

Analisis Sentimen Ulasan Bali Hai Cruises Menggunakan Algoritma Support Vector Machine

(Sentiment Analysis of Bali Hai Cruises Reviews Using Support Vector Machine Algorithm)

Ekka Pujo Ariesanto Akhmad, Didik Purwiyanto
Program Studi Manajemen Pelabuhan dan Logistik Maritim,
Fakultas Vokasi Pelayaran, Universitas Hang Tuah

Abstrak: Tulisan ini menyajikan analisis sentimen dalam ulasan *online* dengan menggunakan algoritma *Support Vector Machine (SVM)*. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan online menjadi positif atau negatif dengan memanfaatkan algoritma *SVM*. Dataset yang digunakan dalam penelitian ini terdiri dari review *online* mulai bulan Mei 2020 hingga bulan April 2024 dari www.tripadvisor.com. Hasilnya menunjukkan efektivitas algoritma *SVM* dalam mengkategorikan sentimen dalam ulasan *online* secara akurat, dengan tingkat presisi dan ingatan yang tinggi. Temuan penelitian ini mempunyai implikasi penting untuk memahami dan menganalisis sentimen dalam konten *online*, dan dapat dimanfaatkan dalam berbagai aplikasi seperti riset pasar, analisis umpan balik pelanggan, dan pengumpulan opini.

Kata kunci: analisis sentimen, ulasan *online*, bali hai cruises, support vector machine

Abstract: This paper presents sentiment analysis in online reviews using Support Vector Machine (SVM) algorithm. This research aims to classify the sentiment of online reviews into positive or negative by utilizing the SVM algorithm. The dataset used in this research consists of online reviews from May 2020 to April 2024 from www.tripadvisor.com. The results show the effectiveness of the SVM algorithm in accurately categorizing sentiment in online reviews, with a high level of precision and recall. The findings of this study have important implications for understanding and analyzing sentiment in online content, and can be utilized in various applications such as market research, customer feedback analysis, and opinion gathering.

Keywords: sentiment analysis, online reviews, bali hai cruises, support vector machine

Alamat Korespondensi:

Ekka Pujo Ariesanto Akhmad, Program Studi Manajemen Pelabuhan dan Logistik Maritim, Fakultas Vokasi Pelayaran, Universitas Hang Tuah, Jalan A. R. Hakim 150, Surabaya. e-mail: jurnal.pdp@hangtuah.ac.id

PENDAHULUAN

Salah satu perusahaan pariwisata bahari di Indonesia, yaitu Bali Hai Cruises, melakukan pelayaran sehari-hari ke Pulau Lembongan, yang menawarkan banyak aktivitas menyenangkan untuk seluruh wisatawan. Selain itu, paket makan malam di kapal pesiar ditawarkan oleh perusahaan termasuk penginapan di Hai Tide Beach Resort di tepi pantai. (balihaicruises.com).

Bagi banyak peneliti, penting untuk mendapatkan pemahaman yang lebih mendalam tentang analisis sentimen wisatawan. Penelitian ini memiliki tujuan

untuk mengklasifikasi komentar secara akurat dalam bisnis pariwisata bahari, dimana pembaruan layanan saat ini menjadi tanggung jawab yang paling krusial dan penting. Banyak orang yang tidak dapat memberikan masukan secara langsung, sehingga mereka mengekspresikan pemikiran mereka melalui kata-kata singkat, emoji, atau simbol-simbol unik. Saat ini, situs web perjalanan adalah komunitas online yang berkembang pesat dimana pengguna dapat memposting komentar singkat atau penilaian untuk perusahaan pariwisata bahari berdasarkan pengalaman mereka.

Pengguna lain dapat membaca dengan teliti ulasan tersebut dan membuat keputusan berdasarkan informasi tentang perusahaan pariwisata bahari mana yang akan dipilih tanpa harus berinteraksi secara fisik dengan perusahaan tersebut (Yadu dan Shukla, 2022).

Penelitian Terdahulu

Penelitian analisis sentimen menggunakan *SVM* dapat ditemukan dalam berbagai penelitian berikut.

Sari, dkk. (2023) membandingkan *SVM* dengan *K-Nearest Neighbors* untuk analisis sentimen tempat wisata populer di Indonesia, menemukan *SVM* lebih unggul.

