

ANALISIS SENTIMEN WISATAWAN LAWANG SEWU MENGGUNAKAN METODE NAÏVE BAYES

Hero Galuh Pyawai¹, Fandy Setyo Utomo²

^{1,2}Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Jawa Tengah, Indonesia
email:¹herogaluhpyawai@gmail.com,²fandy_setyo_utomo@amikompurwokerto.ac.id

Info Artikel

Sejarah artikel:

Diterima 24-01-2024

Revisi 11-11-2024

Diterima 24-12-2024

Kata kunci:

Analisis Sentimen

Cross Validation

K-Fold

TF-IDF

Naïve Bayes

ABSTRAK

Abstract: Lawang Sewu adalah sebuah gedung bersejarah di Semarang, Jawa Tengah, yang kini digunakan sebagai museum oleh PT Kereta Api Indonesia. Analisis sentimen pada ulasan wisatawan di situs TripAdvisor dilakukan untuk mengetahui persepsi pengunjung terhadap Lawang Sewu. Penelitian ini menggunakan algoritma Naïve Bayes dengan library Python MultinomialNB. Data ulasan diproses melalui tahap preprocessing yang mencakup pembersihan data, pelabelan, penghapusan stopword, normalisasi, tokenisasi, dan stemming. Bobot kata dihitung menggunakan metode TF-IDF, dan klasifikasi dilakukan dengan model Naïve Bayes. Untuk validasi, digunakan metode K-Fold Cross Validation dengan 10 fold, menghasilkan akurasi 87,8%. Hasil penelitian menunjukkan bahwa metode Naïve Bayes mampu memberikan wawasan yang berguna bagi pengelola Lawang Sewu untuk meningkatkan pengalaman wisata.

Penulis yang sesuai:

Hero Galuh Pyawai

Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto

Email : herogaluhpyawai@gmail.com

1. PENDAHULUAN

Lawang Sewu adalah sebuah gedung bekas perkantoran yang terletak di Semarang, Jawa Tengah, Indonesia. Gedung ini dibangun sebagai kantor pusat *Nederlandsch-Indische Spoorweg Maatschappij* (NISM) dan kini menjadi milik perusahaan kereta api Kereta Api Indonesia (KAI). Saat ini, bangunan tersebut digunakan sebagai museum dan galeri sejarah perkeretaapian oleh Unit Pelestarian dan Perancangan Arsitektur Pusat dan KAI Wisata. Pada tahun 2023 menurut akurat.co jumlah wisatawan Lawang Sewu mengalami penurunan dikarenakan cuaca panas yang sangat ekstrim pada kota Semarang. Pada situs *Hoops ID Travel* tercatat data total pengunjung Lawang Sewu pada tahun 2022 mencapai 716.000 orang sedangkan pada tahun 2023 mencapai 330.917 orang (hingga Juni 2023). Namun, data dari situs TripAdvisor menunjukkan penurunan jumlah pengunjung pada 2023 dibandingkan 2022. Analisis sentimen dapat memberikan wawasan tentang persepsi wisatawan terhadap destinasi ini, membantu pengelola memahami kekuatan dan kelemahannya.

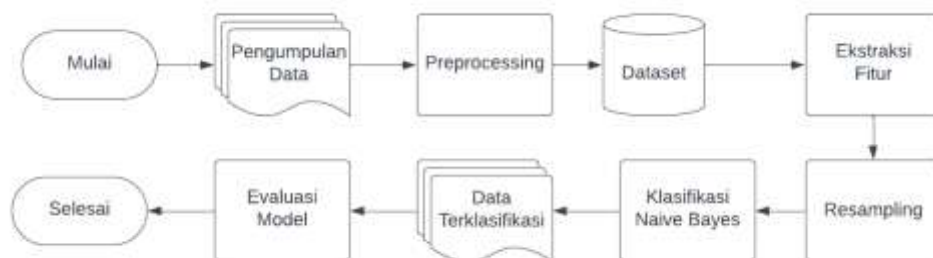
TripAdvisor adalah platform perjalanan *online* yang menyediakan ulasan dan masukan pengguna mengenai hotel, objek wisata, restoran, dan lainnya [1], [2], [3]. Ulasan wisatawan yang pernah berkunjung ke Lawang Sewu dapat memberikan wawasan mengenai pengalamannya di sana. Analisis sentimen adalah teknik

