80%), kemudian dilakukan pengujian algoritma menggunakan operator SMOTE Upsampling serta dilakukan perbandingan dengan hasil pengujian algoritma tanpa menggunakan SMOTE Upsampling untuk menganalisis nilai *Area Under Curve* (AUC); tahap *modeling*, dilakukan klasifikasi dataset menggunakan operator *cross validation* pada algoritma *Naïve Bayes Classifier* (NBC), dan *Support Vector Machine* (SVM) untuk memperoleh gambaran tentang performa algoritma terbaik; tahap *evaluation*, dilakukan analisis nilai *confusion matrix*, *Area Under Curve* (AUC), dan *Receiver Operating Characteristic* (ROC), f-measures, and T-Test. Adapun, hasil klasifikasi data berdasarkan ekstrak sentimen yang divisualisasi menggunakan wordcloud/bar-column akan dihubungkan dengan setiap dimensi dalam metode SERVQUAL yakni dimensi bukti fisik (*tangible*), dimensi daya tanggap (*responsiveness*), dimensi kehandalan (*reliability*), dimensi jaminan (*assurance*), dimensi empati (*empathy*). ; tahap *deployment*, diberikan rekomendasi model dan luaran hasil analisis sentimen berdasarkan CRISP-DM dan SERVQUAL untuk mengoptimalkan produk dan layanan hotel dan resort di Lombok. Adapun, alur dari keseluruhan tahapan dapat dilihat pada gambar 1 berikut ini.



Gambar 1. Tahapan Penelitian sesuai CRISP-DM dan SERVQUAL

Gambar 1 merupakan perpaduan antara metode CRISP-DM dan metode SERVQUAL dalam menganalisis analisis sentimen wisatawan terhadap kualitas layanan hotel dan resort menggunakan algoritma *Naïve Bayes Classifier* (NBC), dan algoritma *Support Vector Machine* (SVM). Berdasarkan hasil pengumpulan data, terdapat enam hotel dengan kualifikasi hotel berbintang dengan wilayah operasional di Lombok, Indonesia. Adapun, hotel dan resort yang dimaksud dapat dilihat pada tabel berikut.

Tabel 1. Data Hotel dan Resort, Jumlah Data dan Sumber Data

Dataset	Jumlah Data (Sebelum Filter)	Jumlah Data (Setelah Filter)	Link Tripadvisor
Aston Sunset Beach Resort	1415	1214	https://www.tripadvisor.com/Hotel_Review-g664666-d6557938-Reviews-Aston_Sunset_Beach_Resort-Gili_Trawangan_Gili_Islands_Pemenang_Lombok_West_Nusa_Tenggara.html
Holiday Resort Lombok	1076	961	https://www.tripadvisor.com/Hotel_Review-g2708413-d299126-Reviews-Holiday_Resort_Lombok-Mangsit_Senggigi_Lombok_West_Nusa_Tenggara.html
Qunci Villas Hotel	2277	2202	https://www.tripadvisor.com/Hotel_Review-g2708413-d547418-Reviews-Qunci_Villas_Hotel-Mangsit_Senggigi_Lombok_West_Nusa_Tenggara.html
Sheraton Senggigi Beach Resort	1687	1497	https://www.tripadvisor.com/Hotel_Review-g1592993-d304529-Reviews-Sheraton_Senggigi_Beach_Resort-Senggigi_Lombok_West_Nusa_Tenggara.html
Sudamala Beach Resort	1017	987	https://www.tripadvisor.com/Hotel_Review-g2708413-d4116487-Reviews-Sudamala_Resort_Senggigi-Mangsit_Senggigi_Lombok_West_Nusa_Tenggara.html
The Oberoi Beach Resort	682	665	https://www.tripadvisor.com/Hotel_Review-g3177245-d302429-Reviews-The_Oberoi_Beach_Resort_Lombok-Tanjung_Lombok_West_Nusa_Tenggara.html

Tabel 1 merupakan data hotel dan resort yang digunakan dalam proses analisis sentimen menggunakan metode CRISP-DM dan SERVQUAL. Dataset yang digunakan ialah sebagai berikut : 1214 data ulasan dari Aston Sunset Beach Resort; 961 data ulasan Holiday Resort Lombok; 2202 data ulasan dari Qunci Villas Hotel; 1497 data ulasan dari Sheraton Senggigi Beach Resort; 987 data ulasan dari Sudamala Beach Resort; 665 data ulasan dari The Oberoi Beach Resort. Jumlah keseluruhan dataset yang akan diproses ialah 7526 data ulasan. Dengan demikian, tahapan *business understanding* dan *data understanding* fokus pada industri perhotelan berdasarkan 7526 data ulasan terhadap

hotel dan resort yang wilayah operasionalnya di Lombok yakni Aston Sunset Beach Resort, Holiday Resort Lombok, Qunci Villas Hotel, Sheraton Senggigi Beach Resort, Sudamala Beach Resort, dan The Oberoi Beach Resort.