Sebuah studi tentang komentar film dari *IMDB* menggunakan *SVM* menunjukkan akurasi 79%, dengan *SVM* mengungguli regresi logistik (Ramadhan, N. G. dan Ramadhan, T. I., 2022).

Lu dan Wu (2019) mengusulkan metode analisis sentimen berdasarkan *SVM* untuk teks ulasan film, mencapai akurasi yang lebih tinggi daripada metode kamus sentimen dasar.

Laoh, dkk. (2019) menggunakan *SVM* untuk klasifikasi sentimen dalam ulasan pariwisata, menunjukkan tingkat akurasi 94%, melampaui metode sebelumnya.

Selain itu, Vinodhini dan Chandrasekaran (2014) memperkenalkan model *SVM-PCA* hibrida untuk penambangan sentimen, mengungguli *SVM* tradisional dalam akurasi klasifikasi.

Karya penelitian yang beragam ini menunjukkan efektivitas *SVM* dalam analisis sentimen di berbagai domain.

Algoritma dan pendekatan untuk analisis sentimen

Analisis sentimen menggunakan berbagai algoritma dan pendekatan untuk

mengekstraksi sentimen dari data tekstual. Model *Machine Learning* klasik seperti *Naive Bayes*, *Decision Tree*, dan *SVM* telah menunjukkan kinerja yang kuat dalam kategorisasi sentimen (Alemayehu, dkk., 2023; Inder, 2022).

Model *Deep Learning*, termasuk *CNN*, *LSTM*, *GRU*, dan kombinasinya, menawarkan kemampuan ekstraksi fitur otomatis, memungkinkan pemrosesan yang efisien dari kumpulan data besar (Malavade dan Sureshkumar, 2023; Hamnah, 2023; Vivek, dkk., 2023).

Selain itu, pendekatan berbasis aturan dan otomatis umumnya digunakan, dengan *Logistic Regression*, *SVM*, *Random Forest*, dan *Naive Bayes* menjadi pilihan populer. Selanjutnya, analisis sentimen memanfaatkan algoritma yang terinspirasi alam, *Machine Learning*, teknik *Deep Learning*, dan Grafik Pengetahuan untuk ekstraksi sentimen yang akurat dari beragam sumber seperti teks, video, dan konten media sosial. Algoritma dan pendekatan yang beragam ini secara kolektif berkontribusi pada kemajuan dan efektivitas sistem analisis sentimen.

Term Frequency - Inverse Document Frequency (TF-IDF)

Pada tahap pembobotan kata, metode *Term Frequency - Inverse Document Frequency (TF-IDF)* digunakan untuk mendapatkan nilai bobot setiap kata pada data yang digunakan. Proses pembobotan kata menggunakan algoritma *TF-IDF*. *TF-IDF* menyajikan nilai frekuensi kata terutama untuk kata-kata yang menarik, misalnya kata-kata yang sering muncul dalam satu dokumen tetapi tidak untuk semua dokumen. Proses ini dilakukan dengan menghitung bobot setiap kata pada data latih dengan menggunakan *TF-IDF* dari sebuah term dihitung dengan mengalikan nilai *TF* dan *IDF* (Alzamzami, dkk., 2020).

Synthetic Minority Oversampling Technique (SMOTE)

Penelitian ini menggunakan *SMOTE*, yaitu metode yang banyak digunakan untuk mengatasi masalah klasifikasi data yang tidak seimbang (Chaoyou, dkk., 2019; Joseph, dkk., 2022; Utkarsh, dkk., 2022; Chawla, dkk., 2002). Ini melibatkan pembuatan sampel sintetis untuk kelas minoritas untuk menyeimbangkan dataset.