yang dapat digunakan untuk menganalisis suatu teks dan menentukan sentimen di baliknya. Metode ini menggabungkan pembelajaran mesin dan Pemrosesan Bahasa Alami (PBA) untuk mencapai hal ini. Melalui analisis sentimen dasar, program dapat memahami apakah sentimen di balik sebuah teks itu positif, negatif, atau netral.

Metode *Naïve Bayes* merupakan metode yang dapat digunakan untuk melakukan pengujian hipotesis dan juga Algoritme klasifikasi dapat berperan dalam melakukan analisis sentimen. Salah satunya adalah dengan menggunakan *Naïve Bayes*, sebuah algoritma dengan konsep probabilitas tertinggi, yang cocok untuk memprediksi kejadian berdasarkan hasil klasifikasi [4], [5], menurut Laia dan Yamin Sistem yang menganalisis sentimen dengan menggunakan metode *Naïve Bayes* ini, dalam satu kali pengujian, mencapai akurasi sebesar 99,5%, presisi sebesar 99,49%, dan *recall* sebesar 100%. Oleh karena itu, dapat disimpulkan bahwa metode klasifikasi algoritma *Naïve Bayes* cukup relevan meski akurasinya belum mencapai 100% [6], sedangkan menurut Singgalen Algoritma *Naïve Bayes Classifier* mendapatkan Hasil evaluasi pengolahan data destinasi wisata Candi Borobudur menunjukkan kinerja yang baik, Hal ini terlihat dari 3850 data evaluasi yang diolah memberikan *score* presisi sebesar 96,36%, skor presisi sebesar 93,23%, dan *recall* sebesar 100%. Ini adalah sistem klasifikasi berdasarkan asumsi independensi antar prediktor yang dikenal sebagai teorema Bayes. Sederhananya, klasifikasi *Naïve Bayes* mengasumsikan bahwa kehadiran beberapa fitur dalam suatu kelas tidak berhubungan dengan kehadiran fitur lainnya [7], [8]. Oleh karena itu, analisis opini wisata di Lawang Sewu melalui analisis data *TripAdvisor* menggunakan metode *Naïve Bayes* dapat memberikan wawasan mengenai opini wisatawan terhadap tempat tersebut. Analisis ini dapat membantu mengidentifikasi kekuatan dan kelemahan serta memberikan masukan yang berguna bagi pengelola Lawang Sewu untuk meningkatkan pengalaman wisata.

2. METODE

Penelitian ini mengadopsi pendekatan kuantitatif dengan menggunakan metode *web scraping* untuk mengumpulkan data ulasan wisatawan mengenai Lawang Sewu di situs *TripAdvisor*. Metode kuantitatif dipilih karena tujuan penelitian adalah untuk menganalisis data numerik dalam bentuk ulasan teks dan untuk memperoleh pola serta informasi yang dapat diukur dari data tersebut. Analisis dilakukan secara otomatis menggunakan algoritma yang dirancang khusus untuk ekstraksi, klasifikasi, dan evaluasi data teks, yang memungkinkan peneliti memperoleh wawasan yang objektif dan dapat diverifikasi.



