

Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel Berbasis Algoritma Decision Tree (DT)

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: 06/12/2023; Accepted: 19/12/2023; Published: 22/12/2023

Abstrak–Pendekatan Cross-Industry Standard Process for Data Mining (CRISP-DM) sangat relevan digunakan dalam mengidentifikasi tantangan bisnis dan menghasilkan rekomendasi berupa model yang tepat untuk menghadapi pelbagai tantangan bisnis. Klasifikasi sentimen dibutuhkan untuk mengidentifikasi dan menganalisis trend dan preferensi konsumen agar dapat merencanakan strategi mitigasi risiko yang berhubungan dengan keberlanjutan usaha. Penelitian ini mengadopsi metode CRISP-DM dalam mengklasifikasi sentimen tamu hotel melalui data ulasan di platform agoda serta menganalisis data sentimen berdasarkan perilaku pembelian produk dan layanan terkait. Adapun, tahapan dalam metode CRISP-DM ialah sebagai berikut : tahap pemahaman konteks bisnis (business understanding); tahap pemahaman karakteristik data (data understanding); tahap pemodelan (modeling); tahap evaluasi (evaluation); dan tahap implementasi (deployment). Hasil penelitian ini menunjukkan bahwa terdapat sepuluh kata yang menjadi perhatian tamu hotel serta didominasi sentimen positif, yaitu shopping, great, stay, staff, clean, location, room, good, mall, hotel. Hasil klasifikasi menggunakan algoritma DT menunjukkan performa yang baik dengan nilai akurasi 93,91%, nilai presisi 90,98%, dan nilai recall 97, 77%. Selanjutnya, nilai AUC sebesar 0,943 atau 94,3% serta nilai f_measure 94, 18%. Selanjutnya, data hasil analisis sentimen dapat dikembangkan menjadi aplikasi pendukung Customer Relationship Management (CRM) untuk menganalisis data riwayat pembelian tamu yang berhubungan dengan sentimen, negara asal, tipe tamu, tipe kamar, lama menginap berdasarkan hari, bulan dan tahun. Dengan demikian, strategi pemasaran layanan akomodasi hotel dapat dioptimalkan untuk personalisasi serta meningkatkan minat serta intensi menginap kembali.

Kata Kunci: CRISP-DM; Klasifikasi; Sentimen; Pemasaran; Akomodasi; CRM

Abstract–The Cross-Industry Standard Process for Data Mining (CRISP-DM) approach is very relevant in identifying business challenges and producing recommendations in the form of appropriate models to face various business challenges. Sentiment classification is needed to identify and analyze consumer trends and preferences in order to plan risk mitigation strategies related to business sustainability. This study adopts the CRISP-DM method in classifying hotel guest sentiment through review data on the Agoda platform and analyzing sentiment data based on the purchase behavior of related products and services. Meanwhile, the stages in the CRISP-DM method are as follows: the stage of understanding the business context (business understanding), the stage of understanding data characteristics (data understanding), the modeling stage (modeling), the evaluation stage, and the implementation stage (deployment). The results of this study show that ten words are the attention of hotel guests and are dominated by positive sentiment, namely shopping, great, stay staff, clean, location, room, good, mall, and hotel. The classification results using the DT algorithm showed good performance with an accuracy value of 93.91%, a precision value of 90.98%, and a recall value of 97.77%. In addition, the AUC value is 0.943 or 94.3%, and the f-measure value is 94.18%. Furthermore, sentiment analysis data can be developed into a Customer Relationship Management (CRM) supporting application to analyze guest purchase history data related to sentiment, country of origin, guest type, room type, and length of stay by day, month, and year. Thus, the marketing strategy of hotel accommodation services can be optimized for personalization and increase interest and intention of returning stays.

Keywords: CRISP-DM; Classification; Sentiment; Marketing; Accommodation; CRM

1. PENDAHULUAN

Pendekatan Cross Industry Standard Process for Data Mining (CRISP-DM) dapat digunakan dalam mengevaluasi performa lembaga serta menghasilkan rekomendasi untuk mengoptimalkan kinerja sesuai dengan target capaian masing-masing. Pendekatan CRISP-DM dapat digunakan untuk menganalisis dan menentukan customer behavior dalam pembelian produk [1]. Disisi lain, metode CRISP-DM dapat digunakan untuk clustering film populer pada platform Netflix sehingga dapat membantu pengguna menentukan pilihan film yang sesuai dengan preferensi [2]. Adapun, metode CRISP-DM melalui model clustering dapat digunakan untuk menganalisis pendapatan bersih bulanan pekerja informal berdasarkan data Kabupaten/Kota menjadi tiga kelompok kategori yakni rendah, sedang dan tinggi [3]. Pendekatan CRISP-DM sangat efektif digunakan, mempertimbangkan alur dan proses yang terstruktur pada setiap tahapan [4]. Hal ini menunjukkan bahwa pendekatan CRISP-DM dapat digunakan sebagai alur atau proses penyelesaian masalah internal organisasi atau perusahaan. Dengan demikian, penelitian ini menggunakan CRISP-DM dalam mengidentifikasi dan menganalisis perilaku pembelian layanan akomodasi hotel berdasarkan studi kasus informasi produk dan layaann digital Menara Peninsula Hotel pada platform agoda.

Inovasi bisnis dan pemasaran produk serta layanan di industri perhotelan memantik perubahan dalam pengolahan data digital yang berhubungan dengan strategi pemasaran. Pendekatan machine learning dapat digunakan untuk memprediksi loyalitas tamu hotel [5]. Disisi lain, Pendekatan text mining dapat menghasilkan

rekomendasi terhadap pihak manajemen hotel terkait atribut-atribut yang menjadi atensi tamu hotel dalam pengambilan keputusan menginap seperti hotel, service, room, food and beverage, location, security, price/value serta atribut lainnya yang berhubungan dengan preferensi dan kebutuhan tamu [6]. Adapun, kepuasan tamu dapat ditinjau dari kualitas layanan, oleh sebab itu penyedia layanan akomodasi perlu memerhatikan persepsi risiko terhadap loyalitas tamu agar dapat mempertahankan pelanggan [7]. Hal ini menunjukkan bahwa persepsi pelanggan terhadap produk dan layanan akomodasi hotel dapat dianalisis melalui fitur ulasan menggunakan pendekatan text mining dan machine learning sehingga dapat menghasilkan rekomendasi sesuai dengan atribut yang menarik atensi tamu hotel. Dengan demikian, pendekatan CRISP-DM dapat digunakan untuk menyelesaikan masalah yang berhubungan dengan kepuasan dan loyalitas tamu hotel dengan merekomendasikan model yang berhubungan dengan Customer Relationship Management (CRM).

Salah satu upaya untuk mempertahankan hubungan pelanggan dengan penyedia layanan akomodasi hotel ialah mengelola rencana strategis untuk mengoptimalkan bidang Customer Relationship Management (CRM). Faktor-faktor yang memengaruhi perilaku konsumen dalam keputusan menginap di Hotel, faktor-faktor yang dimaksud dapat diklasifikasikan berdasarkan budaya, sosial, pribadi dan psikologis [8]. Disisi lain, CRM memiliki pengaruh yang signifikan terhadap kinerja pemasaran hotel, oleh sebab itu pemanfaatan media digital sebagai sarana untuk berinteraksi dengan pelanggan dapat memberikan pengalaman dan meningkatkan kepuasan hingga loyalitas [9]. Adapun, kualitas layanan, CRM, dan citra merek memiliki pengaruh terhadap loyalitas tamu menginap di hotel [10]. Hal ini menunjukkan bahwa pemanfaatan platform digital sebagai strategi pemasaran dapat memberikan pengalaman konsumen sehingga meningkatkan kepuasan dan loyalitas. Dengan demikian, perilaku pembelian layanan akomodasi hotel melalui platform Agoda perlu dikaji secara mendalam untuk menganalisis interaksi antara tamu yang telah mengulas produk dan layanan selama menginap dengan respon dari pihak hotel.

Platform digital yang digunakan sebagai media pemasaran layanan hotel telah banyak dikaji dalam berbagai perspektif keilmuan. Wiastuti dan Wiliam menggunakan data dari platform TripAdvisor dan Agoda untuk mengidentifikasi faktor faktor yang memengaruhi kepuasan dan ketidakpuasan pelanggan [11]. Selain itu, strategi komunikasi pemasaran layanan akomodasi hotel yang paling efektif ialah melalui platform digital seperti TripAdvisor, Agoda, Traveloka, serta Booking.com [12]. Adapun, faktor-faktor seperti authenticity, interactivity, involvement and costumer engagement berperan penting dalam prediksi loyalitas konsumen pada pemesanan layanan akomodasi hotel melalui aplikasi traveloka [13]. Hal ini menunjukkan bahwa platform digital yang digunakan sebagai media pemasaran layanan akomodasi menyimpan pelbagai jenis data pembelian, serta informasi pembeli yang dapat digunakan untuk menganalisis kepuasan dan loyalitas konsumen. Dengan demikian, penelitian ini menggunakan platform digital Agoda sebagai sumber data teks yang akan digunakan dalam menganalisis perilaku pembelian layanan akomodasi hotel, melalui studi kasus pada Menara Peninsula Hotel Jakarta.