K-Fold Cross-Validation

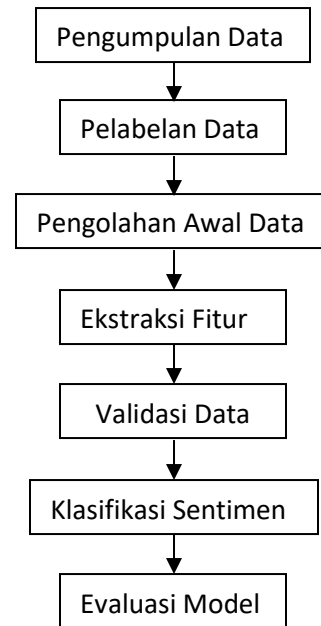
Metode *k-fold cross validation* digunakan dalam penelitian ini untuk menilai performa dan generalisasi model machine learning. Metode ini melibatkan pemisahan dataset menjadi 'k' jumlah subset atau *fold* yang berukuran sama, dan kemudian secara berulang melatih model sebanyak 'k' kali. Setiap kali, satu *fold* digunakan sebagai set validasi (atau tes), sementara *k-1 fold* yang tersisa digabungkan dan digunakan sebagai set pelatihan. (Nguyen, T.T. dan Nguyen, P.C., 2023; Saraswati dan Disha, 2022).

Setelah menjalankan proses untuk semua iterasi 'k', hasilnya dirata-rata untuk mendapatkan ukuran yang komprehensif dari kinerja prediktif model. Teknik ini membantu mengurangi variabilitas dan memberikan penilaian yang lebih akurat dibandingkan dengan menggunakan satu set pelatihan dan pengujian (Linnenbrink, dkk., 2023), sehingga berguna untuk menyempurnakan model dan memverifikasi bahwa model tersebut akan berkinerja baik.

METODE PENELITIAN

Gambar 1 menjelaskan pengumpulan data analisis sentimen diawali dengan melakukan *scraping* atau ekstraksi data dari halaman web Tripadvisor.

Luaran dari *scraping* lalu dipilih kolom yang dibutuhkan dan diperiksa apakah mengandung missing values.



Gambar 1. Alur Penelitian

Kemudian, dataset diberi label *positive* atau *negative* sesuai rating yang diberikan wisatawan.

Berikutnya dataset masuk ke dalam tahap pengolahan awal data. Pengolahan awal data adalah tahap untuk mempersiapkan data yang awalnya tidak terstruktur menjadi data yang lebih terstruktur agar dapat diproses untuk langkah berikutnya. Beberapa tahapan pada pengolahan awal data, yaitu tokenisasi, *data cleaning*, *case folding*, dan *stopwords*.

Setelah melewati proses tersebut, kemudian ekstraksi fitur (pembobotan kata) dikerjakan dengan metode *TF-IDF*.

Sebelum melakukan validasi data, penelitian ini juga menerapkan *SMOTE* untuk mengatasi data yang tidak seimbang dari hasil pelabelan data.

Setelah itu, validasi data dilakukan dengan *k-fold cross validation*.

Klasifikasi data menggunakan empat algoritma *machine learning*, yaitu *SVM*, *Logistic Regression*, *Naive Bayes*, (Yifei, 2023), dan *Random Forest* (Breiman, 2001).

Kinerja model klasifikasi dievaluasi menggunakan *confusion matrix* seperti akurasi, *precision*, dan *recall* (Akhmad, dkk., 2023). Tabel 1 menunjukkan struktur *confusion matrix*.

Tabel 1. Confusion Matrix

Predicted	Actual	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

TP, True Positive; FN, False Negative; FP, False Positive; TN, True Negative

Penelitian ini menggunakan akurasi untuk mengukur seberapa akurat model dapat mengklasifikasikan data dengan benar.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Precision menggambarkan tingkat keakuratan antara data prediksi benar positif yang diminta dengan hasil prediksi yang diberikan oleh model.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

HASIL DAN PEMBAHASAN

Pengumpulan Data

Pengumpulan data ulasan Bali Hai Cruises bulan Mei 2020 sampai dengan bulan April 2024 menggunakan *scraping*. *Software Octoparse* digunakan untuk *web*

scraping dari situs Tripadvisor (www.tripadvisor.com).

Kemudian, penyeleksian fitur dikerjakan pada dataset. Seleksi fitur dikerjakan dengan menghapus kolom Title, Title_URL, Image, bigqs_URL, bigqs, bigqs1, bigqs2, rpecd, Image3, Image4, bmqdv_URL, bmqdv, bigqs5, dan bigqs6, karena dianggap tidak dibutuhkan untuk analisis sentimen. Lalu, dataset disimpan dalam format *excel file* (.xlsx). Luaran *scraping* dapat dilihat pada Gambar 2.