Gambar 1. Metode Penelitian

2.1. Pengumpulan Data

Penelitian ini memiliki alur penelitian yang dimulai dengan melakukan pengumpulan data ulasan wisatawan Lawang Sewu di situs *TripAdvisor* menggunakan teknik crawling, Khder menjelaskan *Web Scraping* atau *Web Crawling* mengacu pada prosedur ekstraksi data secara otomatis dari situs web menggunakan perangkat lunak [9]. Ini adalah proses yang sangat penting dalam bidang-bidang seperti Bisnis Intelijen di era modern. *Web scraping* adalah teknologi yang memungkinkan kita mengekstrak data terstruktur dari teks seperti HTML [9], [10]. *Web scraping* sangat berguna dalam situasi di mana data tidak disediakan dalam format yang dapat dibaca mesin seperti JSON atau XML [9]. Penggunaan *web scraping* untuk mengumpulkan data memungkinkan kita untuk mengumpulkan harga hampir secara real time dari situs toko ritel dan memberikan rincian lebih lanjut, *web scraping* dapat juga dapat digunakan untuk mengumpulkan informasi intelijen dari bisnis terlarang seperti narkoba pasar gelap di *darknet* untuk memberikan penegakan hukum dan penegak hukum dan peneliti data berharga seperti harga dan varietas obat

yang yang tidak akan tersedia dengan metode konvensional [9]. Telah ditemukan bahwa menggunakan program Web *Crawling* akan menghasilkan data yang jauh lebih menyeluruh, akurat, dan konsisten daripada entri manual [9].

2.2. *Preprocessing*

Setelah data ulasan terkumpul, tahap selanjutnya adalah preprocessing, yang bertujuan untuk membersihkan dan menyiapkan data agar siap untuk dianalisis lebih lanjut. Proses preprocessing ini terdiri dari beberapa tahapan:

2.2.1. *Cleaning*

Proses ini melibatkan penghapusan simbol, tanda baca, angka, dan karakter non-alfabet yang tidak memberikan informasi penting bagi analisis.

2.2.2. *Labeling*

Memberikan label sentimen pada setiap ulasan berdasarkan skor sentimen yang telah dihitung. Ulasan akan diberi label positif atau negatif.

2.2.3. *Stopword Removal*

Menghapus kata-kata umum yang tidak memberikan makna signifikan, seperti "dan", "atau", "yang", yang hanya akan menambah noise pada analisis.

2.2.4. *Normalisasi*

Menyamakan teks yang memiliki makna serupa namun ejaan berbeda, misalnya "baik" dan "bagus" akan disamakan.

2.2.5. *Tokenization*

Teks ulasan dipecah menjadi unit-unit kecil yang disebut token (kata-kata individu), yang akan digunakan dalam analisis selanjutnya.

2.2.6. *Stemming*

Mengembalikan kata ke bentuk dasarnya, misalnya "berjalan" menjadi "jalan", untuk mengurangi kompleksitas kata yang berhubungan.

2.3. *Ekstraksi Fitur*

Setelah itu melakukan proses ekstraksi fitur disini berfungsi untuk melakukan pembobotan kata menggunakan TF-IDF, Budiman menjelaskan TF-IDF memberikan bobot pada setiap kata berdasarkan seberapa sering kata tersebut muncul dalam dokumen dan seberapa sering kata tersebut muncul di seluruh dokumen. Bobot ini nantinya digunakan oleh algoritma pembelajaran mesin untuk klasifikasi dokumen. Ekstraksi fitur mengubah data teks yang awalnya tidak terstruktur menjadi format terstruktur, siap diproses dengan algoritma *Naïve Bayes* [11]. Ekstraksi fitur meningkatkan efisiensi dan akurasi model klasifikasi, mengurangi *noise* dan *outlier* dalam data, serta mempermudah interpretasi hasil klasifikasi [11], [12].

2.4. *Resampling*

Selanjutnya adalah proses *Resampling* yang merupakan tahapan untuk melakukan penyeimbangan dataset yang berfungsi agar pada tahapan klasifikasi dan evaluasi model dapat mendapatkan hasil yang akurat [13] proses ini menggunakan metode SMOTE.

