

Sistem Rekomendasi Wisata Berbasis Konten dengan Algoritma Naïve Bayes Multinomial

Arisman Mahfud¹, Ika Nur Fajri²

^{1,2} Sistem Informasi, Universitas Amikom Yogyakarta, Indonesia

Email: arismanmahfud@students.amikom.ac.id¹, fajri@amikom.ac.id²

Abstract

The development of tourism in Indonesia offers vast business opportunities for tourism stakeholders, especially in Yogyakarta, which is rich in natural tourist destinations. However, visitors often face difficulties in selecting suitable destinations due to limited information. This study aims to develop a tourism recommendation system in Yogyakarta by utilizing the Naïve Bayes Multinomial algorithm to improve the accuracy of recommendations and the overall tourism experience. The system leverages word frequency information from textual data to generate better recommendations. Data from 40 tourist destinations, including descriptions, locations, ratings, categories, and facilities, is used in this study. The model's results show an accuracy of 81.2%, precision of 62.5%, recall of 62.5%, and an F1 score of 62.5%. Testing with the keyword "photo spot" yields an accuracy of 50%, precision of 25%, recall of 50%, and an F1 score of 33.33%, while the keyword "natural scenery" achieves 100% accuracy, 100% precision, 100% recall, and an F1 score of 100%. These results demonstrate that the system can provide accurate and useful tourism recommendations for users.

Keywords: tourism, recommendation system, Yogyakarta tourism, Naïve Bayes Multinomial, accuracy.

Abstrak

Pengembangan pariwisata di Indonesia memberikan peluang bisnis yang luas bagi para pelaku pariwisata, teru-tama di Yogyakarta yang kaya akan destinasi wisata alam. Namun, pengunjung sering mengalami kesulitan dalam menentukan destinasi yang sesuai karena terbatasnya informasi. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi wisata di Yogyakarta dengan memanfaatkan algoritma Naïve Bayes Multinomial guna meningkatkan akurasi rekomendasi dan pengalaman wisata. Sistem ini memanfaatkan informasi frekuensi kata dalam data tekstual untuk menghasilkan rekomendasi yang lebih baik. Data dari 40 destinasi wisata, termasuk deskripsi, lokasi, rating, kategori, dan fasilitas, digunakan dalam penelitian ini. Hasil model menunjukkan akurasi sebesar 81,2%, presisi 62,5%, recall 62,5%, dan skor F1 sebesar 62,5%. Pengujian dengan kata kunci "spot foto" menunjukkan akurasi 50%, presisi 25%, recall 50%, dan skor F1 sebesar 33,33%, sedangkan kata kunci "pemandangan alam" mencapai akurasi 100%, presisi 100%, recall 100%, dan skor F1 sebesar 100%. Hasil ini membuktikan bahwa sistem dapat memberikan rekomendasi wisata yang akurat dan bermanfaat bagi pengguna.

Kata kunci: Wisata, Sistem Rekomendasi, Wisata Yogyakarta, Naïve Bayes Multinomial, Akurasi.

1. PENDAHULUAN

Perkembangan pariwisata di Indonesia memberikan peluang bisnis bagi pelaku industri dengan banyaknya daerah yang membangun objek wisata untuk menarik wisatawan. Hal ini tidak hanya meningkatkan popularitas daerah, tetapi juga pendapatan masyarakat lokal. Di Yogyakarta, potensi wisata alam tersebar di seluruh wilayah, memperkaya ragam destinasi yang menarik bagi wisatawan lokal dan asing [1]. Pesona alam Yogyakarta yang menakjubkan memberikan efek menenangkan bagi para pengunjung, sehingga menjadikannya salah satu destinasi wisata favorit. Berdasarkan data dari Badan Pusat Statistik, jumlah wisatawan luar