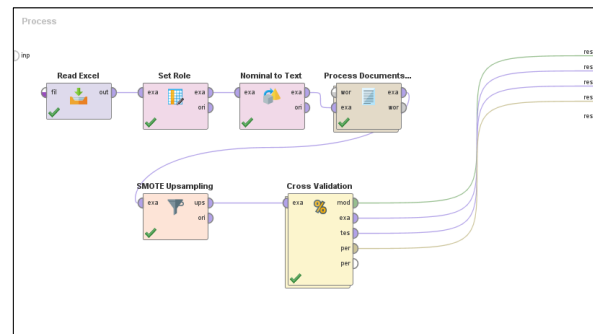
text	label
The porton dock offered excellent activities, the Scuba diving was exceptional and all staff were incredibly friendly. The beaches were gorgeous, and as a 1 vegan and 1 dairy-free dietary requirements, they went above and beyond to cater to us. The food was delicious! Honestly couldn't recommend more. The only thing that could have been better was a longer time on the island and more of a tour.	positive
Amazing staff, great venue, lovely beaches with beautiful clear water so perfect for water activities. Would recommend	positive
100%	positive
Great food, awesome performance. Would go again	positive
Dancing was great, plenty of food 100% would recommend	positive
Got up and danced with the performer	positive
Very nice tour, good food good entertainment. Time flies so fast in this cruise. Staff abd service also very nice.	positive
Organiser Arya is very attentive to all passengers of cruise.	positive
This is a very romantic experience, we had dinner on the boat while watching the sun go down around Nusa Dua beach accompanied by amazing music... we are very satisfied to be able to cruise in Bali hi thank you for your service, for further information you can contact us on whatsapp (+628135338773)	positive
The staff and the experience was amazing thank you so much Bali Hai- Reef Cruise I'll definitely be recommending you	positive
7 guys to friends and family visiting Bali	positive
I had a great good going snorkelling and riding the banana boat, the food was amazing and everyone was extremely nice. I loved it and would enjoy on the boat	positive

Gambar 2. Luaran Scraping

Pelabelan Data

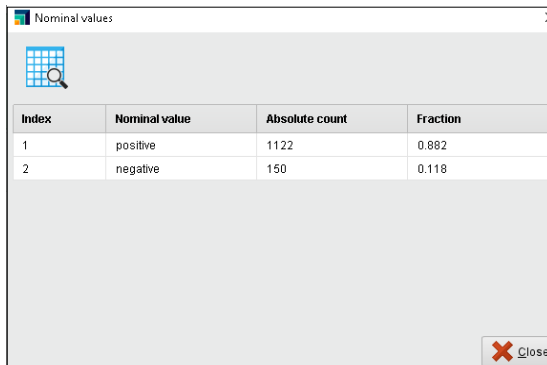
Karena, analisis sentimen menggunakan metode *supervised learning* dibutuhkan dataset yang telah mempunyai label. Labelisasi dikerjakan penulis dengan membubuhkan label *negative* pada data rating 1 dan 2, serta label *positive* pada data rating 3, 4, dan 5.

Gambar 3 menjelaskan *design process* RapidMiner untuk analisis sentimen ulasan Bali Hai Cruises.



Gambar 3. Design Process

Perhitungan jumlah dan pecahan labelisasi data menggunakan *software RapidMiner*. Hasil perbandingan labelisasi data memiliki angka 0,882 untuk sentimen *positive* dan 0,118 untuk sentimen *negative*, atau 1.122 ulasan sentimen *positive* dan 150 ulasan sentimen *negative* dapat dilihat pada Gambar 4.

