

Analisis Sentimen Review Hotel Menggunakan Metode Naïve Bayes pada Hotel di Wilayah Kota Cirebon

Muhamad Jihad Andiana^{1*}, Martanto², Umi Hayati³

^{1,3} Program Studi Teknik Informatika, STMIK IKMI Cirebon

² Program Studi Manajemen Informatika, STMIK IKMI Cirebon

*Email: andianarosdiana@gmail.com

Abstrak

Cirebon, kota di Jawa Barat, Indonesia, dikenal dengan berbagai daya tarik wisata, termasuk kuliner dan situs bersejarah. Namun, menemukan akomodasi yang tepat bisa menjadi tantangan. Untuk mengatasi masalah ini, sebuah studi telah menganalisis 875 ulasan hotel di Cirebon dari Google Maps, menggunakan metode Naive Bayes dan algoritma TF-IDF. Tujuan dari studi ini adalah untuk membantu wisatawan mendapatkan gambaran yang lebih baik dalam memilih hotel. Hasilnya menunjukkan bahwa algoritma ini berhasil mencapai akurasi 90.52% dalam mengidentifikasi apakah ulasan itu positif atau negatif. Bahkan tanpa penggunaan operator SMOTE, akurasi masih tetap tinggi, yaitu 75.66%. Jadi, studi ini memberikan solusi berbasis data untuk memilih hotel di Cirebon.

Kata Kunci : Kota Cirebon, Google Maps, Hotel, Naive Bayes, TF-IDF.

Abstract

Cirebon, a city in West Java, Indonesia, is known for its various tourist attractions, including culinary and historical sites. However, finding the right accommodation can be a challenge. To address this issue, a study has analyzed 875 hotel reviews in Cirebon from Google Maps, using the Naive Bayes method and the TF-IDF algorithm. The aim of this study is to help tourists get a better picture in choosing a hotel. The results show that this algorithm successfully achieved an accuracy of 90.52% in identifying whether the review was positive or negative. Even without the use of the SMOTE operator, the accuracy remains high, at 75.66%. So, this study provides a data-based solution for choosing a hotel in Cirebon.

Keyword : Cirebon City, Google Maps, Hotel, Naive Bayes, TF-IDF.

PENDAHULUAN

Kota Cirebon adalah salah satu kota yang menjadi daya tarik wisatawan untuk berwisata ke daerah tersebut. Kota Cirebon memiliki banyak sekali objek wisata yang dapat dikunjungi oleh para wisatawan seperti kuliner, tempat wisata, wisata religi, dan oleh – oleh. Sebagai kota yang memiliki banyak objek wisata untuk dikunjungi para wisatawan untuk liburan, tentu memerlukan fasilitas untuk memenuhi kebutuhan liburan, salah satunya dengan penginapan atau hotel. Pariwisata adalah salah satu sektor yang berkontribusi besar terhadap perekonomian suatu negara, dan hotel merupakan bagian integral dari industri ini. Di era digital saat ini, ulasan online menjadi salah satu faktor penting yang mempengaruhi keputusan konsumen dalam memilih hotel. Dalam pemesanan hotel pun saat ini bisa dengan sangat mudah dilakukan, yaitu dengan melakukan pemesanan hotel secara online. Dalam hasil analisis yang dilakukan pada paper (Alhamdi, 2023) yang dilakukan dengan menggunakan regresi linear berganda menghasilkan bahwa ada pengaruh

signifikansi online review terhadap keputusan pemesanan hotel. Dalam paper (Jap, Hartanto and Wijaya, 2023) dengan judul “Analisis Kepuasan Dan Ketidakpuasan Tamu Hotel Novotel : Pendekatan Text Mining Atas Ulasan Daring Pada Situs Web Tripadvisor” menyebutkan bahwa hasil penelitian menemukan bahwa hotel Novotel telah konsisten dalam menjalankan SOP terkait kualitas produk dan layanannya. Atribut service menjadi atribut utama penyebab kepuasan dan ketidakpuasan tamu. Setelah itu penelitian yang dilakukan pada paper (Saksana and Thio, 2019) dalam penilaian penting oleh wisatawan dalam memilih hotel menunjukkan bahwa atribut kamar yang bersih dan kamar mandi yang bersih, serta ketersediaan fasilitas merupakan hal yang paling penting bagi wisatawan Indonesia dan Cina. Selain itu, atribut keseluruhan seperti reputasi, lokasi, dan harga kamar juga dianggap penting, diikuti oleh atribut layanan dan fasilitas, serta atribut makanan dan minuman. Wisatawan Cina juga menganggap lokasi hotel yang aman dan kamar hotel yang