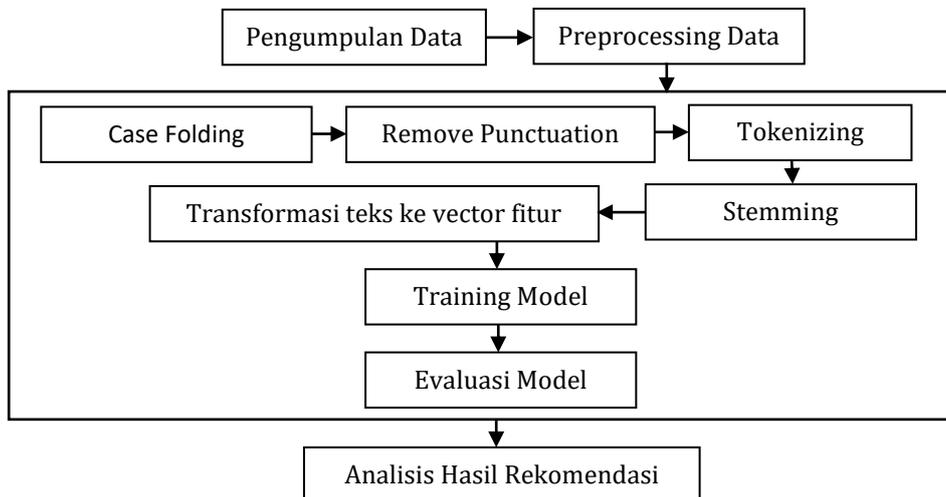
negeri yang mengunjungi Yogyakarta pada bulan Januari 2024 turun 39,33% dibandingkan Desember 2023, dari 11.338 menjadi 6.879 kunjungan [2]. Salah satu daya tarik utama di Yogyakarta adalah Pantai Parangtritis yang terkenal, terutama dari ketinggian bukit paralayang. Tempat ini menawarkan panorama spektakuler dari pantai yang luas dengan deburan ombak yang menggulung dan pasir putih yang panjang [3]. Selain keindahan alamnya, Yogyakarta juga memiliki banyak situs sejarah termasuk Candi Prambanan, yang dikenal sebagai candi Hindu dengan ukiran dominan di Indonesia [4]. Meskipun kekayaan wisata yang melimpah, pengunjung sering kali mengalami kesulitan dalam memilih destinasi wisata yang sesuai, disebabkan oleh keterbatasan informasi yang tersedia. Untuk mengatasi masalah ini, diperlukan pengembangan Sistem Rekomendasi (SR) yang mampu memberikan saran yang selaras untuk pengguna. Beberapa penelitian sebelumnya telah berhasil mengembangkan sistem rekomendasi untuk berbagai keperluan, termasuk wisata. Sebagai contoh, penelitian menggunakan metode Simple Additive Weighting telah menghasilkan aplikasi online yang merekomendasikan pantai-pantai yang tepat beserta detail informasinya seperti nama, alamat, koordinat GPS, dan harga tiket masuk [5]. Selain itu, sistem yang didasarkan pada Expert System dan Algoritma Genetika mampu memberikan rekomendasi wisata yang lebih personal berdasarkan preferensi pengguna [6].

Pada sisi lain, implementasi Metode Weighted Product di Kota Malang telah membuktikan keberhasilannya dalam memberikan rekomendasi lokasi wisata dan kuliner yang optimal untuk wisatawan [7]. Begitu juga dengan penelitian menggunakan metode Topsis di Malang, yang telah menghasilkan sistem informasi web untuk rekomendasi destinasi wisata terbaik [8]. Banyaknya pilihan wisata di Yogyakarta sering membuat pengunjung bingung dalam memilih destinasi, ditambah lagi dengan keterbatasan informasi mengenai tempat-tempat wisata tersebut. Oleh karena itu, diperlukan SR yang mampu memberikan saran wisata di Yogyakarta, sehingga pengunjung dapat menikmati pengalaman wisata yang lebih memuaskan. Penelitian sebelumnya telah membangun sistem rekomendasi wisata Yogyakarta menggunakan metode hybrid filtering, yang merupakan perpaduan antara collaborative filtering dan content-based filtering. Hasil studi ini, berdasarkan hasil uji coba black-box, membuktikan bahwa sistem berfungsi baik. Uji alpha yang dilakukan dengan 10 responden memberikan 52,6% jawaban "Baik", sementara uji beta dengan 40 responden menunjukkan 77,3% jawaban "Setuju". Hal ini mengindikasikan sistem yang dibuat mampu mengakomodasi kebutuhan pengguna [9].