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 :

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

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

Kosasih dan Alberto menunjukkan bahwa proses pembobotan kata dapat dilakukan dengan beberapa tahapan sebagai berikut : pertama, menghitung jumlah Term Frequency (FT) 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. 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 [19]. 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 [20]. 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 Naïve Bayes Classifier

Naive Bayes Classifier (NBC) memiliki keunggulan tersendiri dalam mengklasifikasikan data dengan probabilitas sederhana dengan menerapkan teorema Bayes dengan asumsi independensi yang tinggi [26]. Studi ini didasarkan pada jumlah dataset yang digunakan dalam metode dengan kemampuan klasifikasi yang cepat dan akurat. Pengklasifikasi Naive Bayesian hanya membutuhkan data pelatihan dalam jumlah yang relatif kecil untuk menentukan estimasi parameter yang diperlukan untuk proses klasifikasi. Pada tahap klasifikasi, nilai kelas ditentukan dari data yang didasarkan pada suku yang terjadi, dengan menggunakan persamaan berikut.

$$P(X_k|Y) = \frac{P(Y|X_k)}{\sum_i P(Y|X_i)} \quad (5)$$

Dimana, keadaan posterior (Probabilitas X_k di dalam Y) dapat dihitung dari keadaan prior (Probabilitas Y di dalam X_k) dibagi dengan jumlah dari semua probabilitas Y di dalam semua X_i . Dalam konteks penelitian ini, data teks yang diperoleh dari website Tripadvisor diklasifikasi menggunakan persamaan berikut

$$P(v1|C=c) = \frac{\text{CountTerms}(v1, \text{docsv}(c))}{\text{AllTerms}(\text{docs}(c))} \quad (6)$$

Dimana $v1$ merupakan salah satu suku kata yang muncul dalam ulasan pengguna website Tripadvisor terhadap kualitas produk dan layanan di Destinasi Wisata Candi Borobudur. Sedangkan, $\text{CountTerms}(v1, \text{docsv}(c))$ merujuk pada jumlah kemunculan suatu kata berlabel c (“positif” atau “negatif”). Adapun, $\text{AllTerms}(\text{docs}(c))$ merujuk pada jumlah semua kata berlabel c yang ada pada dataset. Untuk menghindari adanya nilai nol pada probabilitas maka diimplementasikan laplace smoothing, untuk mengurangi probabilitas dari hasil yang terobservasi, dan juga meningkatkan probabilitas hasil yang belum terobservasi. Dengan demikian, persamaan yang digunakan ialah sebagai berikut :

$$P(v1|C=c) = \frac{\text{CountTerms}(v1, \text{docsv}(c)) + 1}{\text{AllTerms}(\text{docs}(c)) + |V|} \quad (7)$$

Dimana $|V|$ merujuk pada jumlah semua kata dalam data ulasan yang ada di dataset. Dengan demikian, proses klasifikasi data ulasan akan menunjukkan kata dengan nilai tertinggi sebagai representasi perhatian pengulas terhadap produk dan jasa layanan pariwisata di Destinasi Candi Borobudur melalui website Tripadvisor.

2.5 Algoritma Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu algoritma yang dapat digunakan untuk klasifikasi data menggunakan hyperplane [27]. Karim menjelaskan bahwa konsep SVM menitikberatkan pada *risk minimization*, yaitu estimasi fungsi dengan cara meminimalisir batas dari *generalization error*, sehingga SVM mampu mengatasi *overfitting* [28]. Adapun, fungsi regresi dari metode SVM adalah sebagai berikut.

$$f(x) = w^T \phi(x) + b \quad (8)$$

Dimana w merupakan vector pembobot, $\phi(x)$ 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 [29]. Hal ini menunjukkan bahwa SVM mampu menghasilkan performa yang baik meskipun dengan jumlah data yang relatif sedikit. Dengan demikian, perlu dilakukan pengujian algoritma SVM dengan dataset ulasan pengunjung Candi Borobudur.

2.6 Evaluasi Klasifikasi (Confusion Matrix)

Pada tahap evaluasi, algoritma NBC dan SVM dievaluasi berdasarkan nilai akurasi, presisi, *recall*, dan nilai f-measure. Kananah 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.

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \quad (9)$$

$$\text{Presisi/Specificity} = \frac{TP}{TP+FP} \quad (10)$$

$$\text{Recall/Sensitivity} = \frac{TP}{TP+FN} \quad (11)$$

$$f - \text{measure} = \frac{2x(\text{Presisi} \times \text{recall})}{\text{presisi} + \text{recall}} \quad (12)$$

Ginatra et al., berpendapat bahwa *confusion matrix* merupakan gambaran akan akurasi, presisi, *recall* dari proses klasifikasi data. 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*/sensitivitas dan presisi sehingga menghasilkan metrik yang efektif untuk mencari kembali informasi dalam himpunan yang tidak seimbang [30]. 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

Industri perhotelan berperan penting dalam meningkatkan kunjungan wisatawan mancanegara ke Indonesia. Faldio et al. berpendapat bahwa hotel berperan penting dalam mendukung sektor pariwisata dengan menyediakan jasa akomodasi bagi wisatawan yang berkunjung di suatu destinasi [31]. Disisi lain, Wijayanto et al. menekankan bahwa opini tamu hotel perlu dianalisis berdasarkan klasifikasi sentimen untuk mengetahui tingkat kepuasan tamu ketika menginap di hotel [32]. Adapun, Fauzzia et al. berpendapat bahwa ketersediaan fasilitas dan kinerja karyawan hotel menentukan persepsi tamu hotel terkait dengan kepuasan dan loyalitas [33]. Hal ini menunjukkan bahwa analisis sentimen wisatawan yang menggunakan jasa akomodasi di suatu hotel, penting untuk dianalisis secara komprehensif dengan mempertimbangkan kualitas layanan. Dengan demikian, metode CRISP-DM dan SERVQUAL menjadi relevan digunakan untuk memperoleh luaran berupa rekomendasi bagi hotel untuk meningkatkan kualitas produk dan layanan.