bersih sebagai indikator penting dalam memilih hotel di Bali, sementara wisatawan Indonesia lebih memperhatikan kebersihan kamar hotel dan kamar mandi. Selain itu ada juga hasil penelitian dari paper (Pakpahan, Kusumah and Rosita, 2020) menunjukkan bahwa ada perbedaan dalam kepuasan dan ketidakpuasan tamu tergantung apakah disampaikan secara tertulis atau lisan. Lebih banyak elemen yang dapat menyebabkan kepuasan atau ketidakpuasan tamu ditemukan ketika disampaikan secara lisan. Ketika tamu menyampaikan secara tertulis, mereka cenderung melakukannya secara singkat, tetapi ketika melakukannya secara lisan, mereka memberikan detail yang lebih lengkap tentang pengalaman mereka. Namun, ketika disampaikan secara lisan, tamu lebih sering menyampaikan kepuasan daripada ketidakpuasan. Oleh karena itu, disarankan agar manajemen hotel bintang empat melakukan survei langsung kepada tamu secara berkala untuk mengetahui kepuasan dan ketidakpuasan mereka, sehingga dapat mempertahankan elemen yang baik dan memperbaiki yang kurang.

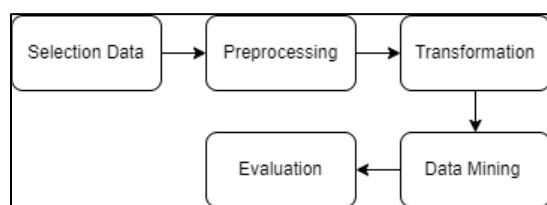
Pada penelitian ini hanya akan fokus untuk mengolah data review sentimen hotel pada aplikasi google maps di wilayah kota cirebon yang di dapatkan melalui website outscrapper.com. Penelitian ini bertujuan untuk memahami bagaimana sentimen dan opini pelanggan tercermin dalam ulasan mereka dan bagaimana informasi ini dapat digunakan oleh industri hotel untuk meningkatkan kualitas layanan mereka. Penelitian ini berpotensi memberikan manfaat signifikan bagi industri perhotelan. Dengan memahami sentimen pelanggan, hotel dapat mengidentifikasi area yang perlu ditingkatkan, sehingga dapat meningkatkan kepuasan pelanggan dan reputasi mereka. Selain itu, hasil penelitian ini juga dapat membantu calon tamu dalam membuat keputusan yang lebih tepat saat memilih hotel. Dengan memahami sentimen umum terhadap suatu hotel, mereka dapat memilih hotel yang paling sesuai dengan kebutuhan dan harapan mereka. Algoritma yang akan digunakan dalam proses mengklasifikasikan teks atau data ke dalam kategori sentimen adalah menggunakan algoritma Naïve Bayes. Dalam jurnal (Syarli and Muin, 2016) menyebutkan

bawa Naïve Bayes adalah salah satu algoritma pembelajaran induktif yang sangat efisien dan efektif dalam machine learning dan data mining. Meskipun menggunakan asumsi keidependan atribut (tidak ada kaitan antar atribut), algoritma ini tetap memberikan performa klasifikasi yang kompetitif. Meskipun jarang terjadi, jika asumsi keidependan atribut dilanggar, performa pengklasifikasian Naïve Bayes tetap tinggi, seperti yang telah dibuktikan dalam berbagai penelitian empiris.

Dalam penelitian ini, peneliti akan menggunakan metode klasifikasi Naïve Bayes untuk menganalisis sentimen dalam ulasan hotel online. Metode penelitian ini melibatkan pengumpulan data dari platform ulasan online, yaitu Google Reviews. Setelah data dikumpulkan, peneliti akan menggunakan teknik Natural Language Processing (NLP) dan Machine Learning untuk menganalisis teks ulasan. NLP akan digunakan untuk memproses teks dan mengidentifikasi kata-kata yang menunjukkan sentimen, sementara Machine Learning akan digunakan untuk mengklasifikasikan ulasan menjadi positif, atau negatif berdasarkan kata-kata tersebut. Selanjutnya, akan menggunakan teknik analisis statistik untuk mengevaluasi hasil dan menarik kesimpulan. Melalui pendekatan ini, peneliti berharap dapat memberikan gambaran yang akurat tentang sentimen pelanggan terhadap hotel dan bagaimana sentimen ini dapat mempengaruhi keputusan pelanggan dalam memilih hotel.

METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini mengacu pada pendekatan yang telah digunakan dalam penelitian sebelumnya yang menyelidiki hal-hal yang serupa, seperti pada Gambar 1.



Gambar 1. Alur Metode Penelitian

Pada penelitian ini menggunakan metode *Knowledge Discovey in Database (KDD)*. Tahapan dalam metode ini meliputi *Selection Data, Preprocessing, Transformation, Data Mining, dan Evaluation*.

2.1 Selection Data

Tahap selection data ini adalah tahap proses pengumpulan data mentah dari *google review*. Data tersebut didapatkan dengan cara *scrapping* dari *google maps review* dengan menggunakan website *outscrapper* dengan mengambil data total 30 hotel yang memiliki review pada rentang waktu awal agustus sampai dengan akhir oktober (3 bulan). Setelah data didapatkan, akan dilakukan *cleaning data* atau pembersihan data seperti menghapus simbol, nomor, *double space* dan *duplicate data*. Setelah dilakukan cleaning data, data yang tadinya memiliki 875 data review pengguna, menjadi 840 komentar.

2.2 Preprocessing

Data yang telah melakukan pembersihan selanjutnya akan melakukan tahap prapemrosesan data. Pra pemrosesan data yang akan dilakukan diantaranya adalah “tokenize”, “transform case”, “filter stopword”, “filter by length”. Semestinya ada metode pembersihan lagi yang bisa digunakan untuk pembersihan data untuk kata yang memiliki imbuhan seperti me, nya, an dll, yang dinamakan stemming. Namun karena data yang didapat lebih banyak mengandung komentar berbahasa indonesia dibandingkan dengan bahasa inggris atau asing, maka pra pemrosesan pada data ini akan menggunakan bahasa Indonesia, karena pada aplikasi rapidminer tidak mendukung stemming dalam berbahasa Indonesia

2.2.1 Tokenize

Tokenize ini berfungsi untuk membagi teks menjadi bagian token atau bagian – bagian tertentu.

2.2.2 Transform Case

Operator transform case ini memiliki fungsi agar teks pada komentar yang memiliki huruf kapital diubah menjadi huruf kecil guna untuk memastikan konsistensi format teks

2.2.3 Filter Stopwords

Operator ini berfungsi agar menghapus teks komentar yang memiliki kata yang tidak berarti.

2.2.4 Filter Tokens(By Length)

Filter token by length adalah proses memfilter token atau kata dalam sebuah teks berdasarkan panjang sesuai dengan yang diinginkan atau butuhkan.

2.3 Transformation Data

Tahap Transformasi adalah saat data disesuaikan untuk digunakan dalam data mining. Data yang telah dibersihkan akan diubah dan disesuaikan dengan metode atau algoritma yang akan digunakan. Pada tahap ini, setiap komentar berbahasa Inggris atau asing akan diubah atau di translate menjadi bahasa Indonesia, selanjutnya data akan diberi label, dan akhirnya setiap kata – kata dalam komentar akan diberi bobot menggunakan metode TF-IDF sehingga data akan berubah dari teks menjadi vektor bobot.

2.4 Data Mining

Setelah selesai melakukan transformation data yang memiliki 3 tahapan diatas, langkah selanjutnya adalah melakukan tahap mengklasifikasi kan data dengan menggunakan metode algoritma naive bayes. Algoritma Naïve Bayes adalah salah satu teknik klasifikasi yang menggunakan model probabilistik dan statistik. Algoritma ini didasarkan pada teorema Bayes dan mengasumsikan bahwa setiap atribut adalah independen satu sama lain (Fikri, Sabrina and Azhar, 2020).

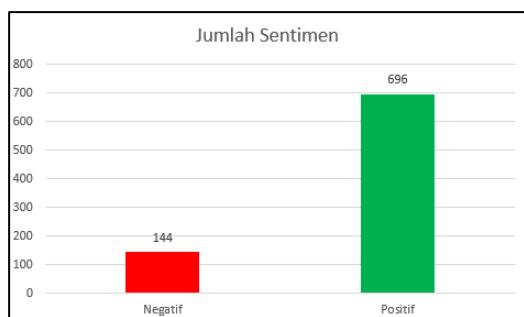
2.5 Evaluation

Tujuan dari fase ini adalah menghitung keakuratan algoritma untuk menganalisis opini menggunakan algoritma Naive Bayes. Perhitungan yang dilakukan pada tahap ini dimaksudkan untuk memastikan keakuratan hasil setiap tahap yang dilakukan dalam penelitian. Selanjutnya hasil algoritma Naive Bayes dievaluasi. Evaluasi hasil klasifikasi yang diperoleh menggunakan matriks konfusi dan laporan klasifikasi berupa presisi, presisi, recall, dan skor F1.