Meskipun nilai akurasi pada penelitian ini sudah tinggi, pemilihan algoritma yang digunakan kurang tepat untuk mencapai akurasi maksimal. Karena itu, penelitian ini bertujuan untuk pengembangan sistem rekomendasi wisata di Yogyakarta menggunakan algoritma Naïve Bayes Multinomial. Diharapkan bahwa dengan kemampuannya dalam mengelola data teks kompleks dan memanfaatkan informasi frekuensi kata, algoritma ini akan meningkatkan akurasi rekomendasi dan memberikan pengalaman wisata yang lebih memuaskan bagi pengguna.

2. METODOLOGI PENELITIAN

Penelitian ini dilakukan dengan persiapan data, prapemrosesan data, klasifikasi dengan algoritma Naive Bayes lalu analisa hasil. Adapun detail dari metodologi penelitian ditampilkan pada Gambar 1 berikut:



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan data

Data dikumpulkan melalui sumber online yang menyediakan informasi tentang tempat wisata di Yogyakarta, terutama dari platform Google dan Google Maps. Data yang dikumpul-kan meliputi deskripsi, lokasi, rating, kategori, dan fasilitas dari setiap tempat wisata. Setelah proses pengum-pulan data selesai, data ini disusun dan disimpan sebagai sebuah dataset yang dinamai "wisata_yogyakarta".

2.2. Case folding data

Setelah data di peroleh, langkah selanjutnya yaitu *Case folding* data, bertujuan untuk membersihkan teks dari noise sehingga menghasilkan data berkualitas untuk pelatihan model Naive Bayes Multinomial [10]. Data tempat wisata di Yogyakarta yang dikumpulkan dari sumber online akan melalui beberapa tahapan prepro-cessing, yaitu:

- a) *Case Folding* proses mengganti huruf dalam teks menjadi huruf kecil [11].
- b) *Remove punctuation* proses untuk menghilangkan tanda baca dari teks seperti koma, titik, kurung, dan karakter khusus lainnya [12].
- c) *Tokenizing* proses yang digunakan untuk memecah sebuah string menjadi kata-kata yang ada di dalamnya. Dengan kata lain, ini adalah cara untuk membagi teks dari kalimat atau paragraf menjadi bagian-bagian yang lebih kecil [13].
- d) *Stemming* proses mengubah teks menjadi kata dasar [14].
- e) Transformasi Teks ke Vektor fitur Dalam tahap ini, teks diubah menjadi representasi vektor numerik menggunakan teknik TF-IDF (Term Frequency-Inverse Document Frequency). TF (Term Frequency) mengukur tingkat

keseringan sebuah term muncul dalam dokumen tersebut [15]. Semakin sering suatu term muncul dalam sebuah dokumen, semakin tinggi bobot atau nilai kesesuaian yang diberikan. IDF (Inverse Document Frequency), Menentukan tingkat kelangkaan sebuah term di seluruh kumpulan dokumen [16]. Nilai IDF semakin besar jika jumlah dokumen yang mengandung term tersebut semakin sedikit. Berikut adalah penjelasan mengenai TF-IDF dan perhitungannya:

Rumus TF-IDF:

$$TF\text{-}IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

Term Frekuensi (TF) mengukur frekuensi kemunculan kata t dalam dokumen d dibandingkan dengan total kata dalam dokumen tersebut.

Rumus TF adalah:

$$TF(t, d) = \frac{f(t, d)}{\sum_{t'} f(t', d)} \quad (2)$$

$f(t, d)$: frekuensi kemunculan term t dalam dokumen d ;

$\sum_{t'} f(t', d)$: total frekuensi semua kata dalam dokumen d .