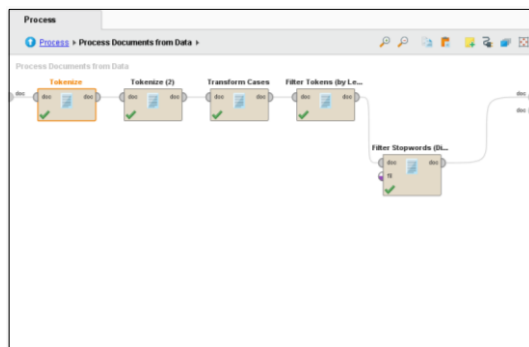


Index	Nominal value	Absolute count	Fraction
1	positive	1122	0.882
2	negative	150	0.118

Gambar 4. Labelisasi Data

Pengolahan Awal Data

Pengolahan awal data dilakukan untuk mengubah dataset yang tidak terstruktur menjadi terstruktur, sehingga memudahkan data untuk diproses dengan melakukan beberapa tahapan, yaitu tokenisasi, *data cleaning*, *case folding*, *filter tokens (by length)*, dan *stopwords*. Tahap pengolahan awal data memproses dokumen dapat dilihat pada Gambar 5.



Gambar 5. Process Document

Tokenisasi

Tokenisasi memakai *regular expression* untuk menemukan karakter yang akan dihapus. Tokenisasi digunakan untuk memecah kalimat menjadi *list* kata.

Data Cleaning

Data Cleaning dipakai untuk menghilangkan angka, beberapa simbol, url, *username* (@username), *hashtag* (#), spasi berlebih, tanda baca, emoji, dan pengulangan karakter yang ada pada kalimat. *Data cleaning* menggunakan *regular expression* untuk menemukan karakter yang akan dihapus.

Case Folding

Case folding digunakan untuk mengubah semua karakter menjadi huruf kecil. Tahapan ini dilakukan karena data yang diperoleh tidak selalu terstruktur dan konsisten dalam penggunaan huruf kapital, maka *case folding* dilakukan untuk menyamaratakan penggunaan huruf kapital.

Filter tokens (by Length)

Filter tokens (by length) digunakan untuk menghilangkan kata-kata dalam teks yang memiliki panjang kurang dari 4 karakter dan panjang lebih dari 712 karakter (*default*-nya 25).

Stopwords

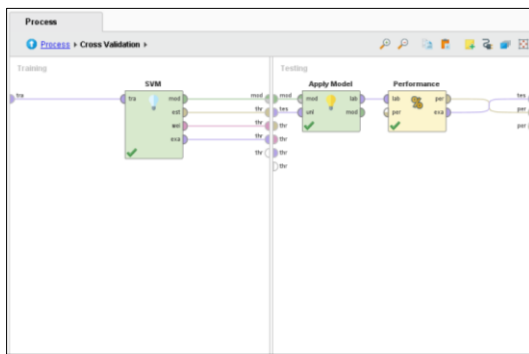
Kata-kata berhenti (*stopwords*) adalah kata-kata dalam *list stop* yang disaring sebelum atau sesudah pemrosesan data bahasa alami (*natural language*), karena dianggap tidak penting. Penelitian ini memakai *stopwords* bahasa Inggris.

Ekstraksi Fitur

Metode TF-IDF digunakan pada penelitian ini untuk memilih fitur sebagai hasil ringkasan, dengan penerapannya pada seleksi fitur bobot kata.

Validasi Data

10-fold cross validation digunakan dalam penelitian ini. Validasi membagi dataset menjadi 10 bagian yang sama, melatih model pada 9 bagian, dan pengujian pada bagian yang tersisa secara berulang. Proses ini membantu dalam menilai kinerja model dengan mengurangi *bias* dan *varians*. Dengan merata-ratakan hasil evaluasi selama 10 iterasi, estimasi yang lebih andal dari metrik kinerja model seperti akurasi, presisi, dan ingatan dapat diperoleh, meningkatkan generalisasi dan ketahanan model di berbagai kumpulan data dan aplikasi. Gambar 6 menjelaskan *design process 10-fold cross validation* dengan algoritma SVM.