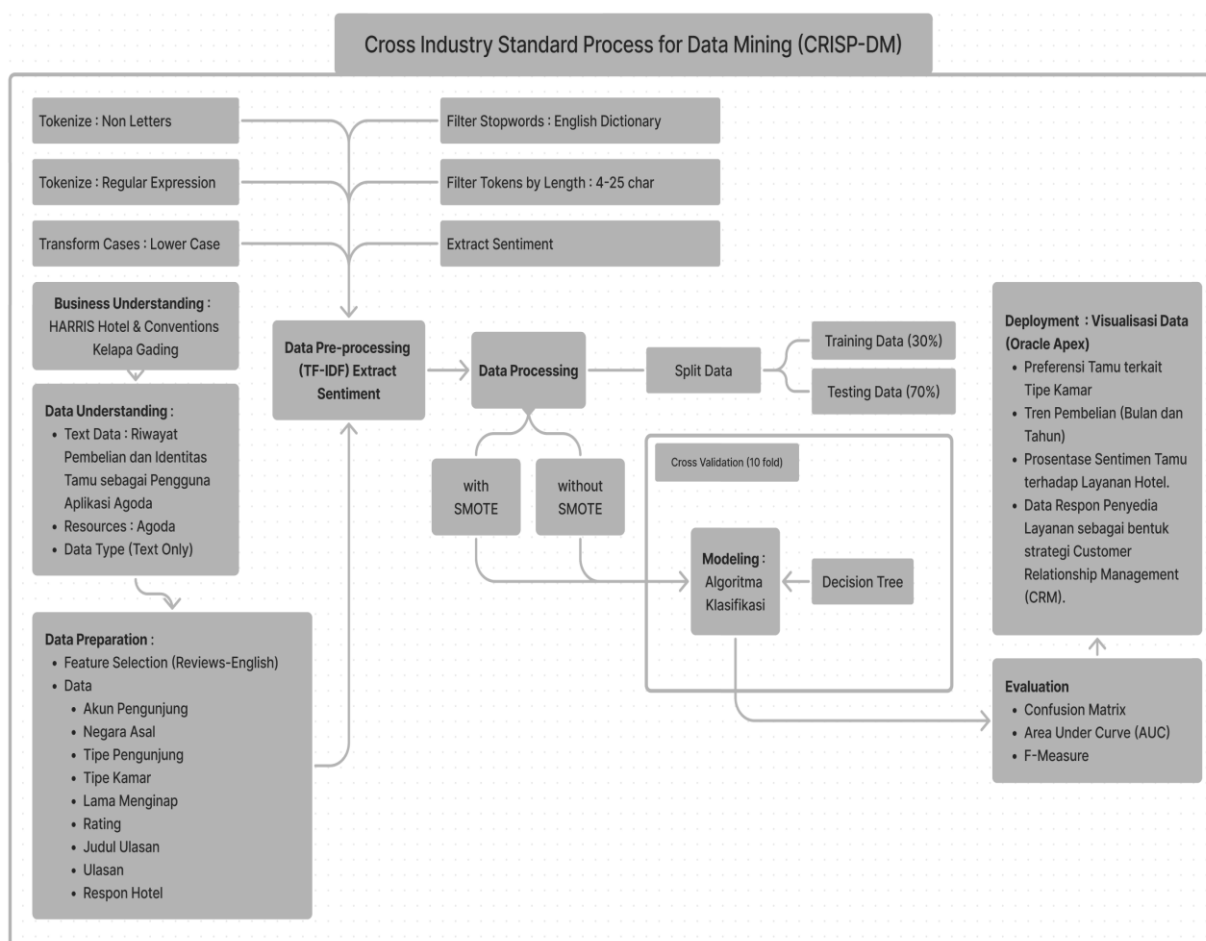
Beberapa penelitian terdahulu menunjukkan bahwa data teks yang dikumpulkan dari platform digital Agoda dapat digunakan dalam analisis dan klasifikasi sentimen pelanggan melalui pendekatan data mining maupun machine learning. Data ulasan dari TripAdvisor dan Agoda untuk melakukan pemodelan TOPIK menggunakan Latent Dirichlet Allocation (LDA) dan algoritma Support Vector Machine (SVM) untuk mengidentifikasi topik-topik yang berhubungan dengan layanan akomodasi hotel serta menjadi aspek penting yang memengaruhi kepuasan dan loyalitas tamu [14]. Analisis sentimen konsumen dapat menggunakan data dari pelbagai platform digital seperti Agoda, Airyrooms, Oyo, dan Reddoorz melalui metode klasifikasi berbasis algoritma Naïve Bayes Classifier (NBC) [15]. Hal ini menunjukkan bahwa perkembangan riset terkait dengan relasi antara penyedia layanan akomodasi hotel dengan pelanggan mengalami perubahan dari pendekatan strategi komunikasi melalui media konvensional menjadi media digital. Dengan demikian, penggunaan data teks berupa ulasan di pelbagai platform penyedia layanan akomodasi dapat dikelola lebih lanjut menggunakan metode CRISP-DM untuk analisis perilaku pembelian pelanggan serta implementasi strategi CRM. Mempertimbangkan hal tersebut maka penelitian ini akan menggunakan data teks ulasan tamu yang memiliki pengalaman reservasi di aplikasi Agoda serta memiliki pengalaman menginap di Menara Peninsula Hotel Jakarta.

Penelitian ini menawarkan perspektif yang berbeda dari beberapa kajian tentang perilaku pembelian layanan akomodasi hotel sebelumnya. Berdasarkan metode yang digunakan, CRISP-DM menawarkan keunggulan dalam proses penyelesaian masalah yang berhubungan dengan pemasaran produk dan layanan akomodasi. Penelitian terdahulu menunjukkan bahwa pendekatan CRISP-DM sangat efektif digunakan dalam menganalisis sentimen tamu hotel dan restoran untuk mengidentifikasi teks populer dalam ulasan tamu tentang produk dan layanan berdasarkan klasifikasi kelas negatif dan positif [16]–[24]. Selanjutnya, pendekatan CRISP-DM digunakan secara luas dalam menganalisis perilaku pembelian produk dan layanan yang secara spesifik berhubungan dengan pengambilan keputusan pembelian layanan akomodasi dan amenities pendukung aktivitas wisata [25]–[27]. Secara konseptual, kerangka kerja CRISP-DM menjadi alur atau proses utama kemudian di dalam proses tersebut terdapat berbagai model pendukung keputusan yang digunakan untuk prediksi pengambilan keputusan sesuai kriteria dan bobot yang ditentukan [28]–[32]. Hal ini menunjukkan bahwa pendekatan CRISP-DM dapat disesuaikan dengan model atau konsep yang berhubungan dengan upaya untuk menyelesaikan masalah bisnis hingga prediksi perilaku konsumsi berbasis data riwayat pembelian. Mempertimbangkan hal tersebut maka penelitian ini menawarkan gagasan untuk mendiskusikan hasil implementasi CRISP-DM dalam analisis perilaku pembelian layanan akomodasi hotel berdasarkan konsep Customer Relationship Management (CRM).