Inverse Document Frequency (IDF) mengevaluasi pentingnya kata t di semua korpus dokumen. Rumus IDF adalah:

$$IDF(t) = \log\left(\frac{N}{df(t)+1}\right)+1 \quad (3)$$

N : jumlah total dokumen dalam korpus

$df(t)$: jumlah dokumen yang mengandung term t

Perhitungan TF-IDF untuk dokumen pertama, salah satu kata didokumen pertama yaitu kata "2011". Frekuensi kemunculan kata "2011" dalam dokumen pertama adalah 1, dan total kata dalam dokumen pertama adalah 52. Jumlah dokumen dalam korpus adalah 40 dan jumlah dokumen yang mengandung kata "2011" adalah 1.

Menghitung TF:

$$TF("2011", d1) = \frac{1}{\text{total kata dalam dokumen pertama}} = \frac{1}{52} = 0.0172$$

Menghitung IDF:

$$IDF("2011") = \log\left(\frac{40}{1+1}\right) + 1 = \log 20 + 1 = 2.3010$$

Menghitung TF-IDF:

$$(2011, d1) = 0.0172 \times 2.3010 = 0.0396$$

Ini adalah perhitungan manual untuk nilai TF-IDF kata "2011" dalam dokumen pertama. Jadi, nilai TF-IDF untuk kata "2011" dalam dokumen pertama adalah 0.0396.

- f) *Training* Model pada tahapan ini, model Naive Bayes Multinomial akan dilatih menggunakan data yang sudah diolah dan diubah menjadi representasi vektor TF-IDF, Algoritma Naive Bayes merupakan algoritma klasifikasi yang menggunakan probabilitas berdasarkan teorema Bayes. Algoritma ini

menganalisis dataset guna memprediksi kelas berdasarkan label, menghitung probabilitas setiap kelas untuk input, dan memprediksi kelas yang belum diketahui [17]. Algoritma ini bekerja berdasarkan kon-sep frekuensi kata, yaitu tingkat kata yang muncul dalam dokumen [18]. Algoritma Naïve Bayes didasarkan pada teorema Bayes, yang dapat digambarkan sebagai rumus berikut:

$$P(C_k|x) = \frac{P(C_k) \cdot P(x|C_k)}{P(x)} \quad (4)$$

Dimana:

$P(C_k|x)$: probabilitas posterior kategori C_k diberikan fitur x ;

$P(C_k)$: probabilitas prior dari kategori C_k Ini adalah seberapa sering kategori C_k muncul dalam dataset dibandingkan dengan kategori lainnya;;

$P(x|C_k)$: probabilitas likelihood dari fitur x diberikan kategori C_k ;

$P(x)$: probabilitas fitur x , yang merupakan konstanta untuk semua kategori dalam prediksi;

C_k : kategori atau kelas ke- k ;

x : fitur atau kata yang muncul.

Karena $P(x)$ dalam konteks klasifikasi, yang sering digunakan adalah bentuk yang lebih sederhana:

$$P(C_k|x) \propto P(C_k) \cdot P(x|C_k) \quad (5)$$

Probabilitas prior dari kategori C_k , kemungkinan bahwa dokumen termasuk dalam kategori C_k tanpa mempertimbangkan fitur lainnya, dihitung dengan rumus :

$$P(C_k) = \frac{N_k}{N} \quad (6)$$

Dimana:

N_k : jumlah dokumen dalam kategori C_k ;

N : jumlah total dokumen.

Probabilitas likelihood dari kata x_i diberikan kategori C_k dalam algoritma Naïve Bayes Multinomial dihitung dengan rumus:

$$P(x_i|C_k) = \frac{N_{i,k} + \alpha}{N_k + \alpha \cdot V} \quad (7)$$

Dimana:

$N_{i,k}$: frekuensi kemunculan kata x_i dalam dokumen kategori C_k ;

N_k : jumlah total kata dalam dokumen kategori C_k ;

α : Parameter smoothing, juga dikenal sebagai Laplace Smoothing. Smoothing digunakan untuk menghindari masalah probabilitas nol ketika sebuah kata dalam dokumen tidak pernah muncul dalam kategori tertentu;

x_i : fitur atau kata ke- i teks dalam dokumen;

V : jumlah total kata unik dalam semua dokumen.