Implementasi metode CRISP-DM dan SERVQUAL dalam pembahasan penelitian ini mengacu pada 7526 data ulasan tamu hotel dan resort yang wilayah operasionalnya di Lombok yakni Aston Sunset Beach Resort, Holiday Resort Lombok, Qunci Villas Hotel, Sheraton Senggigi Beach Resort, Sudamala Beach Resort, dan The Oberoi Beach Resort. Adapun, penelitian ini mengakumulasi seluruh data ulasan untuk diproses menggunakan algoritma NBC dan SVM, sehingga dapat diketahui hubungan antara ulasan tamu hotel dengan masing-masing dimensi dalam metode kualitas layanan yakni dimensi bukti fisik, dimensi kehandalan, dimensi daya tanggap, dimensi jaminan dan dimensi empati. Hal ini menunjukkan bahwa hasil klasifikasi jumlah kata dalam data ulasan tamu hotel berdasarkan dimensi SERVQUAL dapat digunakan sebagai dasar pengembangan kualitas produk dan layanan yang berkesan bagi tamu hotel, serta mampu meningkatkan kepuasan serta loyalitas tamu, sebagaimana tabel berikut.

Tabel 2. Jumlah Kata dalam Data Ulasan Tamu Hotel Berdasarkan Dimensi SERVQUAL

SERVQUAL	Kata	Jumlah Kata dalam Klasifikasi Negatif	Jumlah Kata dalam Klasifikasi Positif
Bukti Fisik (Tangible)	Pool	129	3326
	Room	348	3311
	Bathroom	57	503
	Property	22	366
	Facility	10	372
Daya Tanggap (Responsiveness)	Service	74	1746
	Pleasant	10	194
	Offer	11	145
	Wait	20	148
	Maintained	8	307
Kehandalan (Reliability)	Hospitality	1	142
	Welcoming	3	138
	Management	21	134
	Fast	4	116
	Provide	2	114
Jaminan (Assurance)	Private	13	647
	Setting	6	214
	Reasonable	3	151
	Absolutely	16	284
	Professional	1	165
Empati (Empathy)	Friendly	30	1574
	Kind	12	215
	Smile	3	162
	Attentive	6	256
	Feeling	5	77

Tabel 2 merupakan hasil klasifikasi kata yang paling sering muncul dalam data ulasan tamu berdasarkan metode SERVQUAL, sebagai berikut : Dimensi bukti fisik (*tangible*), terdapat kata *pool*, *room*, *bathroom*, *property*, *facility* yang menunjukkan impresif tamu hotel terhadap ketersediaan fasilitas maupun amenitas di hotel; Dimensi daya tanggap (*responsiveness*), terdapat kata *service*, *pleasant*, *offer*, *wait*, *maintained*, yang menunjukkan bahwa karyawan hotel telah memberikan pelayanan meskipun dalam beberapa kasus dinilai belum optimal oleh tamu hotel; Dimensi kehandalan (*reliability*), terdapat kata *hospitality*, *welcoming*, *management*, *fast*, *provide*, yang menunjukkan bahwa karyawan hotel telah mengelola sistem pelayanan hotel agar dapat menjawab kebutuhan tamu hotel secara efektif dan efisien; Dimensi jaminan (*assurance*), terdapat kata *private*, *setting*, *reasonable*, *absolutely*, *professional*, yang menunjukkan sikap profesional karyawan dalam memberikan layanan bagi tamu dengan pengaturan terbaik dan masuk akal; Dimensi empati (*empathy*), terdapat kata *friendly*, *kind*, *smile*, *attentive*, *feeling*, yang menunjukkan

bahwa karyawan hotel berupaya memberikan layanan dengan empati terhadap tamu hotel sehingga merasa aman dan nyaman selama menginap. Hal ini menunjukkan bahwa Aston Sunset Beach Resort, Holiday Resort Lombok, Qunci Villas Hotel, Sheraton Senggigi Beach Resort, Sudamala Beach Resort, dan The Oberoi Beach Resort telah memiliki manajemen pelayanan yang baik, namun perlu mempertahankan sistem layanan yang prima agar meningkatkan kepuasan dan loyalitas dari tamu hotel.

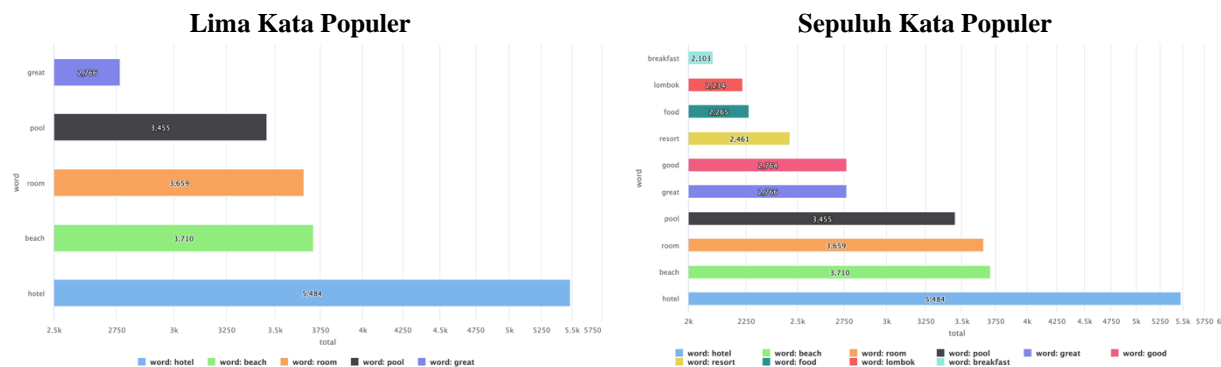
Klasifikasi data ulasan menjadi sentimen positif dan negatif dilakukan melalui proses ekstrak sentimen dari operator *Extract Sentiment* pada aplikasi Rapidminer. Operator *Extract Sentiment* memilih kata dan memberikan bobot kata dalam data teks (data ulasan) sehingga dapat diperoleh nilai positif dan negatif dari *score*. Semakin tinggi nilai atau *score* dalam klasifikasi negatif (*Negativity*) maka ulasan tersebut akan diidentifikasi sebagai ulasan negatif, demikian halnya apabila nilai atau *score* dalam klasifikasi positif (*Positivity*) maka ulasan tersebut akan diidentifikasi sebagai ulasan positif. Hasil dari ekstrak sentimen dapat dilanjutkan ke tahap *labeling* sehingga dapat dikalkulasi menggunakan algoritma NBC dan SVM sebagai model yang digunakan. Adapun, tabel 3 menunjukkan proses ekstrak sentimen untuk klasifikasi ulasan negatif dan positif.