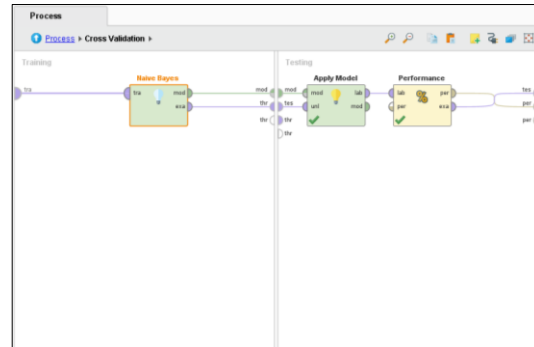


Gambar 6. Cross Validation

Klasifikasi Sentimen

Klasifikasi sentimen melibatkan kategorisasi ulasan ke dalam kelas atau label sentimen positif atau negatif. Beberapa algoritma, yaitu *Logistic Regression (LR)*, *Naive Bayes (NB)*, *Random Forest (RF)*, dan *Support Vector Machine (SVM)* digunakan untuk tujuan ini. Algoritma ini memainkan peran penting dalam pembelajaran yang diawasi, dimana model dilatih pada data berlabel untuk membuat prediksi pada contoh yang tidak terlihat. Pemilihan algoritma yang paling sesuai tergantung pada faktor-faktor seperti presisi, akurasi, waktu konstruksi model, dan ukuran

dataset. Gambar 7 menunjukkan *design process* klasifikasi sentimen dengan algoritma NB. Hal yang sama diterapkan untuk algoritma LR, RF, atau SVM.



Gambar 7. Design Process klasifikasi sentimen

Selain klasifikasi sentimen, Gambar 8 menjelaskan design process untuk wordlist to data. Operator "Wordlist to Data" digunakan untuk mengonversi daftar kata menjadi format data tabular (contohnya, Gambar 9) yang bisa diolah oleh operator-operator lain di RapidMiner.



Gambar 8. Design Process WordList to Data

Urutan daftar kata terbanyak yang muncul pada baris tertentu dari ulasan sentimen dapat dilihat pada Gambar 9.

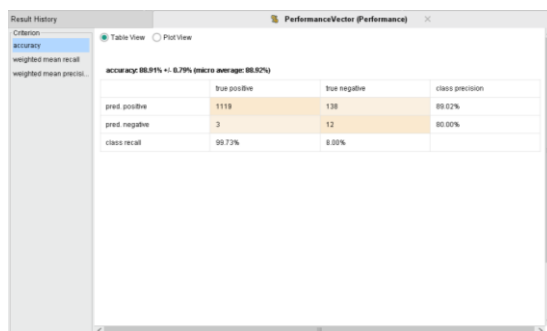
Row No.	word	in documents	total	in class (positive)	in class (negative)
1113	course	589	966	784	202
495	local	415	763	607	156
351	ball	429	614	508	106
1825	food	569	646	542	104
334	back	231	354	253	101
2407	just	183	275	185	90
4584	time	356	510	427	83
2876	money	131	162	84	78
3049	only	155	192	118	74
4706	trip	264	425	351	74
4406	itself	195	247	175	72
4834	hour	217	352	280	72
395	beach	212	395	325	70
1312	album	199	269	202	67
4259	staff	449	523	461	62

Gambar 9. Daftar urutan kata terbanyak

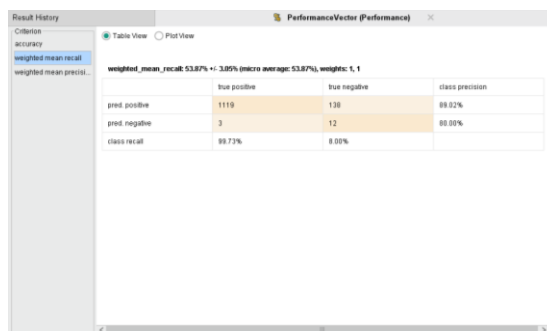
Gambar 9 menerangkan analisis umpan balik pelanggan yang berasal dari daftar kata terbanyak yang timbul dari ulasan sentimen negatif. Manajemen *bali hai cruises* bisa meningkatkan pelayanan jasa di masa depan dengan memperhatikan lima kata-kata terbanyak yang ditulis oleh wisatawan, yakni *cruise*, *boat*, *bali*, *food*, dan *back*.

Evaluasi Model

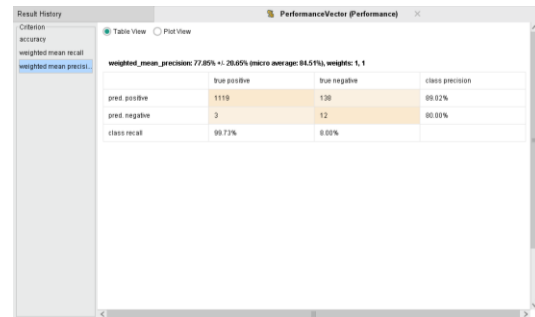
Penelitian ini mengerjakan evaluasi model dalam dua skenario. Skenario pertama mengevaluasi kinerja model klasifikasi sentimen tanpa SMOTE. Gambar 10 menerangkan kinerja (*performance*) algoritma SVM tanpa SMOTE yang terdiri dari, *accuracy*, *recall*, dan *precision*.