Untuk menentukan kategori dari sebuah dokumen berdasarkan model, maka perlu menghitung probabilitas posterior dari semua kategori dan memilih kategori dengan probabilitas tertinggi. Rumusnya adalah:

$$P(C_k|x_1, x_2, \dots, x_n) \propto P(C_k) \cdot \prod_{i=1}^n P(x_i|C_k) \tag{8}$$

x_1, x_2, \dots, x_n : fitur (kata) dari dokumen yang sedang diklasifikasikan kategori dengan probabilitas tertinggi. Untuk menentukan kategori dokumen, maka perlu menghitung probabilitas posterior dari semua kategori dan memilih kategori dengan probabilitas tertinggi dengan rumus:

$$\hat{C} = \arg \max_{C_k} [P(C_k) \cdot \prod_{i=1}^n P(x_i|C_k)] \tag{9}$$

\hat{C} : kategori yang diprediksi atau hasil dari klasifikasi

$\text{Arg max } C_k$: mencari kategori C_k yang memberikan nilai terbesar untuk ekspresi $P(C_k) \cdot \prod_{i=1}^n P(x_i|C_k)$

Proses pelatihan (training) model ini akan melibatkan pembelajaran dari data yang telah dipisahkan sebelumnya menjadi data training dan data testing untuk menghindari overfitting (menghafal) dan memastikan generalisasi model yang baik.

- g) Evaluasi Model, Setelah model dilatih, langkah selanjutnya adalah menilai kinerjanya memakai metrik yang relevan seperti Accuracy, Precision, Recall, dan F1-score. Evaluasi ini bertujuan guna mengukur sejauh mana kinerja model mampu melakukan perkiraan atau memberikan rekomendasi pada data yang masih asing dan belum di eksplorasi [19]. Setelah model dievaluasi, langkah terakhir adalah menggunakan model untuk merekomendasikan tempat wisata di Yogyakarta berdasarkan preferensi pengguna. Model yang telah dilatih akan menghasilkan rekomendasi berdasarkan analisis data yang telah diproses sebelumnya.

3. HASIL DAN PEMBAHASAN

Data yang diperoleh dan digunakan pada studi ini sebanyak 40 data wisata di Yogyakarta, Data yang dikumpulkan meliputi deskripsi, lokasi, rating, ategori, dan fasilitas dari setiap tempat wisata. Berikut Gambar 2 merupakan data yang digunakan:

No	Nama Wisata	Deskripsi	Kategori	Alamat	Rating	Fasilitas
1.	Bukit Klangon	Berada di Desa Glagaharjo, Bukit Klangon...	Wisata sleman	Kalitengah Lor,...	4.6	Toilet,Tempat parkir...
2.	Lava Tour Merapi	Lava Tour Merapi Jogja adalah...	Wisata sleman	Ngipiksari, Hargobinangun...	4.7	Satu mobil jeep untuk 4 orang dewasa....
3.	Obelix Hills	Obelix Hills merupakan salah satu...	Wisata sleman	Jl. Raya Solo Km 17 Kranggan...	4.6	Tempat parkir, Mushola...
4.	Candi Prambanan	Candi Prambanan adalah...	Wisata sleman	Jl. Raya Solo-Yogyakarta Nomor 16...	4.7	Mushola...
40.	Waduk Sermo	Tidak jauh dari Kalibiru...	Wisata kulon progo	Sremo Tengah, Hargowilis...	4.6	Pemancingan ikan...