2. METODOLOGI PENELITIAN

2.1 Metode CRISP-DM

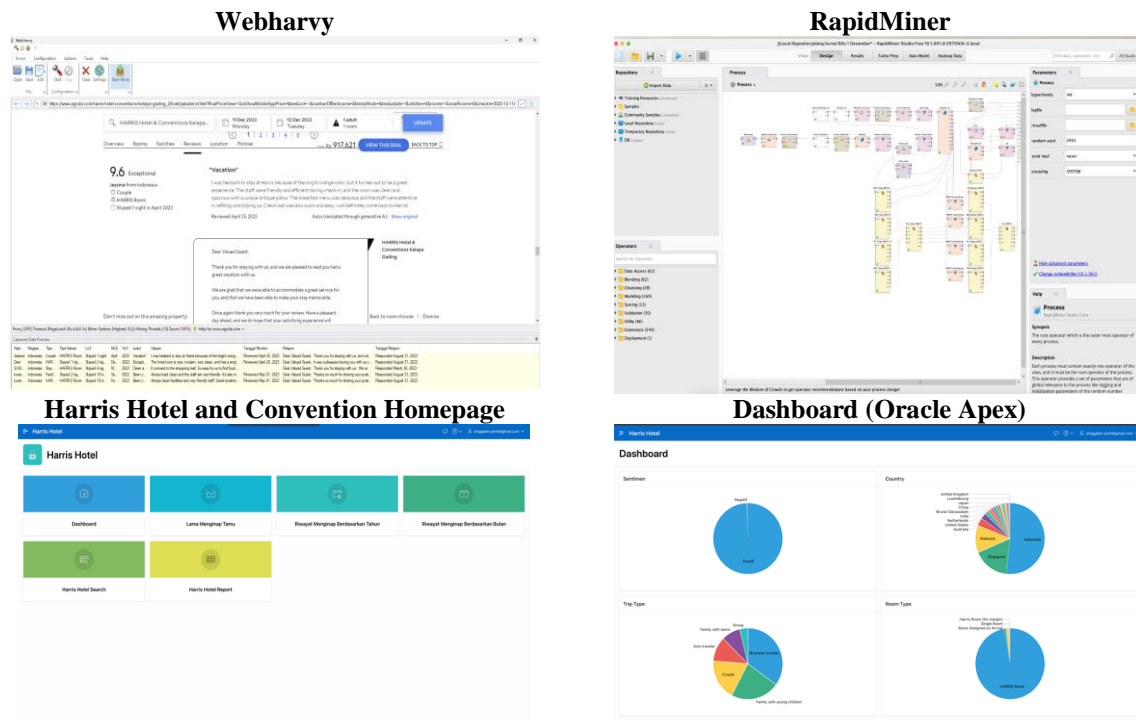
Penelitian ini mengadopsi pendekatan CRISP-DM yang terdiri dari tahapan berikut : pertama, business understanding; kedua, data understanding; ketiga, modelling; keempat, evaluation; kelima, deployment. Terdapat beberapa pertimbangan penggunaan pendekatan CRISP-DM dalam analisis perilaku pembelian layanan akomodasi sebagai berikut : pertama, proses pengumpulan data dari platform Agoda menyediakan data teks terkait riwayat pembelian yang berhubungan dengan identitas tamu, asal negara, lama waktu menginap, hingga jenis kamar yang digunakan; kedua, data teks dapat dikumpulkan serta dibersihkan menggunakan instrumen pendukung seperti webharvy dan Rapidminer untuk menganalisis proses evaluasi performa pembelajaran mesin (machine learning); ketiga, pada tahap pemodelan dapat dievaluasi model klasifikasi atau algoritma dengan tingkat akurasi, presisi, recall, serta Area Under Curve (AUC) yang terbaik; keempat, evaluasi hasil pengolahan data dapat menghasilkan rekomendasi produk dan layanan populer berdasarkan perilaku pembelian, serta segmen pasar yang relevan; kelima, implementasi hasil analisis perilaku pembelian layanan akomodasi dapat menjadi strategi alternatif untuk menyelesaikan masalah pemasaran di unit bisnis. Dengan demikian, pendekatan CRISP-DM sangat relevan digunakan dalam penelitian ini. Adapun, alur atau proses CRISP-DM dapat dilihat pada gambar 1 berikut.



Gambar 1. Cross-Industry Standard Process for Data Mining (CRISP-DM)

Gambar 1 merupakan proses implementasi pendekatan CRISP-DM dalam klasifikasi sentimen dan analisis perilaku pembelian layanan akomodasi hotel berbasis algoritma Decision Tree (DT). Pada tahap business understanding, data terkait dengan riwayat pembelian layanan akomodasi di aplikasi Agoda dikumpulkan menggunakan aplikasi Webharvy dengan konfigurasi data teks yang berhubungan dengan nama akun pengguna, negara asal pengguna, tipe pengunjung atau tamu, tipe kamar yang digunakan, lama menginap berdasarkan jumlah hari, bulan, dan tahun, serta penilaian atau rating. Selanjutnya, data teks berupa judul ulasan, detail ulasan, tanggal ulasan, respon pihak hotel terhadap ulasan konsumen, dan tanggal respon dikumpulkan untuk divisualisasikan menggunakan aplikasi oracle apex. Pada tahap ini, tujuan pengumpulan dan pengolahan data teks berdasarkan riwayat pembelian layanan akomodasi, ialah untuk menghasilkan informasi terkait preferensi tamu terhadap produk dan layanan yang disediakan, memperoleh informasi tentang rata-rata lama menginap tamu berdasarkan hari, bulan dan tahun. Adapun, respon pihak penyedia layanan menanggapi ulasan tamu dapat menjadi strategi

untuk menjaga relasi antara penyedia layanan dengan tamu agar dapat meningkatkan angka menginap kembali. Dengan demikian, proses bisnis penyedia layanan akomodasi dapat dikaji secara komprehensif sehingga dapat dioptimalkan melalui kajian ini. Adapun, instrumen dan aplikasi pendukung yang digunakan dalam implementasi pendekatan CRISP-DM dapat dilihat pada gambar 2 berikut.



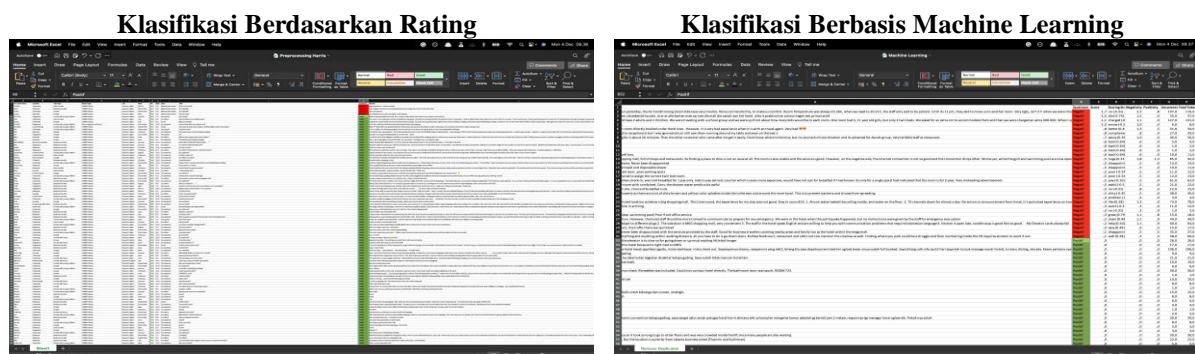
Gambar 2. Aplikasi Pendukung dalam Implementasi CRISP-DM Pada Tahap Business Understanding

Gambar 2 merupakan aplikasi pendukung yang digunakan dalam proses pengumpulan, pengolahan dan visualisasi data hasil implementasi pendekatan CRISP-DM. Selanjutnya, pada tahap data understanding hasil pengumpulan data disiapkan dengan memanfaatkan beberapa operator dalam aplikasi Rapidminer untuk membersihkan data yang telah dikumpulkan dalam format .xlsx maupun .csv. Sebelum diproses menggunakan aplikasi Rapidminer, data duplikat telah dihilangkan menggunakan aplikasi Microsoft Excel dengan memanfaatkan tools remove duplicate, selanjutnya data disortir dan diseleksi agar tidak ada data yang kosong (not null). Selanjutnya, hasil pembersihan data di Microsoft Excel dapat dilanjutkan menggunakan aplikasi Rapidminer melalui beberapa operator sebagai berikut : pertama, operator tokenize pertama, untuk menghilangkan simbol (non-letters); kedua, operator tokenize kedua, untuk menghilangkan simbol (regular expression); ketiga, operator transform cases, untuk mengubah teks menjadi lowercase; keempat, operator filter tokens by length, membatasi jumlah karakter kata (4-25 char); kelima, operator filter stopwords, menghilangkan kata-kata yang tidak memiliki makna dalam analisis sentimen; keenam, operator extract sentiment, untuk memberikan score terhadap setiap kata yang mengandung makna negatif dan positif. Selain itu, terdapat operator word to list yang digunakan untuk menampilkan data dalam bentuk 10 kata populer yang tergolong kelas negatif dan kelas positif. Dengan demikian, tahap data understanding menekankan pemahaman karakteristik data, serta fungsi dari operator yang digunakan untuk memperdalam analisis sentimen dan perilaku pembelian layanan akomodasi hotel. Berdasarkan hasil pemahaman karakteristik data, maka dapat dilanjutkan ke tahap data processing untuk pemodelan berbasis algoritma Decision Tree (DT). Data yang digunakan adalah data sentimen yang telah diekstrak menggunakan operator extract sentimen dengan pengaturan jumlah data uji sebanyak 70% (402) dan data latih sebanyak 30% (172) dari total data teks yang telah diekstrak sebanyak 574 untuk pemodelan. Meskipun demikian, total data yang telah dikumpulkan berjumlah 20210 ulasan, yang telah dibersihkan dan diseleksi untuk pemodelan data menjadi 574 ulasan.