Tabel 3. Proses Ekstrak Sentimen dari Ulasan Wisatawan

Review					
<p><i>Not too sure about anyone else but after our stay. Not impressed... It was a chore to get a towel by the pool. They were always drying. The staff was rude and unapologetic about any poor situation we found ourselves in and one staff member said "too bad" when the stairwell to a neighboring room smelled like sewage; this stairwell was right outside our room. They unplugged the fridge to plug in the DVD player without telling us so the desserts we brought back were warm and my partner was sick from it. After my partner was sick, I told the hotel manager about our issues with the stay. Novi (manager) then asked to see my partner who was sick to confirm her identity and condition which we felt was invasive and inappropriate. After confirming that my partner was indeed sick from the hotel's poor service, she was a lot more apologetic. We only stayed 3 nights, and every moment we were there something else would go wrong. Even the shower tiles were sharp and uncomfortable to stand on! In the morning I went to check out of the hotel, but the staff forgot my coffee order so I waited 30 minutes for a coffee. During the trip, the pizza we ordered for lunch took over an hour to come out... At 5:00 pm they had run out of beers at the pool bar which was frustrating... The breakfast was terrible. The pancakes were hard around the edges, the fruit wasn't covered, and swarming with insects... I'm not a winey occupant, very easy going but this really put a downer on my trip to Gili... Hiring a bike was a definite chore due to language barriers and we ended up having to just take a bike so we didn't miss the sunset... For what we paid we should have stayed at Le pirate... Do not stay here... a Horrible experience...</i></p>					
String Scoring					
<p>sure (0.33) impressed (0.54) rude (-0.51) poor (-0.54) bad (-0.64) like (0.38) warm (0.23) sick (-0.59) sick (-0.59) sick (-0.59) sick (-0.59) poor (-0.54) wrong (-0.54) uncomfortable (-0.41) frustrating (-0.49) terrible (-0.54) hard (-0.10) easy (0.49) definite (0.28) miss (-0.15) horrible (-0.64)</p>					
Sentiment	Score	Negativity	Positivity	Uncovered Tokens	Tokens
Negatif	-1,9	6,3	4,4	285	305
Review					
<p><i>This is one of the best beach front hotels we have stayed out and for the price and value outstanding...kind of dont wont to write a review so we can keep this place a secret :). We traveled as a couple and their was a few families staying however you would hardly know. With 3 amazing infinity pools and one as an adult only you have plenty to pick from. From the moment you wake up and head for breakfast all you can do is smile and relax. The breakfast are fresh and amazing, I loved the fruit, fresh juice and then being able to pick al carte. If you got up early enough or got lucky you could get a seat right at the edge of the water and if not there is plenty of room in the restaurant with great views of the ocean and surrounding area. The whole place is so well looked after and so clean and tidy, every time we got back to our room after breakfast it was cleaned it was like they know we had popped out. If all you are after is lounging by the pool and eating great food from one of the restaurants or by the pool this place is for you...if you are looking for adventure this place is also for you as easy to hire and get on boats to the islands etc. We booked a day trip to the gillis island via the hotel and it was easy to book and the trip was amazing and great value - we such a great day out on the boat and visiting the gillis islands and the people they used for the tour were great (sorry can't remember his name). All the 3 restaurants are lovely (only tip if you won't beach front at night you need to book in advance - sometimes is a little frustrating when it is all booked out as the restaurants are super popular and really good)however there are loads of options so one night we had pizza, the other lovely pasta, then another Spanish tapas and another Indonesian...so don't stress loads to pick from and really well priced and if you are fussy they will do their best to accommodate you. We were lucky to stay in an individual garden cottage room. lovely room and amazing outdoor shower. Best outdoor shower we have seen. The best thing about this place apart from the amazing staff we just the location...nothing else really around apart from a couple of local restaurants and just pure blue ocean that you could step right out onto and swim/ snorkel. We loved every morning going for a walk along the beach and feeling alive. So if you are looking for lovely accommodation, amazing fresh yummy food, great location, awesome infinity pools x 3 and sometime to relax read a book and also head out on a few day trips then this is the place for you. We are already planning on heading back next year :)</i></p>					
String Scoring					
<p>best (0.82) value (0.36) outstanding (0.77) kind (0.62) amazing (0.72) smile (0.38) relax (0.49) fresh (0.33) amazing (0.72) I (0.51) loved (0.74) fresh (0.33) lucky (0.46) great (0.79) well (0.28) clean (0.44) like (0.38) great (0.79) adventure (0.33) easy (0.49) easy (0.49) amazing (0.72) great (0.79) value (0.36) great (0.79) great (0.79) sorry (-0.08) lovely (0.72) frustrating (-0.49) super (0.74) popular (0.46) good (0.49) lovely (0.72) stress (-0.46) well (0.28) best (0.82) lucky (0.46) lovely (0.72) amazing (0.72) best (0.82) best (0.82) amazing (0.72) loved (0.74) feeling (0.13) alive (0.41) lovely (0.72) amazing (0.72) fresh (0.33) yummy (0.62) great (0.79) awesome (0.79) relax (0.49)</p>					

Sentiment Positif	Score 27,9	Negativity 1	Positivity 28,9	Uncovered Tokens 499	Tokens 551
----------------------	---------------	-----------------	--------------------	-------------------------	---------------