Gambar 2. Dataset Wisata Yogyakarta

Setelah dataset diperoleh, langkah selanjutnya pada kolom Deskripsi dilakukan *Case folding* data, yang merupakan sumber data utama untuk sistem rekomendasi. Tahapan *Case folding* data meliputi *case folding*, *Remove punctuation*, *Tokenizing*, *Stemming*, dan transformasi teks ke vektor fitur. Berikut adalah penjelasan dan hasil dari setiap tahap pengolahan data yang telah dilakukan:

Case folding pada tahap ini teks dalam kolom Deskripsi diubah menjadi huruf kecil untuk memastikan konsistensi dalam analisis data. Hasil dari tahap *case folding* disimpan dalam kolom *lwr*, berikut hasil nya ditampilkan pada Gambar 3:

```

Hasil case folding:
                                lwr
0 berada di desa glagaharjo, bukit klangon adala...
1 lava tour merapi jogja adalah aktivitas wisata...
2 obelix hills merupakan salah satu objek wisata...
3 candi prambanan adalah bangunan candi bercora...
4 ledok sambu merupakan salah satu wisata di yog...
5 museum gunung merapi ini menampilkan berbagai ...
    
```

Gambar 3. Hasil *Case Folding*

Kemudian *Remove punctuation*, teks di hapus tanda bacanya dan disimpan dalam kolom *clean_punct*, berikut hasil nya ditampilkan pada Gambar 4:

```

                                clean_punct
0 berada di desa glagaharjo bukit klangon adalah...
1 lava tour merapi jogja adalah aktivitas wisata...
2 obelix hills merupakan salah satu objek wisata...
3 candi prambanan adalah bangunan candi bercora...
4 ledok sambu merupakan salah satu wisata di yog...
5 museum gunung merapi ini menampilkan berbagai ...
    
```

Gambar 4. Hasil *Remove punctuation*

Kemudian *Tokenizing*, teks dipecah menjadi token-token kata melalui proses *Tokenizing*. Hasil dari tahap ini disimpan dalam kolom *tokenized*, berikut hasil nya ditampilkan pada Gambar 5:

```

Hasil tokenizing:
                                tokenized
0 [berada, di, desa, glagaharjo, bukit, klangon,...
1 [lava, tour, merapi, jogja, adalah, aktivitas,...
2 [obelix, hills, merupakan, salah, satu, objek,...
3 [candi, prambanan, adalah, bangunan, candi, be...
4 [ledok, sambu, merupakan, salah, satu, wisata,...
5 [museum, gunung, merapi, ini, menampilkan, ber...
    
```

Gambar 5. Hasil *Tokenized*

Kemudian *Stemming*, setiap token diubah menjadi bentuk dasar kata untuk menyederhanakan analisis, Hasil *Stemming* disimpan dalam kolom *stemmed*, berikut hasil nya diperlihatkan pada Gambar 6:



Hasil stemming:

```

                                stemmed
0  [berada, di, desa, glagaharjo, ,, bukit, klang...
1  [lava, tour, merapi, jogja, adalah, aktivita, ...
2  [obelix, hill, merupakan, salah, satu, objek, ...
3  [candi, prambanan, adalah, bangunan, candi, be...
4  [ledok, sambu, merupakan, salah, satu, wisata,...
5  [museum, gunung, merapi, ini, menampilkan, ber...
  
```

Gambar 6. Hasil Stemmed

Kemudian Transformasi teks ke vector fitur, Teknik TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk mengonversi teks menjadi representasi numerik yang dapat digunakan dalam model pembelajaran mesin. Setelah melewati semua tahapan *Case folding* data, Kolom Deskripsi siap digunakan untuk rekomendasi. Hasil Training Model dan Evaluasi Model memperoleh data training terdiri dari 80% dari total data, sementara data testing mencakup 20% dari total data. Pembagian ini dilakukan menggunakan metode stratified sampling untuk memastikan proporsi kategori yang sama antara data training dan data testing. tabel 1 dibawah ini adalah tabel pembagian data training dan data testing:

Tabel 1. Pembagian data *training* dan data *testing*

Kategori	Data Training	Data Testing
Wisata Sleman	8	2
Wisata Bantul	8	2
Wisata Gunung Kidul	8	2
Wisata Kulon Progo	8	2
total	32	8