2.2 Analisis Sentimen Menggunakan Algoritma Decision Tree (DT)

Pada tahap pemodelan data, terdapat 574 data ulasan yang akan digunakan dalam proses pengukuran performa algoritma Decision Tree (DT). Pada tahap pengujian, operator Synthetic Minority Oversampling Technique (SMOTE) digunakan untuk mengatasi permasalahan ketimpangan data sehingga dapat meningkatkan akurasi dari performa algoritma Decision Tree (DT). Apabila kelas sentimen ditentukan berdasarkan rating yang diberikan oleh pengguna aplikasi, maka terdapat 4 data ulasan yang memiliki nilai di bawah lima sehingga diklasifikasikan sebagai sentimen negatif, sedangkan jumlah data ulasan di atas nilai lima ialah 570, sehingga terjadi ketidakseimbangan data ketika hendak dilanjutkan ke proses pengukuran performa algoritma. Penggunaan

operator extract sentiment merupakan bagian dari proses machine learning dimana mesin secara otomatis akan melakukan proses kalkulasi score pada setiap kata yang mengandung unsur negatif dan positif. Dengan demikian, hasil kalkulasi data menunjukkan bahwa operator extract sentiment dapat mendeteksi 30 data ulasan sebagai sentimen negatif, dan 515 data ulasan sebagai sentimen positif. Dengan demikian, hasil pembelajaran mesin (machine learning) dapat diukur performanya melalui algoritma DT untuk meninjau nilai akurasi, presisi, recall, hingga Area Under Curve (AUC), sebelum dijadikan model yang relevan dengan konteks dataset yang digunakan.



Gambar 3. Perbandingan Hasil Klasifikasi Berdasarkan Rating dan Machine Learning

Gambar 3 merupakan gambaran tentang hasil klasifikasi berdasarkan penilaian pengguna (rating), dan hasil klasifikasi berdasarkan machine learning melalui operator extract sentiment. Setelah hasil scoring maka dilakukan pengujian performa algoritma berdasarkan pembagian data 30% data latih dan 70% data uji. Adapun, untuk operator SMOTE digunakan untuk mengatasi persoalan ketidakseimbangan data sebelum diproses ke tahap pengukuran performa algoritma DT. 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 [33]. 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 [34]. 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 :

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

Persamaan (1) 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 (2) sebagai berikut.

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

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 konstansa. 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. Selanjutnya, dilakukan proses evaluasi menggunakan algoritma DT.

Salah satu keunggulan DT ialah Iterative Dichotomizer version (ID3), yaitu model klasifikasi yang berupa pohon keputusan secara top-down dengan cara kerja mengevaluasi semua atribut menggunakan suatu ukuran statistik berupa information gain untuk mengukur efektifitas suatu atribut dalam mengklasifikasi sample data. Dalam algoritma ini, dibutuhkan nilai entropy dan gain, dimana entropy merupakan parameter untuk mengukur jumlah keberagaman atau keberadaan dalam sebuah himpunan data, sedangkan gain merupakan perolehan informasi sebagai ukuran efektifitas suatu atribut. Berikut adalah persamaan untuk mendapatkan nilai entropy dan gain.

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i \times \log_2 p_i \quad (3)$$

$$\text{Gain (S, A)} = S - \sum_{i=1}^n \frac{|S_i|}{|S|} \times S_i \quad (4)$$

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 adlah subset dari nilai Entropy. Selanjutnya, 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.

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

$$\text{Presisi/Specificity} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Recall/Sensitivity} = \frac{TP}{TP+FN} \quad (7)$$

$$f - \text{measure} = \frac{2 \times (\text{Presisi} \times \text{recall})}{\text{presisi} + \text{recall}} \quad (8)$$

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/sensivitas dan presisi sehingga menghasilkan metrik yang efektif untuk mencari kembali informasi dalam himpunan yang tidak seimbang [35]. Dengan demikian dapat diketahui bagaimana performa algoritma DT yang dapat dijadikan sebagai model klasifikasi sentimen sesuai dengan dataset yang digunakan, serta dapat dihubungkan dengan analisis perilaku pembelian berdasarkan trend atau popularitas dan pola konsumsi layanan akomodasi hotel.

3. HASIL DAN PEMBAHASAN

Perilaku pembelian dapat dianalisis secara komprehensif untuk memperoleh gambaran tentang pola serta trend pasar yang memberikan petunjuk kepada pengelola usaha untuk mengoptimalkan strategi pemasaran. Sarudin dan Ismail melakukan tinjauan terhadap hasil ulasan pengguna aplikasi TripAdvisor untuk menganalisis minat pembelian layanan akomodasi hotel [36]. Disisi lain, Sanra menunjukkan bahwa pemenuhan informasi berperan sebagai mediasi antara media pemasaran yang digunakan terhadap minat pembelian layanan akomodasi [37]. Adapun, Hasianah dan Priatma menunjukkan bahwa strategi pemasaran dapat dioptimalkan dengan memanfaatkan data atau informasi terkait dengan produk, harga dan lokasi melalui pelbagai media yang diakses publik [38]. Hal ini berarti bahwa konsumen membutuhkan informasi yang lengkap sebelum mengambil keputusan pembelian, sehingga penyedia produk dan layanan yang mampu mengoptimalkan media pemasaran digital dengan informasi yang jelas dapat memantik minat atau intensi pembelian. Dengan demikian, dapat diketahui bahwa perilaku pembelian dapat dianalisis berdasarkan data riwayat pembelian atau riwayat konsumsi produk melalui kolom ulasan di pelbagai platform e-commerce. Dalam konteks pembelian layanan akomodasi hotel, pelbagai platform seperti TripAdvisor, booking.com serta Agoda telah menyediakan kolom ulasan sebagai strategi untuk meninjau produk dan layanan yang diterima serta memberikan respon kepada tamu agar dapat meningkatkan citra positif penyedia layanan.

Dalam konteks penelitian ini, data yang digunakan merupakan data yang berhubungan dengan riwayat pembelian layanan akomodasi hotel dengan studi kasus di hotel HARRIS Hotel and Convention Gading Kelapa Gading. Data awal yang diperoleh sebelum proses pembersihan data ialah sejumlah 2209 data ulasan serta informasi lain terkait dengan nama akun pengguna, negara asal pengguna sistem, tipe kamar yang digunakan, lama waktu menginap berdasarkan jumlah hari, bulan dan tahun. Selanjutnya, terdapat data berupa penilaian (rating) serta tanggal mengulas. Selain itu, terdapat data berupa tanggapan pihak hotel atas ulasan konsumen sebagai bentuk pengelolaan relasi dengan konsumen (Customer Relationship Management). Berdasarkan hasil penggunaan operator extract sentiment, dapat diketahui bahwa kinerja mesin dalam proses klasifikasi, mengalami perkembangan pesat sehingga dapat digunakan dalam pelbagai upaya untuk analisis sentimen konsumen atau pelanggan hotel terkait produk dan layanan yang spesifik. Pada tabel 1, dapat dilihat proses klasifikasi sentimen positif dan negatif berbasis operator extract sentiment.

Tabel 1. Proses Klasifikasi Operator Extract Sentiment di Rapidminer

Klasifikasi	Ulasan	Score	Detail
Negatif	Usually, no problem when if i were here. But yesterday, the Air Conditioning (room 626)	- 1,71794871794872	no (-0.31) problem (-0.44) trouble (-0.44)

Klasifikasi	Ulasan	Score	Detail
Positif	was very trouble . None can handle this, to make us comfort . Room Temperature was always 25-26C, when we need to 20-21C, the staff only said to be patient. Until its 11 pm, they said to move us to another room. Very Ugly , isn't it!! when we were check on, no body can explain me about this problem . Just be carefull and pls the your temperature condition before you enter your room. When its hot, please request another room.		comfort (0.38) ugly (-0.59) no (-0.31) problem (-0.44) pls (0.08) please (0.33)
	Hotel is well located as it is attached to the Kelapa Gading Mall with shopping, restaurants, beauty outlets (nails, eyelashes) and the movies all right there. Breakfast was varied and plentiful. The omlets and fresh fruit were great . The rooms are basic but have all the amenities required for a comfortable stay. Bed was really comfortable with good pillows, shower was very good . Staff were super friendly and helpful . They seem to have been well trained in customer service. Plan to return when next in Jakarta	6,333333333333333	well (0.28) beauty (0.72) fresh (0.33) great (0.79) comfortable (0.59) comfortable (0.59) good (0.49) good (0.49) super (0.74) friendly (0.56) helpful (0.46) well (0.28)