Tabel 3 merupakan proses ekstrak sentimen untuk klasifikasi data ulasan menjadi kelas negatif dan positif. Operator Extract Sentiment di aplikasi Rapidminer memberikan bobot pada setiap kata yang mengandung unsur positif dan negatif kemudian dikalkulasi untuk mendapatkan nilai *score* negatif maupun positif sehingga dapat diklasifikasi menjadi sentimen negatif dan positif. Setiap ulasan wisatawan terhadap kualitas layanan di hotel diklasifikasikan sebagai sentimen negatif apabila terdapat keluhan terkait dengan bukti fisik, kehandalan, daya tanggap, jaminan dan empati. Demikian halnya, ulasan diklasifikasi sebagai positif apabila terdapat apresiasi wisatawan terkait dengan dengan bukti fisik, kehandalan, daya tanggap, jaminan dan empati yang diperoleh saat menginap di hotel. Adapun, lima dan sepuluh kata teratas yang diperoleh dari hasil klasifikasi sentimen dapat dilihat pada gambar berikut.



Gambar 2. Lima dan Sepuluh Kata Populer dari Data Ulasan Wisatawan

Gambar 2 menunjukkan bahwa lima kata populer dalam data ulasan wisatawan terhadap kualitas layanan hotel dan resort di Lombok ialah *great*, *pool*, *room*, *beach*, dan *hotel* yang menunjukkan kepuasan terhadap fasilitas ruangan dan amenities hotel, serta pemandangan sekitar. Selanjutnya, sepuluh kata populer dalam data ulasan wisatawan terhadap kualitas layanan hotel dan resort di Lombok ialah *breakfast*, *lombok*, *food*, *resort*, *good*, *great*, *pool*, *room*, *beach*, dan *hotel* yang dapat diinterpretasikan bahwa produk berupa makanan yang disajikan kepada tamu hotel mendapatkan apresiasi yang positif. Hal ini menunjukkan implementasi pelayanan prima di Aston Sunset Beach Resort, Holiday Resort Lombok, Qunci Villas Hotel, Sheraton Senggigi Beach Resort, Sudamala Beach Resort, dan The Oberoi Beach Resort, perlu dipertahankan dan ditingkatkan. Selanjutnya, hasil implementasi metode klasifikasi menggunakan algoritma NBC dapat dilihat pada gambar berikut.

accuracy: 97.92% +/- 0.73% (micro average: 97.92%)

	true Negatif	true Positif	class precision
pred. Negatif	3142	131	96.00%
pred. Positif	0	3011	100.00%
class recall	100.00%	95.83%	

precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Positif)

	true Negatif	true Positif	class precision
pred. Negatif	3142	131	96.00%
pred. Positif	0	3011	100.00%
class recall	100.00%	95.83%	

recall: 95.83% +/- 1.45% (micro average: 95.83%) (positive class: Positif)

	true Negatif	true Positif	class precision
pred. Negatif	3142	131	96.00%
pred. Positif	0	3011	100.00%
class recall	100.00%	95.83%	

f_measure: 97.87% +/- 0.76% (micro average: 97.87%) (positive class: Positif)

	true Negatif	true Positif	class precision
pred. Negatif	3142	131	96.00%
pred. Positif	0	3011	100.00%
class recall	100.00%	95.83%	

Gambar 3. Confusion Matrix algoritma NBC dengan Operator SMOTE

Gambar 3 merupakan hasil evaluasi kinerja algoritma NBC dengan menggunakan Operator SMOTE Upsampling dimana nilai akurasi berjumlah 97.92%, nilai presisi berjumlah 100%, nilai recall berjumlah 95.83%, dan nilai *f-measure* berjumlah 97.87%. Selanjutnya nilai *true Negatif* dan *pred. Negatif* berjumlah 3142, nilai *true Positif* dan *pred. Negatif* berjumlah 131, nilai *true Negatif* dan *pred. Positif* berjumlah 0, dan nilai *true Positif* dan *pred. Positif* berjumlah 3011. Meskipun demikian, nilai AUC ialah 54,7% dan nilai t-Test berjumlah 97,9%. Berbeda halnya dengan hasil evaluasi kinerja algoritma NBC tanpa menggunakan Operator SMOTE Upsampling, dimana nilai akurasi berjumlah 91.72%, nilai presisi berjumlah 95,79%, nilai recall berjumlah 95,55%, dan nilai *f-measure* berjumlah 95,66%. Selanjutnya nilai *true Negatif* dan *pred. Negatif* berjumlah 11, nilai *true Positif* dan *pred. Negatif* berjumlah 140, nilai *true Negatif* dan *pred. Positif* berjumlah 132, dan nilai *true Positif* dan *pred. Positif* berjumlah 3002. Selain itu, nilai AUC ialah 50% dan nilai t-Test berjumlah 91,7%. Selain itu, hasil evaluasi performa algoritma SVM dengan menggunakan Operator SMOTE Upsampling dapat dilihat pada gambar berikut.

accuracy: 98.57% +/- 0.50% (micro average: 98.57%)

	true Negatif	true Positif	class precision
pred. Negatif	3142	90	97.22%
pred. Positif	0	3052	100.00%
class recall	100.00%	97.14%	

precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Positif)

	true Negatif	true Positif	class precision
pred. Negatif	3142	90	97.22%
pred. Positif	0	3052	100.00%
class recall	100.00%	97.14%	

recall: 97.14% +/- 0.99% (micro average: 97.14%) (positive class: Positif)