HASIL DAN PEMBAHASAN

Data review yang didapatkan dari google maps untuk penelitian ini adalah data review tentang hotel yang berada di sekitar kota Cirebon. Hasil dari penelitian disini akan menjelaskan secara rinci langkah – langkah yang digunakan untuk penelitian ini. Dimulai

dari proses pengambilan data, preprocessing data, transformation data, data mining, evaluasi, dan hasil dari penelitian ini. Selain itu, untuk mempermudah proses pemahaman pembaca peneliti akan membuat visualisasi seperti table, diagram, cloudword, ataupun yang lainnya.



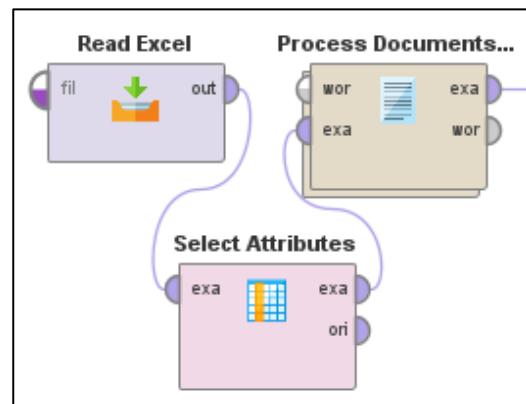
Gambar 2. Column Chart Sentimen

Data hasil scrapping selanjutnya akan dilakukan *selection data* untuk memilih atribut data yang akan dinilai penting dalam analisis seperti review atau komentar dan rating untuk menentukan nilai sentimen dari review. Dari penelitian sebelumnya yang dilakukan dalam paper (Fathullah, Sari and Adikara, 2020) menyebutkan bahwa rating dapat dijadikan acuan untuk menentukan label positif ataupun negatif pada sebuah komentar atau ulasan.

Dalam melakukan pengujian, akan digunakan 840 data komentar yang telah dilakukan *cleaning* dan *preprocessing* pada tahap sebelumnya. Karena data komentar negatif dan positif memiliki perbedaan yang jauh seperti yang tersaji dalam Gambar 2, maka dalam penelitian ini akan digunakan metode *SMOTE upsampling* untuk menyeimbangkan data komentar negatif dan positif. Dengan menggunakan metode SMOTE, sampel yang baru akan disintesis dari kelas minoritas yang kemudian akan menyeimbangkan dataset, dengan cara membangun sebuah instance baru dari kelas minoritas dengan tujuan meningkatkan performance dari metode klasifikasi (Sulistiyono *et al.*, 2021).

Langkah pertama adalah menambahkan operator untuk mengimport data yang telah dilakukan *cleaning* karena process preprocessing akan kita gunakan kembali

pada proses ini. Sebelum menambahkan *process* untuk *preprocessing*, tambahkan operator *select* atribut untuk memilih atribut yang hanya akan digunakan untuk penelitian. Pilih atribut review dan sentimen untuk dilakukan analisis.



Gambar 3. Process Preprocessing

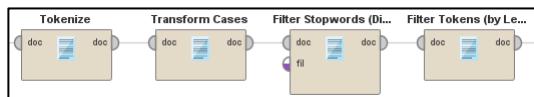
Pada parameters *read excel* lakukan penyesuaian untuk mengatur *role* dan *type* pada kolom *dataset metadata information* dari atribut yang akan digunakan untuk penelitian.