Gambar 10 a. Accuracy – SVM tanpa SMOTE



Gambar 10 b. Recall – SVM tanpa SMOTE



Gambar 10 c. Precision – SVM tanpa SMOTE

Hasil kinerja (*performance*) selain algoritma SVM, yakni LR, NB, dan RF tanpa SMOTE dapat dilihat pada Tabel 1.

Sedangkan, skenario kedua mengevaluasi model klasifikasi sentimen dengan SMOTE. Hasil kinerja (*performance*) algoritma SVM, LR, NB, dan RF dengan SMOTE dapat dilihat pada Tabel 2.

Tabel 1

Confusion Matrix Tanpa SMOTE

Algoritma	Accuracy	Recall	Precision
SVM	88,91%	53,87%	77,85%
NB	75,62%	54,42%	52,91%
RF	88,21%	50%	44,10%
LR	81,76%	80,13%	66,87%

Tabel 2

Confusion Matrix Dengan SMOTE

Algoritma	Accuracy	Recall	Precision
SVM	98,26%	98,26%	98,31%
NB	92,51%	92,51%	93,46%
RF	95,90%	95,90%	95,95%
LR	97,64%	97,64%	97,74%

Berdasarkan Tabel 1 dan 2, *Accuracy* dan *Precision* algoritma SVM mempunyai nilai tertinggi pada dua skenario evaluasi model klasifikasi sentimen. Nilai tertinggi *Recall* untuk skenario tanpa SMOTE ditempati algoritma LR. Sedang, nilai tertinggi *Recall* untuk skenario dengan SMOTE diperoleh algoritma SVM.

Penerapan metode SMOTE *Up Sampling* memberikan dampak kenaikan nilai *confusion matrix* bagi tiap algoritma. Hal tersebut dapat dilihat pada Tabel 2.

KESIMPULAN DAN SARAN

Kesimpulan

Analisis sentimen dengan algoritma Support Vector Machine (SVM) dapat membantu Bali Hai Cruises dalam memahami pengalaman wisatawan dan meningkatkan kualitas layanannya. Perusahaan dapat menggunakan hasil analisis sentimen ini untuk mengidentifikasi aspek-aspek yang perlu diperbaiki dan untuk mengembangkan strategi pemasaran yang lebih efektif.

Saran

Penelitian ini dapat dikembangkan dengan menggunakan metode klasifikasi teks lain yang lebih canggih, seperti LightGBM atau Bidirectional Encoder Representations from Transformers (BERT). Selain itu, penelitian ini juga dapat diperluas dengan menganalisis sentimen ulasan wisatawan terhadap perusahaan pariwisata bahari lain di Bali.