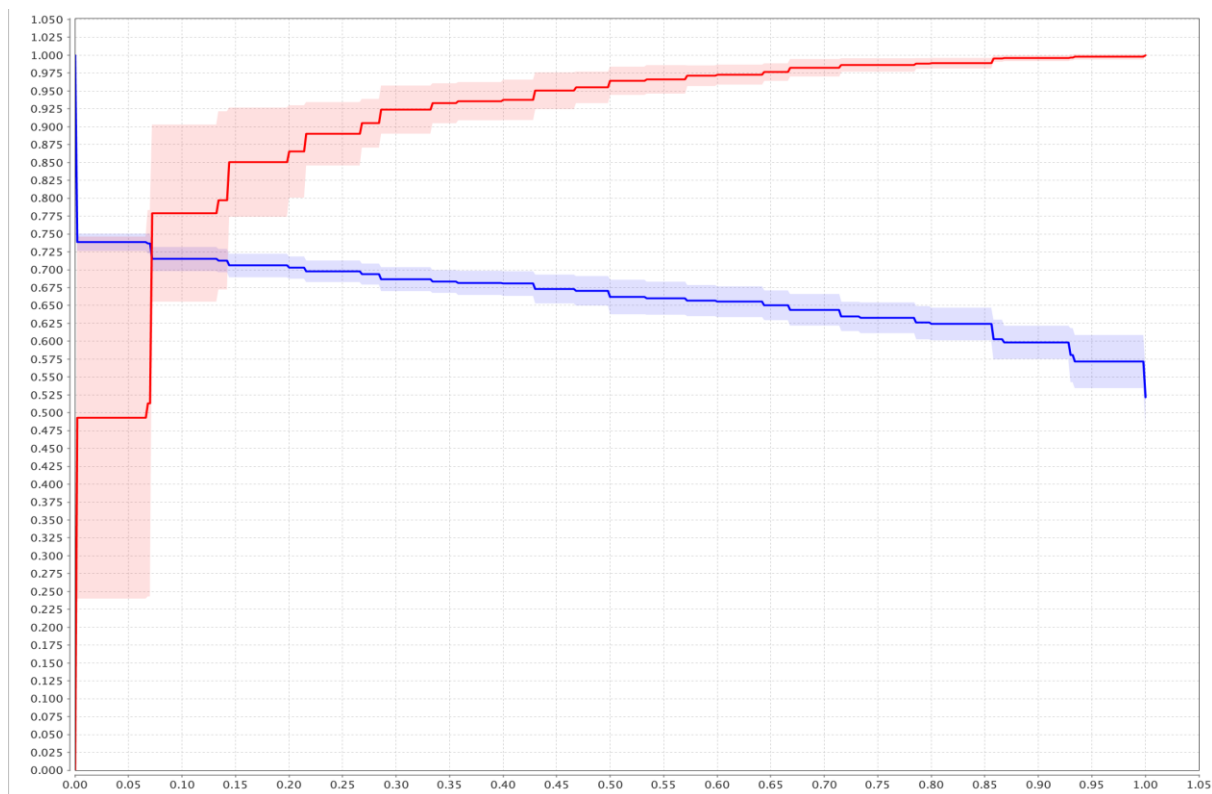
	true Negatif	true Positif	class precision
pred. Negatif	3142	90	97.22%
pred. Positif	0	3052	100.00%
class recall	100.00%	97.14%	

f_measure: 98.54% +/- 0.51% (micro average: 98.55%) (positive class: Positif)

	true Negatif	true Positif	class precision
pred. Negatif	3142	90	97.22%
pred. Positif	0	3052	100.00%
class recall	100.00%	97.14%	

Gambar 4. Confusion Matrix algoritma SVM dengan Operator SMOTE

Gambar 4 merupakan hasil evaluasi kinerja algoritma SVM dengan menggunakan Operator SMOTE Upsampling dimana nilai akurasi berjumlah 98.57%, nilai presisi berjumlah 100%, nilai recall berjumlah 97,14%, dan nilai *f-measure* berjumlah 98,54%. Selanjutnya nilai *true Negatif* dan *pred. Negatif* berjumlah 3142, nilai *true Positif* dan *pred. Negatif* berjumlah 90, nilai *true Negatif* dan *pred. Positif* berjumlah 0, dan nilai *true Positif* dan *pred. Positif* berjumlah 3052. Adapun nilai AUC ialah 100% dan nilai t-Test berjumlah 98,6%. Berbeda halnya dengan hasil evaluasi kinerja algoritma SVM tanpa menggunakan Operator SMOTE Upsampling, dimana nilai akurasi berjumlah 95.71%, nilai presisi berjumlah 95,71%, nilai recall berjumlah 100%, dan nilai *f-measure* berjumlah 97,81%. Selanjutnya nilai *true Negatif* dan *pred. Negatif* berjumlah 2, nilai *true Positif* dan *pred. Negatif* berjumlah 0, nilai *true Negatif* dan *pred. Positif* berjumlah 141, dan nilai *true Positif* dan *pred. Positif* berjumlah 3142. Selain itu, nilai AUC ialah 91,1% dan nilai t-Test berjumlah 95,7%. Hal ini menunjukkan bahwa algoritma SVM menunjukkan performa yang lebih baik dibandingkan dengan NBC, hal tersebut dapat dilihat dari nilai ROC pada gambar berikut.