Review	<input checked="" type="checkbox"/> column ...	text	attribute
Nama_Hotel	<input checked="" type="checkbox"/> column ...	polynomi...	attribute
Nama	<input checked="" type="checkbox"/> column ...	polynomi...	attribute
SENTIMEN	<input checked="" type="checkbox"/> column ...	polynomi...	label
Rating	<input checked="" type="checkbox"/> column ...	integer	attribute

Gambar 4. Penyesuaian Parameter Read Excel

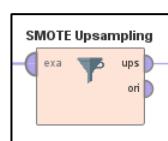
Selanjutnya adalah tahap *preprocessing* dengan menggunakan operator *process documents from data*. Dalam operator *preprocessing* data memiliki bagian dalam untuk melakukan *preprocessing* data yaitu, *tokenize* yang ini berfungsi untuk membagi teks menjadi bagian token atau bagian – bagian tertentu, *transform case* berfungsi agar teks pada komentar yang memiliki huruf kapital diubah menjadi huruf kecil guna untuk memastikan konsistensi format teks, *filter stopword* berfungsi agar menghapus teks komentar yang memiliki kata yang tidak berarti, dan *filter tokens by length* adalah

proses memfilter token atau kata dalam sebuah teks berdasarkan panjang sesuai dengan yang diinginkan atau butuhkan.



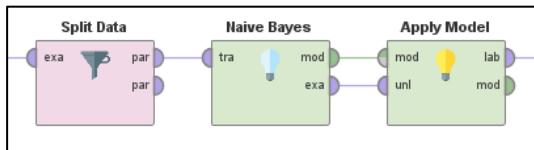
Gambar 5. Tahap Preprocessing Data

Setelah data dilakukan preprocessing, data akan diolah dengan menggunakan metode SMOTE dan juga tanpa metode SMOTE untuk menyeimbangkan data komentar negatif dan positif.



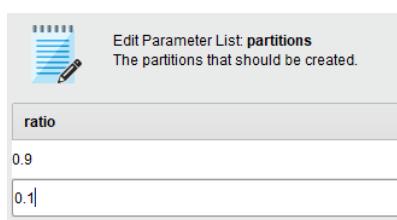
Gambar 6. Operator SMOTE

Setelah data dilakukan penyeimbangan antara komentar negatif dan juga komentar positif, selanjutnya dilakukan adalah menambahkan operator *split data* yang berguna untuk membagi data menjadi data latih dan data uji.



Gambar 7. Process Klasifikasi Data

Dalam operator split data tambahkan dua data baru. Kolom pertama adalah untuk data latih sementara kolom kedua adalah data uji. Split data yang dilakukan untuk pengujian akan menggunakan split data 90 :10, maka dapat dituliskan untuk operator *split data* seperti berikut.



Gambar 8. Process Split Data

Selanjutnya adalah penerapan model naive bayes operator *apply model* dan untuk menghitung tingkat keakurasiannya model *naive bayes* tambahkan operator *performance*. Berikut adalah hasil perbandingan dari penelitian naive bayes dengan menggunakan metode SMOTE dan tanpa menggunakan SMOTE.

Tabel 1. Perbandingan Naive Bayes + SMOTE dan Naive Bayes Tanpa SMOTE

	Naïve Bayes + SMOTE	Naïve Bayes Tanpa SMOTE
Accuracy	87.54%	75.66%
Precision	80.05%	41.40%
Recall	100.00%	100.00%
AUC	0.751	0.708

Hasil pada Tabel 1 membuktikan bahwa hasil klasifikasi sentimen dengan menggunakan metode SMOTE dan tanpa SMOTE memiliki perbedaan yang cukup signifikan. Dimana model naive bayes dengan menggunakan metode SMOTE dapat menghasilkan tingkat *accuracy*, *precision*, dan *AUC* yang lebih tinggi mencapai 87.54% pada *accuracy*, 80.05% pada *precision*, dan pada *AUC* 0.751 dibandingkan dengan tanpa tambahan metode SMOTE yang hanya mendapatkan *accuracy* 75.66%, *precision* 41.40%, dan *AUC* 0.708.

Pada penelitian ini, akan dilakukan 5 model split data yang akan dilakukan uji coba untuk mendapatkan model split data yang mendapatkan akurasi yang paling besar. Split data yang akan dilakukan dibagi menjadi model seperti pada Tabel 2.

Tabel 2. Pembagian Model Split Data

Model	Data Latih	Data Uji
Model 1	90%	10%
Model 2	80%	20%
Model 3	70%	30%
Model 4	60%	40%
Model 5	50%	50%

Setelah dilakukan pengujian pada kelima model, model 5 mendapatkan hasil yang paling tinggi dengan akurasi mencapai 90.51% dengan menggunakan metode SMOTE dan model lainnya memiliki nilai

akurasi dibawah model kelima yang dapat dilihat pada Tabel 3.