2.5. *Klasifikasi*



Proses selanjutnya adalah melakukan klasifikasi *Naïve Bayes* algoritma *Naïve Bayes* memiliki kecepatan kategorisasi yang cepat dan akurasi yang tinggi ketika berurusan dengan set data berukuran besar [14], [15]. Dalam melakukan klasifikasi digunakan data hasil pada tahap *preprocessing* sampai tahap pembobotan kata lalu kita akan mendapatkan data *training* yang kemudian akan kita lakukan pengujian dengan menggunakan data *test* untuk menguji hasil keakuratan dari klasifikasi yang telah dilakukan.

2.6. Evaluasi

Tahapan yang terakhir yaitu Evaluasi model yang digunakan untuk mengetahui kinerja model yang digunakan menggunakan matriks konfusi untuk memperoleh nilai *accuracy*, *precision*, *recall*, *f1-score* dan dilakukan pengujian kedua menggunakan *k-fold cross validation* untuk menentukan hasil evaluasi yang maksimal.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Penelitian ini menggunakan data ulasan wisatawan Lawang Sewu pada situs *TripAdvisor* dengan jumlah data sekitar 1237 dimana data tersebut di dapatkan melalui proses *Crawling* pada situs *TripAdvisor* menggunakan *Instant Data Scraper* yang merupakan ekstensi google chrome.

3.2. Preprocessing

Tahap selanjutnya yaitu *preprocessing* tahap ini dilakukan untuk mendapatkan dataset yang bersih, disini terdapat *library* yang di perlukan dan di dalam melakukan *preprocessing* terdapat beberapa tahap yaitu *Cleaning*, *Labeling*, *Normalisasi*, *Remove Stopword*, *Tokenisasi*, *Stemming*.

3.2.1. Cleaning

Berfungsi untuk penghapusan tanda khusus pada teks seperti simbol yang tidak diperlukan, tanda baca, karakter khusus, dan merubah teks menjadi huruf kecil semua, dirujuk pada tabel 1.

Tabel 1. Cleaning

Sebelum	Sesudah
Halo!!! Apa kabar? Sudahkah kamu makan?#pertanyaan #hariIni	halo apa kabar sudahkah kamu makan pertanyaan hariini
Kamu siapa???	kamu siapa
No rumah 20	no rumah

3.2.2. Labeling

Pada proses ini berfungsi untuk memberikan label agar kita dapat mengetahui *sentiment* yang terdapat pada dataset yang sudah kita *cleaning* sebelumnya, dirujuk pada tabel 2.

Tabel 2. Labeling

Ulasan	Sentiment Score	Sentiment
terimakasih	5	positif
sesal	-4	negatif
kenamaan	5	positif
membunuh	-4	negatif
mencederai	-4	negatif
tidak segan	-3	negatif

3.2.3. Remove Stopword

Tahapan ini berfungsi untuk menghapus kata-kata umum yang terdapat pada dataset, dirujuk pada tabel 3.



Tabel 3. Remove Stopword

Sebelum	Sesudah
halo apa kabar sudahkah kamu makan	halo kabar sudahkah makan
ini adalah contoh kalimat dengan tanda baca yang berlebihan	ini contoh kalimat tanda baca berlebihan
apa kabar pertanyaan seseorang	kabar pertanyaan seseorang

3.2.4. Normalisasi

Proses ini digunakan untuk menyamaratakan text yang memiliki makna yang sama tetapi penulisannya berbeda, dirujuk pada tabel 4.

Tabel 4. Normalisasi

Sebelum	Sesudah
ying	yang
hi	hai
jgn	jangan

3.2.5. Tokenisasi

Proses ini dilakukan untuk membagi text menjadi unit-unit lebih kecil atau disebut juga dengan token, dirujuk pada tabel 5.