Gambar 5. Nilai ROC algoritma NBC dan SVM

Gambar 5 menunjukkan bahwa berdasarkan nilai ROC algoritma SVM memiliki performa yang lebih optimal dibandingkan dengan NBC. Meskipun demikian, kedua algoritma menunjukkan performa yang baik apabila menggunakan operator SMOTE sebagai penyeimbang data. Dengan demikian, dapat diketahui bahwa pada tahap pemodelan, algoritma SVM dapat digunakan sebagai model untuk menganalisis sentimen wisatawan terhadap kualitas layanan di hotel dan resort berdasarkan data teks (ulasan) sesuai pengalaman menggunakan jasa akomodasi di Aston Sunset Beach Resort, Holiday Resort Lombok, Qunci Villas Hotel, Sheraton Senggigi Beach Resort, Sudamala Beach Resort, dan The Oberoi Beach Resort. Hal ini menunjukkan bahwa penggunaan algoritma SVM dalam analisis sentimen wisatawan terhadap kualitas layanan hotel dan resort berdasarkan 7526 data ulasan yang telah diproses menggunakan aplikasi Rapidminer menunjukkan bahwa sentimen positif lebih dominan dibandingkan dengan sentimen negatif. Dengan demikian dapat diketahui bahwa manajemen pelayanan hotel dan resort di Lombok, mampu meningkatkan kunjungan wisatawan mancanegara sekaligus meningkatkan citra pariwisata Lombok dan Indonesia.

Rekomendasi untuk mengoptimalkan manajemen pelayanan di hotel dan resort berdasarkan studi kasus penelitian ini ialah sebagai berikut : pertama, sentimen dan ulasan wisatawan perlu diklasifikasi berdasarkan setiap dimensi dalam metode SERVQUAL yakni bukti fisik (*tangibles*), daya tanggap (*responsiveness*), kehandalan (*reliability*), jaminan (*assurance*), empati (*empathy*) sehingga dapat dijadikan sebagai rujukan masing-masing bidang atau divisi di hotel atau resort; kedua, hotel dan resort di Lombok perlu memperhatikan ulasan terkait dengan jaminan (*assurance*), empati (*empathy*) dalam layanan yang diberikan; ketiga, optimalisasi manajemen perlu dilakukan secara berkala untuk mengidentifikasi, mengevaluasi, dan mengantisipasi perubahan preferensi wisatawan terkait dengan jasa akomodasi. Dengan demikian, dapat diketahui bahwa metode CRISP-DM dan SERVQUAL dapat digunakan untuk memberikan rekomendasi manajemen layanan dan optimalisasi kualitas produk di industri perhotelan, menggunakan algoritma SVM.

4. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa algoritma SVM memiliki performa yang lebih baik dibandingkan NBC, dimana nilai akurasi berjumlah 98,57%, nilai presisi berjumlah 100%, nilai recall berjumlah 97,14%, dan nilai *f-measure* berjumlah 98,54%. Adapun nilai AUC ialah 100% dan nilai t-Test berjumlah 98,6%. Berbeda halnya dengan hasil evaluasi kinerja algoritma NBC tanpa menggunakan Operator SMOTE Upsampling, dimana nilai akurasi berjumlah 95,71%, nilai presisi berjumlah 95,71%, nilai recall berjumlah 100%, dan nilai *f-measure* berjumlah 97,81%. Selain itu, nilai AUC ialah 91,1% dan nilai t-Test berjumlah 95,7%. Selanjutnya, berdasarkan metode SERVQUAL terdapat lima kata populer dalam data ulasan wisatawan terhadap kualitas layanan hotel dan resort di Lombok ialah *great*, *pool*, *room*, *beach*, dan *hotel* yang menunjukkan kepuasan terhadap fasilitas ruangan dan amenitas hotel, serta pemandangan sekitar. Selanjutnya, sepuluh kata populer dalam data ulasan wisatawan terhadap

kualitas layanan hotel dan resort di Lombok ialah *breakfast, lombok, food, resort, good, great, pool, room, beach*, dan *hotel* yang dapat diinterpretasikan bahwa produk berupa makanan yang disajikan kepada tamu hotel mendapatkan apresiasi yang positif. Hal ini menunjukkan implementasi pelayanan prima di Aston Sunset Beach Resort, Holiday Resort Lombok, Qunci Villas Hotel, Sheraton Senggigi Beach Resort, Sudamala Beach Resort, dan The Oberoi Beach Resort, perlu dipertahankan dan ditingkatkan.