Tabel 3. Hasil Model Split Data

Model	Hasil Akurasi
Model 1	87.54%
Model 2	87.79%
Model 3	88.91%
Model 4	89.95%
Model 5	90.52%

Oleh karena itu dengan hasil akurasi pada 5 model dapat disimpulkan bahwa perhitungan model 5 memiliki akurasi yang baik dalam mengklasifikasi analisis terkait dengan komentar review pada hotel di kota Cirebon.

SIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa algoritma naive bayes dapat melakukan klasifikasi data pada *google maps review* dengan nilai akurasi mencapai 90.52% dengan menggunakan tambahan metode SMOTE. Namun dalam penelitian ini belum mencari tahu tentang keefektivitasan metode lain seperti random forest, K-NN, SVM, dan metode lainnya dalam mengklasifikasikan data sentimen hotel di wilayah cirebon.

Dengan total sentimen 875 komentar yang diambil dari rentang waktu awal agustus sampai dengan akhir oktober (3 bulan) pada aplikasi google maps dengan pencarian “hotel bintang 3 di Cirebon”, sentimen yang mendominasi adalah sentimen positif dengan total sentimen positif mencapai 729 komentar, sedangkan komentar negatif hanya mendapat 146 komentar. Maka dapat disimpulkan bawah hotel di daerah wilayah kota Cirebon ini layak untuk ditempati bagi para pengunjung atau tamu. Dan hasil ini bisa menjadi acuan manager hotel untuk mempromosikan hotelnya.

Untuk penelitian selanjutnya dapat mencoba metode lain seperti *random forest*, K-NN, SVM, dan metode lainnya agar dapat membandingkanya dengan metode *Naive Bayes* ini. Dan gunakan preprocessing data *stemming* agar data yang akan digunakan untuk klasifikasi lebih siap untuk digunakan.

DAFTAR PUSTAKA

- Alhamdi, R. (2023) ‘Pengaruh Online Review Dan Harga Terhadap Keputusan Pemesanan Kamar Hotel Di Online Travel Agent (Studi Kasus Kota Batam)’, *Jurnal Manajemen Perhotelan*, 9(2), pp. 63–70. Available at: <https://doi.org/10.9744/jmp.9.2.63-70>.
- Fathullah, N.S., Sari, Y.A. and Adikara, P.P. (2020) ‘Analisis Sentimen Terhadap Rating dan Ulasan Film dengan menggunakan Metode Klasifikasi Naïve Bayes dengan Fitur Lexicon-Based’, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4(2), pp. 590–593. Available at: <https://j-ptik.ub.ac.id/index.php/j-ptik/article/view/6987>.
- Fikri, M.I., Sabrina, T.S. and Azhar, Y. (2020) ‘Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter’, *Smatika Jurnal*, 10(02), pp. 71–76. Available at: <https://doi.org/10.32664/smatika.v10i02.455>.
- Jap, V., Hartanto, J. and Wijaya, S. (2023) ‘Analisis Kepuasan Dan Ketidakpuasan Tamu Hotel Novotel: Pendekatan Text Mining Atas Ulasan Daring Pada Situs Web Tripadvisor’, *Jurnal Manajemen Perhotelan*, 9(2), pp. 71–80. Available at: <https://doi.org/10.9744/jmp.9.2.71-80>.
- Pakpahan, D.T.U., Kusumah, A.H.. and Rosita, R. (2020) ‘Komparasi Persepsi Kepuasan Tamu Secara Lisan Dan Tertulis Ketika Menginap Di Hotel Bintang Empat’, *Jurnal Manajemen Perhotelan*, 5(2), pp. 94–101. Available at: <https://doi.org/10.9744/jmp.5.2.94-101>.
- Saksana, A.C. and Thio, S. (2019) ‘Atribut-Atribut Yang Dinilai Penting Oleh Wisatawan Domestik Dan Wisatawan Cina Dalam Memilih Hotel Di Bali’, *Jurnal Manajemen Perhotelan*, 5(1), pp. 9–17. Available at:

<https://doi.org/10.9744/jmp.5.1.9-17>.

Sulistiyono, M. *et al.* (2021) ‘Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi’, *Sistemasi*, 10(2), p. 445. Available at: <https://doi.org/10.32520/stmsi.v10i2.1303>.

Syarli and Muin, A.A. (2016) ‘Metode Naive Bayes Untuk Prediksi Kelulusan’, *Jurnal Ilmiah Ilmu Komputer*, 2(1), pp. 22–26. Available at: <https://media.neliti.com/media/publications/283828-metode-naive-bayes-untuk-prediksi-kelulu-139fcfea.pdf>.