DAFTAR PUSTAKA

- Akhmad, E. P. A., Adi, K., Widodo, A. P. (2023). Machine learning approach to customer sentiment analysis in twitter airline reviews. *E3S Web of Conferences* **448**, 02044 <https://doi.org/10.1051/e3sconf/202344802044>
- Alemayehu, F., Meshesha, M., Abate, J. (2023). Amharic Political Sentiment Analysis Using Deep Learning Approaches. *Scientific Reports* <https://doi.org/10.1038/s41598-023-45137-9>
- Anita, Wulan, Sari., Teguh, Iman, Hermanto., Meriska, Defriani. (2023). Sentiment Analysis of Tourist Reviews Using K-Nearest Neighbors Algorithm and Support Vector Machine. *Sinkron : jurnal dan penelitian teknik informatika*, doi: 10.33395/sinkron.v8i3.12447
- Alzamzami, Fatimah. M. Hoda, A. El Saddik. (2020) "Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation". *IEEE Access*.
- Balihaicruises. (2024). Tentang Kami. Diakses dari <https://balihaicruises.com/about-us/> tanggal 16/04/2024.
- Breiman, L.E.O. (2001). Random Forests. *Mach. Learn.* **45**, p.5-32.
- Chaoyou, Guo., Ma, Yankun., Zhe, Xu., Cao, Mengmeng., Qian, Yao. (2019). An Improved Oversampling Method for imbalanced Data—SMOTE Based on Canopy and K-means. doi: 10.1109/CAC48633.2019.8997367
- Enrico, Laoh., Isti, Surjandari., Nadhila, Idzni, Prabaningtyas. (2019). Enhancing Hospitality Sentiment Reviews Analysis Performance using SVM N-Grams Method. doi: 10.1109/ICSSSM.2019.8887662
- G., Vinodhini., R., M., Chandrasekaran. (2014). Sentiment Mining Using SVM-Based Hybrid Classification Model. doi: 10.1007/978-81-322-1680-3_18
- Hamnah, Rao. (2023). Sentiment Analysis using Deep Learning: A Domain Independent Approach. doi:10.1109/ICEARS56392.2023.10085676
- Inder, M. D. (2022). Sentiment Analysis Techniques in Recent Works Using GRA Methodology. *REST Journal on Banking, Accounting and Business* Vol: 1(3), September 2022 DOI: <https://doi.org/10.46632/jbab/1/3/8>

- Jan Linnenbrink, Carles Milà, Marvin Ludwig, and Hanna Meyer. (2023). kNNDM: k-fold Nearest Neighbour Distance Matching Cross-Validation for map accuracy estimation. doi: 10.5194/egusphere-2023-1308
- Joseph, R., Barr., Marcus, Sobel., Tyler, Thatcher. (2022). Upsampling, a comparative study with new ideas. doi:10.1109/ICSC52841.2022.00059
- Kui, Lu., Jiesheng, Wu. (2019). Sentiment Analysis of Film Review Texts Based on Sentiment Dictionary and SVM. doi: 10.1145/3319921.3319966
- N. Chawla, K. Bowyer, L. Hall, W. Kegelmeyer, (2002). "SMOTE: Synthetic Minority OverSampling Technique," *J. Artif. Intell. Res.*, vol. **16**, p. 321–357.
- Nguyen, TT., Nguyen, PC. (2023). K-Fold Cross-Validation Technique for Predicting Ultimate Compressive Strength of Circular CFST Columns. ICSCEA 2021. Lecture Notes in Civil Engineering, vol 268. Springer, Singapore. https://doi.org/10.1007/978-981-19-3303-5_79
- Ramadhan, N. G. ., & Ramadhan, T. I. . (2022). Analysis Sentiment Based on IMDB Aspects from Movie Reviews using SVM. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 6(1), 39-45. <https://doi.org/10.33395/sinkron.v7i1.11204>
- Saraswati, Kajal Rai & Handa, Disha. (2022). Comparative Analysis of KNN Classifier with K-Fold Cross-Validation in Acoustic-Based Gender Recognition. Lecture Notes in Networks and Systems, vol 445. Springer, Singapore. https://doi.org/10.1007/978-981-19-1412-6_34
- Utkarsh, Pandey., Sanyam, Shukla., Rajesh, Wadhvani. (2022). OCSK-SMOTE: An improved minority oversampling technique for unbalanced dataset using One-Class SVM and K-means. doi: 10.1109/icacite53722.2022.9823814
- V. Malavade and B. Sureshkumar. (2023). "Comparative Analysis of Artificial Intelligence Techniques for Sentiment Analysis and its Challenges," *7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, pp. 670-675, doi: 10.1109/ICICCS56967.2023.10142264.
- Vivek, J., Joshi., Sanjay, W, Patel., Radhika, Agarwal., Himanshu, Arora. (2023). Sentiments Analysis using Machine Learning Algorithms. doi: 10.1109/ICEARS56392.2023.10085432
- Yadu, R. dan Shukla, R. (2022). A Hybrid Model Integrating Adaboost Approach for Sentimental Analysis of Airline Tweets. *Revue d'Intelligence Artificielle* Vol. 36, No. 4, August, 2022, pp. 519-528.
- Yifei, Zhao. (2023). Overview of Deep Learning Methods for Sentiment Analysis. *Advances in Engineering Technology Research*, doi: 10.56028/aetr.5.1.367.2023