REFERENCES

- [1] D. Firmansyah, D. Saepuloh, and Dede, "Daya Saing : Literasi Digital dan Transformasi Digital," *J. Financ. Bus. Digit.*, vol. 1, no. 3, pp. 237–250, 2022, doi: 10.55927/jfbd.v1i3.1348.
- [2] R. A. Purnama, A. A. Rifai, I. Firmansyah, and R. Rijanto, "Model Kinerja Kewirausahaan di Industri 4.0 : Kompetensi Inti dan Transformasi Digital," *SEIKO J. Manag. Bus.*, vol. 5, no. 2, pp. 121–137, 2022, doi: 10.37531/sejaman.v5i2.3010.
- [3] S. Asrol, R. Lidyah, T. Hartini, and Muhammadinah, "Peran Percepatan Transformasi Digital untuk Keberlanjutan Bisnis dan Pemulihan Ekonomi Pelaku UMKM di Kota Palembang Pascapandemi," *J. Intelekt. Keislaman, Sos. dan Sains*, vol. 11, no. 2, pp. 242–246, 2022, doi: 10.19109/intelektualita.v11i2.14685.
- [4] N. Ainun, V. Sihombing, and Masrizal, "Application of the Serqual Method to Analyzed the Satisfaction Level of Hotel Services and Facilities," *Sink. J. dan Penelit. Tek. Inform.*, vol. 5, no. 2, pp. 239–245, 2021, doi: 10.33395/sinkron.v5i2.10862.
- [5] Y. Prananda, D. R. Lucitasari, and M. S. A. Khannan, "Penerapan Metode Service Quality (SERVQUAL) Untuk Peningkatan Kualitas Pelayanan Pelanggan," *Opsi*, vol. 12, no. 1, pp. 1–11, 2019, doi: 10.31315/opsi.v12i1.2827.
- [6] A. Sibua, A. Umar, W. Sibua, and D. Weka, "Analysis of Service Quality using Service Quality and SWOT Method in Border Area (Case Study : PT. JABABEKA Morotai)," *J. Ind. Eng. Manag.*, vol. 7, no. 3, pp. 250–260, 2022.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] M. A. Burhanuddin, R. Ismail, N. Izzaimah, A. A.-J. Mohammed, and N. Zainol, "Analysis of Mobile Service Providers Performance Using Naive Bayes Data Mining Technique," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 6, p. 5153, 2018, doi: 10.11591/ijece.v8i6.pp5153-5161.
- [12] A. Aakash and A. Gupta Aggarwal, "Assessment of Hotel Performance and Guest Satisfaction through eWOM: Big Data for Better Insights," *Int. J. Hosp. Tour. Adm.*, vol. 23, no. 2, pp. 317–346, 2022, doi: 10.1080/15256480.2020.1746218.
- [13] A. Firmansyah and Eriswandi, "Analisis Sentimen Terhadap Operator Seluler Telkomsel Menggunakan Algoritma Naive Bayes," *SIGMA – J. Teknol. Pelita Bangsa*, vol. 10, no. 4, pp. 108–116, 2020.
- [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] 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] A. Darmawan, N. Kustian, and W. Rahayu, "Implementasi Data Mining Menggunakan Model SVM untuk Prediksi Kepuasan Pengunjung Taman Tabebuaya," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 2, no. 3, p. 299, 2018, doi: 10.30998/string.v2i3.2439.
- [17] N. P. A. Aryawati, "Digital Tourism: Strategi Menarik Wisatawan Milenial Ke NTB," *J. Ilm. Hosp.*, vol. 9, no. 1, pp. 87–92, 2020, [Online]. Available: <http://stp-mataram.e-journal.id/JHI>
- [18] R. Kurniansah and J. J. Purnama, "Komponen-Komponen Pendukung Pariwisata Kuta Lombok Provinsi Nusa Tenggara Barat," *J. Hosp. dan Pariwisata*, vol. 6, no. 1, pp. 1–6, 2020, doi: 10.30813/jhp.v6i1.2029.
- [19] M. O. Tumbade, K. D. Hartomo, and M. N. N. Sitokdana, "Sistem Informasi Dalam Mendukung Kematangan Planning Smart City E-Government Untuk Peningkatan Kapabilitas Pelayanan Pemerintah Lombok Timur Terhadap Masyarakat," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2137–2152, 2022, doi: 10.35957/jatisi.v9i3.2225.
- [20] A. P. Haster and K. D. Hartomo, "Analisis Tingkat Kematangan Smart City Kabupaten Lombok Utara Menggunakan COBIT 2019," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1459, 2022, doi: 10.30865/mib.v6i3.4344.
- [21] K. M. Sukiahy and C. V. R. Jummi, "Klasifikasi Dan Visualisasi Pariwisata Aceh Dengan Menggunakan Data Pada Twitter," *JUSIM (Jurnal Sist. Inf. Musirawas)*, vol. 5, no. 2, pp. 148–157, 2020, doi: 10.32767/jusim.v5i02.1062.
- [22] D. A. Muthia, "Komparasi Algoritma Klasifikasi Text Mining Untuk Analisis Sentimen Pada Review Restoran," *J. PILAR Nusa Mandiri*, vol. 14, no. 1, pp. 69–74, 2018.
- [23] M. Kusmira, "Penerapan Data Mining Pengajuan Pembiayaan Perumahan (Consumer Financing) Individual Menggunakan Algoritma C4.5," *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 66–71, 2019, doi: 10.31294/jki.v7i1.126.
- [24] 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.
- [25] R. W. Astuti, S. Puspitorini, and F. K. Akbar, "Implementasi RapidMiner pada Analisa Ketepatan Bidang Pekerjaan Alumni Prodi Teknik Informatika STMIK Nurdin Hamzah Jambi," *Fortech*, vol. 3, no. 2, pp. 58–64, 2019.
- [26] M. W. Pertiwi, M. F. Adiwisatra, and D. Supriadi, "Analisa Komparasi Menggunakan 5 Metode Data Mining dalam Klasifikasi Persentase Wanita Sudah menikah di Usia 15-49 yang Memakai Alat KB (Keluarga Berencana)," *J. Khatulistiwa*



- Inform.*, vol. 7, no. 1, pp. 37–42, 2019, doi: 10.31294/jki.v7i1.5741.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] F. Nalle, R. P. C. Fanggidae, and P. Y. Amtiran, “Analisis Harga Terhadap Tingkat Hunian Hotel dalam Mendukung Industri Pariwisata di Kota Kupang,” *GLORY J. Ekon. Ilmu Sos.*, vol. 3, no. 1, pp. 1–13, 2022, [Online]. Available: <https://ejurnal.undana.ac.id/index.php/glory/article/view/7664>
- [32] S. Wijayanto, D. A. Prabowo, D. Y. Kristiyanto, and M. Y. Fathoni, “Analisis Sentimen Berbasis Aspek pada Layanan Hotel di Wilayah Kabupaten Banyumas dengan Word2Vec dan Random Forest,” *JPIT J. Pengemb. IT*, vol. 8, no. 1, pp. 1–3, 2023.
- [33] W. Fauzzia, R. Rahmayani, and R. Handayani, “Peranan Standar Operasional Prosedur dalam Meningkatkan Kinerja Cook di Kitchen Department Hotel Harper Purwakarta by Aston,” *J. Manaj. Perhotelan Dan Pariwisata*, vol. 5, no. 2, pp. 100–104, 2022, [Online]. Available: <https://ejournal.undiksha.ac.id/index.php/JMPP/article/view/46858>