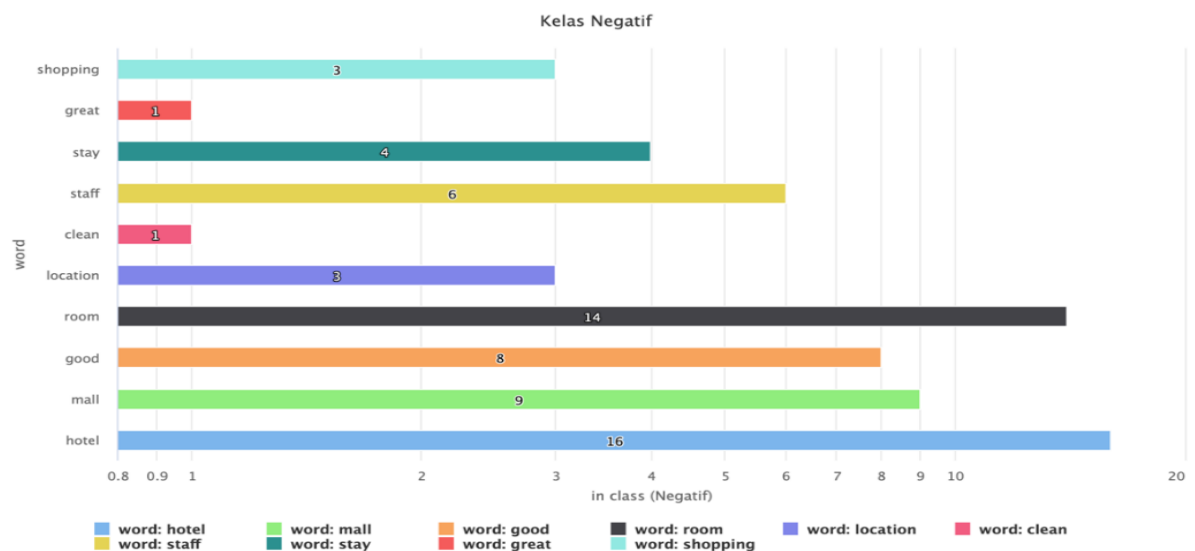
Tabel 1 merupakan proses klasifikasi operator extract sentiment di Rapidminer dimana kata-kata dalam ulasan konsumen akan diberikan score, kemudian diakumulasi sehingga memperoleh total score yang diklasifikasikan ke dalam kelas negatif dan kelas positif. Berdasarkan proses klasifikasi operator extract sentiment di Rapidminer terdapat masalah yang perlu diatasi untuk mengoptimalkan proses analisis sentimen, dimana standar atau dokumen berupa kata dan score yang berhubungan dengan unsur sentimen (negatif dan positif) dibutuhkan mendeteksi setiap kata dalam kalimat maupun paragraf. Selain itu, ulasan dalam bahasa Indonesia yang mengandung unsur sindiran masih belum bisa diproses secara optimal oleh mesin. Meskipun demikian, berdasarakan hasil uji performa algoritma DT dapat diketahui nilai confusion matrix dan AUC sebagaimana tabel 2 berikut.

Tabel 2. Confusion Matrix Algoritma Decision Tree (DT)

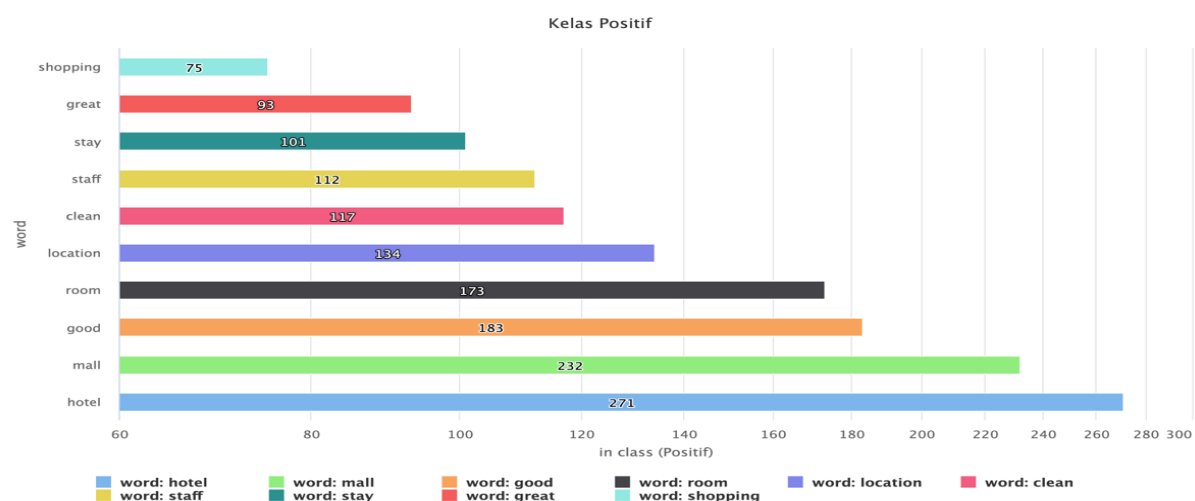
Tanpa Operator SMOTE	Menggunakan Operator SMOTE
PerformanceVector: accuracy: 94.77% +/- 2.12% (micro average: 94.76%) ConfusionMatrix: True: Negatif Positif Negatif: 3 2 Positif: 18 359 AUC (optimistic): 0.995 +/- 0.012 (micro average: 0.995) (positive class: Positif) AUC: 0.572 +/- 0.123 (micro average: 0.572) (positive class: Positif) AUC (pessimistic): 0.150 +/- 0.242 (micro average: 0.150) (positive class: Positif) precision: 95.24% +/- 1.64% (micro average: 95.23%) (positive class: Positif) ConfusionMatrix: True: Negatif Positif Negatif: 3 2 Positif: 18 359 recall: 99.45% +/- 1.16% (micro average: 99.45%) (positive class: Positif) ConfusionMatrix: True: Negatif Positif Negatif: 3 2 Positif: 18 359	PerformanceVector: accuracy: 93.91% +/- 2.63% (micro average: 93.91%) ConfusionMatrix: True: Negatif Positif Negatif: 325 8 Positif: 36 353 AUC (optimistic): 0.996 +/- 0.005 (micro average: 0.996) (positive class: Positif) AUC: 0.943 +/- 0.025 (micro average: 0.943) (positive class: Positif) AUC (pessimistic): 0.890 +/- 0.048 (micro average: 0.890) (positive class: Positif) precision: 90.98% +/- 4.48% (micro average: 90.75%) (positive class: Positif) ConfusionMatrix: True: Negatif Positif Negatif: 325 8 Positif: 36 353 recall: 97.77% +/- 2.22% (micro average: 97.78%) (positive class: Positif) ConfusionMatrix: True: Negatif Positif Negatif: 325 8 Positif: 36 353

Tanpa Operator SMOTE	Menggunakan Operator SMOTE
f_measure: 97.29% +/- 1.10% (micro average: 97.29%) (positive class: Positif)	f_measure: 94.18% +/- 2.37% (micro average: 94.13%) (positive class: Positif)
ConfusionMatrix:	ConfusionMatrix:
True: Negatif Positif	True: Negatif Positif
Negatif: 3 2	Negatif: 325 8
Positif: 18 359	Positif: 36 353

Tabel 3 merupakan kinerja algoritma DT melalui confusion matrix dan AUC (tanpa operator SMOTE dan menggunakan operator SMOTE) dalam klasifikasi data ulasan berbasis operator extract sentiment. Berdasarkan performa algoritma DT tanpa operator SMOTE menunjukkan bahwa nilai akurasi 94,77%, nilai presisi 95,24%, dan nilai recall 99,45%. Selanjutnya, nilai AUC sebesar 0,572 atau 57,2% serta nilai f_measure 97,29%. Hal ini menunjukkan bahwa meskipun nilai akurasi tinggi, namun AUC masih tergolong rendah sehingga perlu digunakan operator SMOTE. Adapun, performa algoritma DT setelah menggunakan SMOTE menunjukkan bahwa nilai akurasi 93,91%, nilai presisi 90,98%, dan nilai recall 97,77%. Selanjutnya, nilai AUC sebesar 0,943 atau 94,3% serta nilai f_measure 94,18%. Berdasarkan nilai confusion matrix algoritma DT setelah menggunakan operator SMOTE dapat diketahui nilai akurasi, presisi dan recall mengalami penurunan namun nilai AUC mengalami peningkatan. Dengan demikian dapat diketahui bahwa pada tahap pemodelan, algoritma DT berbasis operator SMOTE dapat meningkatkan kinerja algoritma sehingga relevan digunakan sebagai model dalam analisis sentimen. Selanjutnya, dapat diketahui 10 kata populer yang diklasifikasikan dalam kelas negatif dan positif.