Tabel 5. Tokenisasi

Sebelum	Sesudah
halo kabar sudahkah makan	["halo", "kabar", "sudahkah", "makan"]
ini contoh kalimat tanda baca berlebihan	["ini", "contoh", "kalimat", "tanda", "baca", "berlebihan"]
kabar pertanyaan seseorang	["kabar", "pertanyaan", "seseorang"]

3.2.6. Stemming

Ini adalah tahapan terakhir pada *preprocessing*, tahapan ini dilakukan untuk mengembalikan kata ke bentuk dasarnya, dirujuk pada tabel 6.

Tabel 6. Ekstraksi Fitur

Sebelum	Sesudah
halo kabar sudahkah makan	halo kabar sudah makan
ini contoh kalimat tanda baca berlebihan	ini contoh kalimat tanda baca lebih
kabar pertanyaan seseorang	kabar tanya orang

3.3. Ekstraksi Fitur

Pada tahap ini kita akan melakukan pembobotan kata menggunakan TF-IDF, fungsi tahap TF adalah menghitung kemunculan kata dalam suatu dokumen. Fitur frekuensi istilah memungkinkan Anda menentukan bobot suatu kata dalam suatu dokumen berdasarkan berapa kali kata tersebut muncul dalam dokumen. Cara memberi bobot pada persamaan *frekuensi term* (TF) dan frekuensi dokumen invers (IDF) [16], [17]. dapat dilihat pada persamaan berikut:

$$W_{dt} = TF_{dt} * IDF_{ft} \quad (1)$$

TF-IDF ini digunakan untuk menemukan representasi nilai dari kumpulan data pelatihan. Ini membentuk vektor antara dokumen dan kata, yang digabungkan berdasarkan kesamaan antara dokumen dan kata [16].



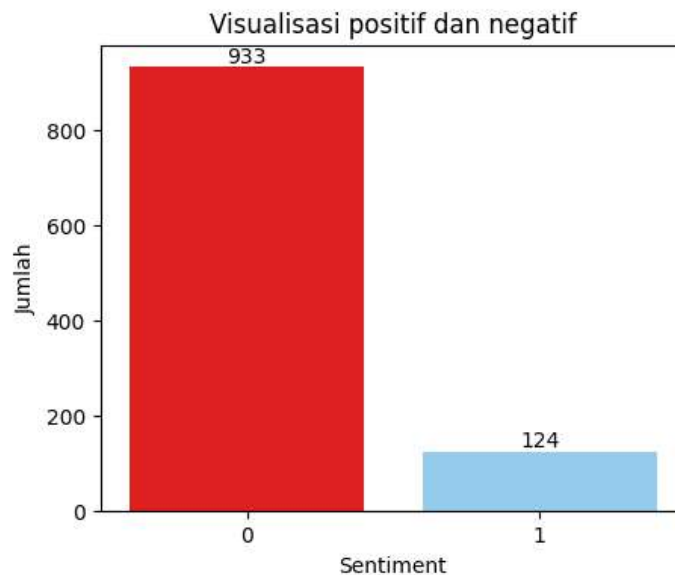
```
[7.27098843 7.27098843 6.01822546 ... 6.86552332 3.95680243 7.27098843]
(0, 3197) 0.19356976691234032
(0, 3006) 0.27238073790898704
(0, 3002) 0.3070319752435114
(0, 2880) 0.15543441776153052
(0, 2851) 0.1275478409977281
(0, 2491) 0.21901019044293155
```

Gambar 2. Hasil TF-IDF

Dirujuk pada gambar 2 kolom pertama berarti jumlah dokumen, kolom kedua berarti letak kata kolom ketiga berarti hasil dari tf

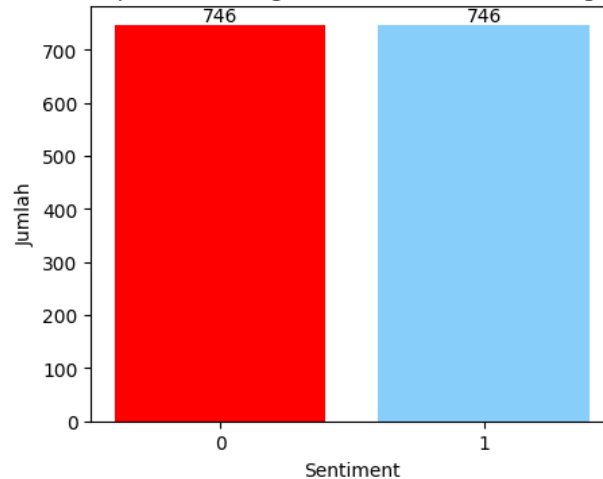
3.4. Resampling

Pada proses *resampling* kita akan menggunakan fungsi SMOTE. Metode SMOTE merupakan cara ampuh untuk mengatasi masalah data yang tidak seimbang dan telah terbukti berhasil dalam berbagai kasus penggunaan [13]. *Synthetic Minority Over-sampling* (SMOTE) dapat berguna ketika Anda menangani masalah ketidakseimbangan kelas, masalah yang sangat umum terjadi pada tugas klasifikasi. SMOTE mencoba mengatasi keterbatasan dalam teknik pengambilan sampel secara acak. *Random over-sampling* membuat sampel baru dengan mengkloning sampel minoritas yang sudah ada; hal ini meningkatkan volume sampel minoritas dalam dataset tetapi tidak memasukkan sampel unik ke dalam dataset. Hal ini dapat menyebabkan *over-fitting* pada algoritme pembelajaran karena algoritme ini sekarang melihat lebih banyak salinan dari sampel minoritas yang sama. Alih-alih membuat replika tambahan dari sampel minoritas asli, SMOTE menghasilkan sampel minoritas sintetis yang tidak identik dengan sampel yang ada. Sampel SMOTE tinggal berdekatan dengan sampel minoritas asli. Cara kerja SMOTE dapat di jelaskan sebagai berikut: laju sampling ini adalah nilai persen yang memberi tahu SMOTE berapa banyak sampel minoritas tambahan yang akan dihasilkan untuk setiap sampel minoritas asli. Sebagai contoh, jika mengatur tingkat pengambilan sampel ke 200%, SMOTE akan menghasilkan 2 sampel minoritas baru untuk setiap sampel minoritas asli. Jika menetapkan nilai ini kurang dari 100%, misalnya 50%, SMOTE akan secara acak memilih 50% sampel minoritas asli dan hanya untuk masing-masing sampel minoritas yang akan menghasilkan sampel minoritas baru, dirujuk pada gambar 3 dan gambar 4.



Gambar 3. Sebelum dilakukan SMOTE

Visualisasi positif dan negatif setelah imbalance dengan SMOTE



Gambar 4. Setelah dilakukan SMOTE

3.5. Klasifikasi Naïve Bayes

Naïve Bayes adalah pengklasifikasi probabilitas sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data tertentu. *Naïve Bayes* didasarkan pada asumsi yang disederhanakan bahwa nilai atribut tidak bergantung pada kondisi ketika diberikan nilai keluaran [18]. *Naïve Bayes* juga diartikan sebagai klasifikasi menggunakan metode probabilistik dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang masa depan dari pengalaman masa lalu [19]. Untuk menjelaskan metode *Naïve Bayes*, perlu diperhatikan bahwa proses klasifikasi memerlukan beberapa petunjuk untuk menentukan kelas mana yang sesuai untuk sampel yang dianalisis. Jadi metode Bayesian di atas diadaptasi sebagai berikut:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Akurasi Model Naïve Bayes : 0.8726415094339622

Gambar 5. Hasil akurasi Naïve Bayes

Dirujuk pada gambar 5 hasil akurasi dari metode Naïve Bayes mendapatkan hasil akurasi sebesar 0.872 atau sebesar 87,2%