(a)



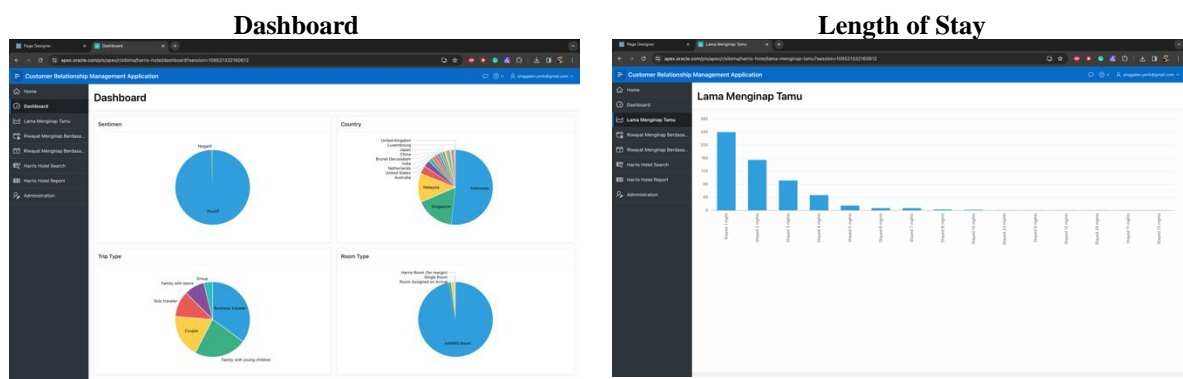
(b)

Gambar 4. Perbandingan Jumlah Kata dalam Data Ulasan Berdasarkan Kelas Negatif (a) dan Positif (b)

Gambar 4 merupakan hasil perbandingan jumlah kata dalam data ulasan berdasarkan kelas negatif dan positif, dapat diketahui bahwa jumlah kelas positif lebih dominan dibandingkan dengan kelas negatif. Meskipun

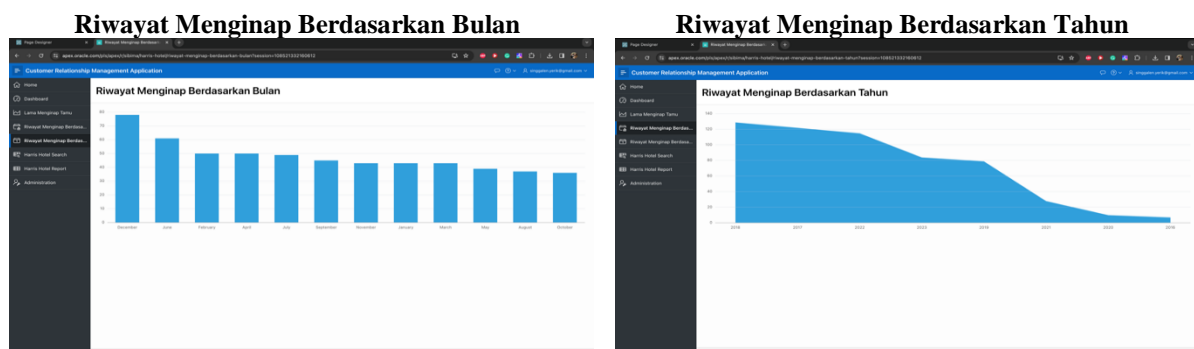
demikian, kata yang tergolong kelas negatif perlu diperhatikan dan dievaluasi untuk mengoptimalkan produk dan layanan akomodasi hotel. Berdasarkan 10 kata populer yang paling sering muncul dalam ulasan ialah sebagai berikut : shopping, great, stay, staff, clean, location, room, good, mall, hotel. Hal ini menunjukkan bahwa pihak hotel perlu meningkatkan kinerja karyawan sehingga dapat memberikan pelayanan prima kepada tamu, memberikan pengalaman kepada tamu selama menginap di hotel, menjaga kebersihan lingkungan hotel, dan kamar yang digunakan. Beberapa pengunjung memberikan ulasan terkait dengan lokasi hotel yang dinilai strategis, terhubung dengan pusat perbelanjaan sehingga memantik minat dan keputusan tamu untuk menginap.

Beberapa penelitian terdahulu menunjukkan bahwa promosi, fasilitas, produk dan harga memiliki pengaruh terhadap minat dan keputusan menginap tamu. Apsarini et al. menunjukkan bahwa faktor-faktor yang memengaruhi keputusan menginap tamu ialah sebagai berikut : harga, promosi, fasilitas, keamanan, lokasi, dan ulasan tamu [39]. Selain itu, Chairani dan Wulansari menunjukkan bahwa brand image memiliki pengaruh terhadap keputusan tamu untuk menginap di hotel [40]. Adapun, Silalahi et al. menunjukkan bahwa keputusan menginap tamu juga dipengaruhi oleh karakteristik demografi [41]. Hal ini menunjukkan bahwa data yang berhubungan dengan demografi dan riwayat pembelian layanan akomodasi dapat digunakan untuk mengidentifikasi produk dan layanan yang paling diminati oleh tamu. Dengan demikian, maka hasil analisis sentimen dapat dikembangkan sebagai aplikasi pendukung untuk mengoptimalkan Customer Relationship Management (CRM), sebagaimana gambar berikut.



Gambar 5. Aplikasi Pendukung Customer Relationship Management (CRM)

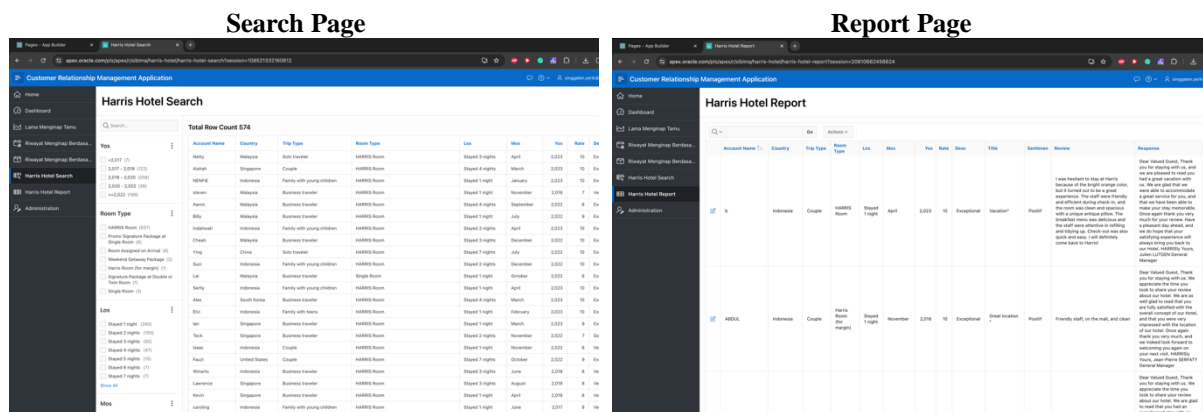
Gambar 5 merupakan pengembangan aplikasi pendukung Customer Relationship Management (CRM) yang diintegrasikan dengan hasil analisis sentimen tamu berdasarkan sepuluh kata populer. Dalam konteks penelitian ini, hasil analisis sentimen menunjukkan bahwa terdapat kata-kata yang menjadi atensi tamu selama menginap yaitu shopping, great, stay, staff, clean, location, room, good, mall, hotel. Selanjutnya, dapat dianalisis data riwayat pembelian tamu yang berhubungan dengan sentimen, negara asal, tipe tamu, tipe kamar, lama menginap berdasarkan hari, bulan dan tahun. Berdasarkan studi kasus penelitian ini, dapat diketahui sebagai berikut : pertama, penilaian tamu didominasi oleh sentimen positif; kedua, mayoritas tamu berasal dari Indonesia, Singapore dan Malaysia; ketiga, tipe tamu yang menginap didominasi kepentingan bisnis, bersama keluarga, bersama pasangan, dan sendiri; keempat, tipe kamar yang paling banyak digunakan ialah Harris Room; kelima, lama menginap tamu yang teridentifikasi memiliki kuantitas lebih tinggi ialah satu hari; keenam, secara kuantitas bulan menginap tamu yang teridentifikasi ialah pada bulan Desember; ketujuh, berdasarkan tahun pada dataset yang digunakan, dapat diketahui bahwa riwayat pembelian layanan akomodasi hotel tertinggi pada Tahun 2018. Adapun, visualisasi lama menginap berdasarkan bulan dan tahun dapat dilihat pada gambar berikut.