3.6. Evaluasi Model

Pada tahapan ini dilakukan evaluasi model menggunakan *K-Fold Cross Validation* yang merupakan salah satu teknik yang digunakan untuk mengklasifikasikan data menjadi data latih dan data uji [20]. Teknik ini banyak digunakan oleh para peneliti karena terbukti dapat mengurangi bias yang terjadi pada saat pengambilan sampel. Validasi silang *K-fold* secara terus menerus membagi data menjadi data pelatihan dan data pengujian, memberikan kesempatan kepada semua data untuk menjadi data pengujian. *K* adalah jumlah nomor pengurutan data yang digunakan untuk membagi pelatihan dan pengujian. Dirujuk pada tabel 6 memperlihatkan partisi data menggunakan metode validasi silang *K-Fold*.

Tabel 6. Partisi data menggunakan metode K-Fold Cross Validation

Fold 1	Fold 2	Fold 3	...	Fold K
Test	Train	Train	...	Train
Train	Test	Train	...	Train
Train	Train	Test	...	Train
...
Train	Train	Train	...	Test

Dari proses *Cross Validation* menggunakan *K-Fold* yang menggunakan 10 *Fold* di dapatkan hasil akurasi sebagai berikut:

```
Akurasi setiap fold: 0.8823529411764706
Akurasi setiap fold: 0.8823529411764706
Akurasi setiap fold: 0.8705882352941177
Akurasi setiap fold: 0.8352941176470589
Akurasi setiap fold: 0.8352941176470589
Akurasi setiap fold: 0.9047619047619048
Akurasi setiap fold: 0.8809523809523809
Akurasi setiap fold: 0.8690476190476191
Akurasi setiap fold: 0.9166666666666666
Akurasi setiap fold: 0.9047619047619048

Akurasi rata-rata: 0.8782072829131652
```

Gambar 6. Hasil akurasi K-Fold Cross Validation

Dirujuk pada gambar 6 hasil dari akurasi dari *K-Fold Cross Validation* mendapatkan hasil akurasi sebesar 0.878 atau 87,8%

4. KESIMPULAN

Berdasarkan hasil penelitian Analisis Sentimen Wisatawan Lawang Sewu menggunakan Metode *Naïve Bayes* yang telah diuraikan pada bab sebelumnya, dapat disimpulkan bahwa tingkat akurasi *Naïve Bayes* dalam penelitian ini mencapai 0.872 atau 87,2%. Untuk memverifikasi hasil akurasi *Naïve Bayes*, pengecekan ulang juga dilakukan menggunakan metode *K-Fold Cross Validation*, yang menghasilkan tingkat akurasi sebesar 0.878 atau 87,8%.

REFERENSI

- [1] A. H. Suasapha and P. P. Bali, "Jurnal kajian bali," no. 2, pp. 410–438, 2023.
- [2] I. Y. Saragih and G. I. Bhaskara, "Pencitraan Sosial Media," *J. Destin. Pariwisata*, vol. 7, no. 2, p. 231, 2019, doi: 10.24843/jdepar.2019.v07.i02.p04.
- [3] R. Sarudin, "Analisis Online Review Tripadvisor . Com Terhadap Minat Pembelian Produk Jasa Akomodasi Di Hotel Manhattan Tripadvisor . Com Review Online Analysis on the Interest of Buying Accommodation Services in Hotel," *J. Hosp. dan Pariwisata*, vol. 7, pp. 33–46, 2021.
- [4] F. Amaliah and I. K. Dwi Nuryana, "Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter," *J. Informatics Comput. Sci.*, vol. 3, no. 03, pp. 384–393, 2022, doi: 10.26740/jinacs.v3n03.p384-393.
- [5] S. Samsir, K. Kusmanto, A. H. Dalimunthe, R. Aditiya, and R. Watrianthos, "Implementation Naïve Bayes Classification for Sentiment Analysis on Internet Movie Database," *Build. Informatics, Technol. Sci.*, vol. 4, no. 1, pp. 1–6, 2022, doi: 10.47065/bits.v4i1.1468.
- [6] E. Laia and M. Yamin, "Penerapan Algoritma Naïve Bayes dalam Menganalisis Sentimen pada Review Pengguna E-Commerce," *Media Online*, vol. 4, no. 1, pp. 305–316, 2023, doi: 10.30865/klik.v4i1.1186.
- [7] Y. A. Singgalan, "Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, 2022, doi: 10.47065/bits.v4i3.2486.
- [8] D. N. N. Husnina, D. E. Ratnawati, and ..., "Analisis Sentimen Pengguna Aplikasi RedBus berdasarkan Ulasan di Google Play Store menggunakan Metode Naïve Bayes," ... *Teknol. Inf. dan ...*, vol. 7, no. 2, 2023, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12297%0Ahttps://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/12297/5593>
- [9] M. A. Khder, "Web scraping or web crawling: State of art, techniques, approaches and application," *Int. J. Adv. Soft Comput. its Appl.*, vol. 13, no. 3, pp. 144–168, 2021, doi:

- 10.15849/ijasca.211128.11.
- [10] J. A. Septian, T. M. Fachrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [11] irwan budiman, M. R. Faisal, and D. T. Nugrahadi, “Studi Ekstraksi Fitur Berbasis Vektor Word2Vec pada Pembentukan Fitur Berdimensi Rendah,” *J. Komputasi*, vol. 8, no. 1, 2020, doi: 10.23960/komputasi.v8i1.2517.
- [12] Regina, T. H. Saragih, and D. Kartini, “Analisis Sentimen Brand Ambassador Bts Terhadap Tokopedia Menggunakan Klasifikasi Bayesian Network Dengan Ekstraksi Fitur Tf-Idf,” *J. Inform. Polinema*, vol. 9, no. 4, pp. 383–390, 2023, doi: 10.33795/jip.v9i4.1333.
- [13] H. Utami, “Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network,” *Indones. J. Appl. Stat.*, vol. 5, no. 1, p. 31, 2022, doi: 10.13057/ijas.v5i1.56825.
- [14] H. N. Ed, *Cyberspace Data and Intelligence, Syndrome, and Health*. 2019. doi: 10.1007/978-981-15-1922-2.
- [15] 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.
- [16] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, “Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter,” *J. Inf. Syst. Manag.*, vol. 3, no. 2, pp. 44–49, 2021, doi: 10.24076/joism.2021v3i2.558.
- [17] V. Rupapara, F. Rustam, H. F. Shahzad, A. Mehmood, I. Ashraf, and G. S. Choi, “Impact of SMOTE on Imbalanced Text Features for Toxic Comments Classification Using RVVC Model,” *IEEE Access*, vol. 9, pp. 78621–78634, 2021, doi: 10.1109/ACCESS.2021.3083638.
- [18] P. T. Torini *et al.*, “JINTEKS (Jurnal Informatika Teknologi dan Sains) ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI PERPANJANGAN KONTRAK KERJA Program Studi Teknik Informatika , Fakultas Teknik , Universitas Pelita Bangsa PENDAHULUAN Kontrak kerja merupakan sesuatu perjanjian ant,” vol. 5, no. 3, pp. 516–524, 2023.
- [19] A. Saleh, “Klasifikasi Metode Naive Bayes Dalam Data Mining Untuk Menentukan Konsentrasi Siswa (Studi Kasus Di Mas Pab 2 Medan),” 2014.
- [20] D. Abriha, P. K. Srivastava, and S. Szabó, “Smaller is better? Unduly nice accuracy assessments in roof detection using remote sensing data with machine learning and k-fold cross-validation,” *Heliyon*, vol. 9, no. 3, 2023, doi: 10.1016/j.heliyon.2023.e14045.