Gambar 6. Data Riwayat Menginap Berdasarkan Bulan dan Tahun

Gambar 6 merupakan visualisasi data riwayat menginap berdasarkan bulan dan tahun pada aplikasi pendukung Customer Relationship Management (CRM). Melalui pengembangan aplikasi tersebut maka dapat dianalisis perilaku pembelian layanan akomodasi hotel berdasarkan season, sehingga dapat disiapkan strategi

pelayanan serta strategi pemasaran yang tepat dengan memberikan promosi harga hingga potongan harga yang sesuai. Selanjutnya, untuk menjalin relasi jangka panjang dengan pelanggan, maka pihak pengelola hotel dapat menanggapi ulasan tamu melalui pesan yang dapat memberikan jaminan (assurance), atau meyakinkan tamu tentang pelbagai upaya penyedia layanan dalam mempertahankan hingga meningkatkan fasilitas dan layanan. Adapun, strategi yang ditetapkan dapat menggunakan data demografi pelanggan hingga data riwayat pembelian layanan akomodasi. Beberapa studi terdahulu menunjukkan bahwa salah satu strategi untuk memberikan kesan yang positif pada seluruh tamu hotel ialah dengan memberikan pengalaman personalisasi selama menginap [42], [43]. Hal ini menunjukkan bahwa aplikasi pendukung Customer Relationship Management (CRM) sangat bermanfaat digunakan dalam mengoptimalkan pelayanan tamu hotel berdasarkan data riwayat pembelian, serta data demografi pelanggan. Dengan demikian, diperlukan fitur dan fungsi aplikasi yang mempermudah akses terhadap database, serta mempermudah proses penelusuran informasi berdasarkan tanggal ulasan pengguna sebagaimana gambar berikut.



Gambar 7. Navigation and Report Page

Gambar 7 merupakan Navigation dan Report Page yang dapat digunakan oleh Guest Relationship Officer penyedia layanan akomodasi Hotel untuk mengakses informasi yang berhubungan dengan ulasan hingga data riwayat pembelian tamu. Dengan demikian, dapat dikelola data ulasan berdasarkan riwayat pembelian dengan strategi komunikasi pemasaran yang tepat sehingga memberikan kesan personalisasi bagi pelanggan, serta memantapkan keputusan menginap kembali. Lestari et al. menunjukkan bahwa fasilitas yang digunakan tamu selama menginap menjadi salah satu faktor yang memengaruhi keputusan menginap kembali [44]. Selain itu, Syafri dan Wulandari berpendapat bahwa performa pelayanan hotel memiliki pengaruh terhadap intensi menginap kembali [45]. Adapun, citra merek dan lokasi hotel juga memiliki pengaruh terhadap keputusan menginap kembali [46]. Hal ini menunjukkan bahwa personalisasi terhadap pengalaman menginap hotel, perlu dioptimalkan dengan pengelolaan strategi komunikasi pemasaran yang tepat. Dengan demikian, maka pendekatan analisis sentimen dan perilaku pembelian layanan akomodasi hotel berdasarkan metode CRISP-DM melalui algoritma Decision Tree, sangat relevan digunakan dalam upaya mengoptimalkan performa bidang Customer Relationship Management (CRM) di hotel.

4. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa metode CRISP-DM melalui algoritma Decision Tree, sangat relevan digunakan dalam upaya mengoptimalkan performa bidang Customer Relationship Management (CRM) di hotel. Secara spesifik, hasil klasifikasi sentimen menunjukkan bahwa terdapat sepuluh kata yang menjadi perhatian tamu hotel serta didominasi sentimen positif, yaitu shopping, great, stay, staff, clean, location, room, good, mall, hotel. Hasil klasifikasi menggunakan algoritma DT menunjukkan performa yang baik dengan nilai akurasi 93,91%, nilai presisi 90,98%, dan nilai recall 97, 77%. Selanjutnya, nilai AUC sebesar 0,943 atau 94,3% serta nilai f_measure 94, 18%. Selanjutnya, data hasil analisis sentimen dapat dikembangkan menjadi aplikasi pendukung Customer Relationship Management (CRM) untuk menganalisis data riwayat pembelian tamu yang berhubungan dengan sentimen, negara asal, tipe tamu, tipe kamar, lama menginap berdasarkan hari, bulan dan tahun. Dengan demikian, strategi pemasaran layanan akomodasi hotel dapat dioptimalkan untuk personalisasi serta meningkatkan minat serta intensi menginap kembali.

REFERENCES

- [1] N. C. Sastya and I. Nugraha, "Penerapan Metode CRISP-DM dalam Menganalisis Data untuk Menentukan Customer Behavior di MeatSolution," J. Pendidik. dan Apl. Ind., vol. 10, no. 2, pp. 103–115, 2023.
- [2] I. Fitrianti, A. Voutama, and Y. Umaidah, "Clustering Film Populer pada Aplikasi Netflix dengan Menggunakan Algoritma K-Means dan Metode CRISP-DM," JTISI, vol. 4, no. 2, pp. 301–311, 2023, [Online]. Available:

- <https://jurnal.mdp.ac.id/index.php/jtsi/article/view/4929%0Ahttps://jurnal.mdp.ac.id/index.php/jtsi/article/download/4929/1546>
- [3] F. Salsabila, I. Fitrianti, Y. Umaidah, and N. Heryana, "Penerapan Metode CRISP-DM untuk Analisa Pendapatan Bersih Bulanan Pekerja Informal di Provinsi Jawa Barat dengan Algoritma K-Means," *DINAMIK*, vol. 28, no. 2, pp. 97–104, 2023.
 - [4] Y. Yudianta, A. Y. Agustina, and N. Khofifah, "Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan," *IJEB Indones. J. Islam. Econ. Bus.*, vol. 8, no. 1, pp. 1–20, 2023, [Online]. Available: <http://e-journal.lp2m.uinjambi.ac.id/ojp/index.php/ijoieb>
 - [5] I. Z. P. Hamdan and M. Othman, "Predicting Customer Loyalty Using Machine Learning for Hotel Industry," *J. Soft Comput. Data Min.*, vol. 3, no. 2, pp. 31–42, 2022.
 - [6] V. Jap, J. Hartanto, and S. Wijaya, "Analisis Kepuasan Dan Ketidakpuasan Tamu Hotel Novotel: Pendekatan Text Mining Atas Ulasan Daring Pada Situs Web Tripadvisor," *J. Manaj. Perhotelan*, vol. 9, no. 2, pp. 71–80, 2023, doi: 10.9744/jmp.9.2.71-80.
 - [7] P. A. D. S. Dieby, "Pengaruh Kualitas Pelayanan, Citra Merek, dan Persepsi Risiko terhadap Loyalitas Melalui Kepuasan Tamu yang Menginap pada Masa Pandemi Covid-19 di Hotel Accor Group Bali," *JIMPS J. Ilm. Mhs. Pendidik. Sej.*, vol. 8, no. 4, pp. 5104–5117, 2023.
 - [8] N. Wulansari, Y. Abrian, and A. Adrian, "Pengaruh Faktor-Faktor Perilaku Konsumen Terhadap Keputusan Menginap," *J. Manaj. Terap. dan Keuang.*, vol. 10, no. 2, pp. 171–182, 2021.
 - [9] R. Setiadi, "Social Media Marketing and Customized Customer Relationship Management: The Implication For Marketing Performance of Budget Hotel Industry," *Indones. J. Bus. Entrep.*, vol. 9, no. 3, pp. 478–490, 2023, doi: 10.17358/ijbe.9.3.478.
 - [10] N. Yulianti and V. Cyasmoro, "Pengaruh Kualitas Pelayanan, CRM dan Citra Merek Terhadap Loyalitas Tamu Menginap di Hotel Horison Ultima Bekasi," *eCo-Buss*, vol. 6, no. 1, pp. 238–252, 2023.
 - [11] R. D. Wastuti and N. H. Wiliam, "Analisis Isi International Luxury Hotel di Bandung Berdasarkan TripAdvisor dan Agoda," *J. Tour. Hosp. Essentials J.*, vol. 8, no. 2, pp. 111–120, 2018, doi: 10.17509/thej.v8i2.13745.
 - [12] Wisyani and V. Okditazeini, "Strategi Komunikasi Pemasaran Hotel Emersia Batusangkar Dalam Meningkatkan Pengunjung Pasca Pandemi Covid-19," *J. Komun. dan Penyiaran*, vol. 3, no. 1, pp. 61–77, 2023, doi: 10.55397/cps.v3i1.38.
 - [13] Valdio and M. P. Berlianto, "Pengaruh Authenticity, Interactivity, Involvement dan Customer Engagement Dalam Memprediksi Customer Loyalty Pada Pemesan Hotel Melalui Traveloka.com," in *NCBMA 2023 (Universitas Peilta Harapan Indonesia)*, 2023, pp. 158–172.
 - [14] Erniyati, P. Harsani, M. Mulyati, and L. D. Fahriza, "Topic Modeling LDA and SVM in Sentiment Analysis of Hotel Reviews," *Komputasi J. Ilm. Ilmu Komput. dan Mat.*, vol. 20, no. 2, pp. 93–100, 2023, doi: 10.33751/komputasi.v20i2.7604.
 - [15] H. Febuariyanti, M. Firmansyah, J. S. Wibowo, and M. S. Utomo, "Analisis Sentimen Tanggapan Terhadap Aplikasi Layanan Informasi Penginapan Menggunakan Metode Klasifikasi Naive Bayes," *semanTIK*, vol. 6, no. 2, pp. 115–124, 2020.
 - [16] Y. A. Singgalen, "Penerapan Metode CRISP-DM untuk Optimalisasi Strategi Pemasaran STP (Segmenting , Targeting , Positioning) Layanan Akomodasi Hotel , Homestay , dan Resort," *J. Media Inform. Budidarma*, vol. 7, no. 4, pp. 1980–1993, 2023, doi: 10.30865/mib.v7i4.6896.
 - [17] Y. A. Singgalen, "Analisis Sentimen dan Sistem Pendukung Keputusan Menginap di Hotel Menggunakan Metode CRISP-DM dan SAW," *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1343–1353, 2023, doi: 10.47065/josh.v4i4.3917.
 - [18] Y. A. Singgalen, "Analisis Sentimen Konsumen terhadap Food , Services , and Value di Restoran dan Rumah Makan Populer Kota Makassar Berdasarkan Rekomendasi TripAdvisor Menggunakan Metode CRISP-DM dan," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 1899–1914, 2023, doi: 10.47065/bits.v4i4.3231.
 - [19] Y. A. Singgalen, "Analisis Sentimen Wisatawan terhadap Kualitas Layanan Hotel dan Resort di Lombok Menggunakan SERVQUAL dan CRISP-DM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 1870–1882, 2023, doi: 10.47065/bits.v4i4.3199.
 - [20] Y. A. Singgalen, "Analisis Perilaku Wisatawan Berdasarkan Data Ulasan di Website TripAdvisor Menggunakan CRISP-DM : Wisata Minat Khusus Pendakian Gunung Rinjani dan Gunung Bromo," *J. Comput. Syst. Informatics*, vol. 4, no. 2, pp. 326–338, 2023, doi: 10.47065/josyc.v4i2.3042.
 - [21] Y. A. Singgalen, "Analisis Sentimen Wisatawan terhadap Taman Nasional Bunaken dan Top 10 Hotel Rekomendasi TripAdvisor Menggunakan Algoritma SVM dan DT berbasis CRISP-DM," *J. Comput. Syst. Informatics*, vol. 4, no. 2, pp. 367–379, 2023, doi: 10.47065/josyc.v4i2.3092.
 - [22] 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.
 - [23] 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.
 - [24] Y. A. Singgalen, "Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT)," *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1551–1562, 2023, doi: 10.30865/mib.v7i3.6461.
 - [25] Y. A. Singgalen, "Perbandingan Metode ARAS dan EDAS dalam Menghasilkan Rekomendasi Layanan Akomodasi Hotel," *J. Comput. Syst. Informatics*, vol. 5, no. 1, pp. 155–164, 2023, doi: 10.47065/josyc.v5i1.4574.
 - [26] Y. A. Singgalen, "Penerapan Metode Additive Ratio Assessment (ARAS) dan Ranking of Centroid (ROC) dalam Pemilihan Layanan Akomodasi dan Local Cuisine," *J. Comput. Syst. Informatics*, vol. 5, no. 1, pp. 51–60, 2023, doi: 10.47065/josyc.v5i1.4569.
 - [27] Y. A. Singgalen, "Penerapan Metode Evaluation based on Distance from Average Solution (EDAS) dalam Optimalisasi Layanan dan Pemasaran Coffeeshop," *J. Comput. Syst. Informatics*, vol. 5, no. 1, pp. 1–9, 2023, doi:

- 10.47065/josyc.v5i1.4460.
- [28] Y. A. Singgalen, “Analisis Perbandingan Top 10 Best-Value dan Top 10 Traveler- Ranked Hotel Menggunakan MOORA,” *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 899–911, 2023, doi: 10.47065/josyc.v4i4.4070.
 - [29] Y. A. Singgalen, “Analisis Sentimen Top 10 Traveler Ranked Hotel di Kota Makassar Menggunakan Algoritma Decision Tree dan Support Vector Machine,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 323–332, 2023, doi: 10.30865/klik.v4i1.1153.
 - [30] Y. A. Singgalen, “Analisis Model Pendukung Keputusan Simple Additive Weighting (SAW) terhadap Top 10 Traveler Ranked Hotel,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 418–426, 2023, doi: 10.30865/klik.v4i1.1159.
 - [31] Y. A. Singgalen, “Analisis Multi-Objective Optimization on the Basis of Ratio Analysis (MOORA) Terhadap Data Top 10 Best Value Hotel,” *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 841–849, 2023, doi: 10.47065/josyc.v4i4.4053.
 - [32] Y. A. Singgalen, “Penerapan Metode TOPSIS Sebagai Pendukung Keputusan Pemilihan Layanan Akomodasi di Destinasi Wisata Pulau,” *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1386–1394, 2023, doi: 10.30865/mib.v7i3.6530.
 - [33] 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.
 - [34] 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.
 - [35] 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.
 - [36] R. Sarudin and A. Ismail, “Tripadvisor.Com Review Online Analysis On The Interest Of Buying Accommodation Services In Hotel Manhattan,” *J. Hosp. dan Pariwisata*, vol. 7, no. 4, pp. 33–43, 2021, [Online]. Available: <https://journal.ubm.ac.id/index.php/hospitality-pariwisata>
 - [37] S. W. Zanra, “Pengaruh Media Instagram Terhadap Minat Pembelian Kamar Pada Hotel Khas Pekanbaru dengan Pemenuhan Informasi Sebagai Variabel Mediasi,” *J. Manaj. dan Bisnis*, vol. 6, no. 1, pp. 1–10, 2023, [Online]. Available: <http://journal.stieindragiri.ac.id/index.php/jmbi%0APENGARUH>
 - [38] N. Hasanah and A. Priatama, “Strategi Pemasaran Terhadap Minat Beli Perhotelan Di Kota Tarakan,” *J. Manaj. dan Bisnis*, vol. 4, no. 3, pp. 409–416, 2022.
 - [39] G. D. Apsarini and E. Ervina, “Keputusan Wisatawan Backpacker dalam Memilih Hostel di Kota Bandung,” *J. Ilm. Univ. Muhammadiyah But.*, vol. 9, no. 4, pp. 864–874, 2023.
 - [40] F. Chairani and N. Wulansari, “Pengaruh Brand Image terhadap Keputusan Menginap di Four Points By Sheraton Batam,” *Student Sci. Creat. J.*, vol. 1, no. 6, pp. 210–2018, 2023.
 - [41] G. L. Silalahi, E. Ervina, and R. Taufiq, “Hubungan Karakteristik Tamu dengan Keputusan Tamu Menginap di Virtual Hotel Operator di Kota Bandung,” *JIIP - J. Ilm. Ilmu Pendidik.*, vol. 6, no. 11, pp. 9167–9172, 2023, doi: 10.54371/jiip.v6i11.2937.
 - [42] G. Yulanda and E. D. H. Putri, “Strategi Guest Relation Officer Dalam Penanganan Tamu Yang Menunggu Pelayanan Di Hotel Santika Premiere Yogyakarta,” *J. Khasanah Ilmu*, vol. 8, no. 1, pp. 20–30, 2017, [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/khasanah/article/view/1413>
 - [43] F. M. Suarka, P. R. Pertiwi, and I. B. D. Setiawan, “Upaya Guest Relation Officer dalam Menangani Keluhan Tamu Peserta MICE di Hotel Kawasan ITDC Nusa Dua Bali,” *J. Kepariwisata dan Hosp.*, vol. 2, no. 2, pp. 144–156, 2018.
 - [44] W. D. Lestari, Ramadania, and E. Listiana, “Pengaruh Pengalaman Pelanggan dan Kondisi Fasilitas Terhadap Kepuasan Pelanggan dan Pengaruhnya Terhadap Niat Menginap Kembali,” *J. Bening*, vol. 10, no. 2, pp. 74–85, 2023.
 - [45] Yolanda Syafri and D. P. Wulandari, “Pengaruh Service Performance terhadap Repurchase Intention di The Axana Hotel Padang,” *J. Manaj. Perhotelan dan Pariwisata*, vol. 6, no. 2, pp. 662–669, 2023, doi: 10.23887/jmpp.v6i2.67104.
 - [46] B. Handrianto, R. A. Vildayanti, Noermijati, S. Aisjah, and A. S. Hussein, “Pengaruh Citra Merek, Persepsi Harga dan Lokasi Terhadap Keputusan Menginap Kembali,” *J. Jimek*, vol. 6, no. 1, pp. 23–35, 2023.