Tabel 1 menunjukkan bahwa data training terdiri dari 32 sampel dan data testing terdiri dari 8 sampel, dengan distribusi yang seimbang di setiap kategori wisata. Dari hasil training model didapatkan tabel kontingensi yang menunjukkan jumlah True Positives (TP), False Positives (FP), False Negatives (FN), dan True Negatives (TN) untuk masing-masing kategori. Tabel ini memungkinkan untuk melihat secara detail performa model pada setiap kategori secara individual serta total keseluruhan.. berikut hasil yang didapatkan disajikan pada Gambar 7:

TP, FP, FN, TN per Kategori

Kategori	TP	FP	FN	TN
0 wisata sleman	0.0	2.0	0.0	6.0
1 wisata bantul	1.0	1.0	0.0	6.0
2 wisata kulon progo	2.0	0.0	3.0	3.0
3 wisata gunung kidul	2.0	0.0	0.0	6.0
Total Total	5.0	3.0	3.0	21.0

Gambar 7. Tabel Kontingensi

Untuk menghitung hasil evaluasi model menggunakan tabel kontingensi yang di dapat dari training model. Berikut hasil nya yang ditampilkan dalam Gambar 8:

	Metrik	Nilai Manual	Nilai (Persen) Manual
0	Akurasi	0.8125	81.2%
1	Presiisi	0.6250	62.5%
2	Recall	0.6250	62.5%
3	F1-Score	0.6250	62.5%

Gambar 8. Hasil dari Evaluasi Model

Gambar 8 memperlihatkan bahwa model Naive Bayes Multinomial mencapai keakuratan sebesar 81.2%, presisi 62.5%, recall 62.5%, dan F1-score 62.5% pada data testing. Berikut untuk pengujian manual menggunakan rumus sebagai berikut:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{12}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision}+\text{recall}} \tag{13}$$

TP: Prediksi Positif dan Sebenarnya Positif, FP: Prediksi Positif tetapi Sebenarnya Negatif, FN: Prediksi Negatif tetapi Sebenarnya Positif, TN: Prediksi Negatif dan sebenarnya Negatif.

Total TP, TN, FP, FN = TP total: 0 + 1 + 2 + 2 = 5, FP total: 2 + 1 + 0 + 0 = 3, FN total: 0 + 0 + 3 + 0 = 3, TN total: 6 + 6 + 3 + 6 = 21

$$\text{Accuracy} : \frac{TP+TN}{TP+TN+FP+FN} = \frac{5+21}{5+1+3+3} = \frac{26}{32} = 0.8125 \text{ atau } 81.25\%$$

$$\text{Precision} : \frac{TP}{TP+FP} = \frac{5}{5+3} = \frac{5}{8} = 0.625 \text{ atau } 62.5\%$$

$$\text{Recall} : \frac{TP}{TP+FN} = \frac{5}{5+3} = \frac{5}{8} = 0.625 \text{ atau } 62.5\%$$

$$\text{F1-Score} : 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision}+\text{recall}} = 2 \times \frac{0.625 \times 0.625}{0.625+0.625} = 2 \times \frac{0.390625}{1.25} = 0.625 \text{ atau } 62.5\%$$

Untuk menguji hasil dari sistem rekomendasi yang telah dibuat, penulis menguji dengan menginputkan kata "spot foto" dan "pemandangan alam" untuk kata "spot foto" mendapatkan hasil yang ditampilkan pada Gambar 9 berikut:

Nama Wisata	Deskripsi	Kategori	Alamat	Rating	Fasilitas
Hutan Pinus Mangunan	Sejak 2014, Pinus Mangunan mulai ramai dikunjungi wisatawan. Selain menikmati latar berfoto yang unik berupa jajaran hutan pinus, keserunya sepih sehingga nyaman untuk refreshing. Beberapa spot foto yang bisa wisatawan manfaatkan, di antaranya seperti rumah pohon dan panggung pertunjukan. Ada juga taman bunga di sisi selatan kawasan Pinus Mangunan. Wisatawan bisa menikmati pemandangan ka-ran selatan yang terbuka dan menenangkan sepanjang perbukitan di selatan Yogyakarta.	wisata hutan	Jl raya Jl. Hutan Pinus Ngajir, Mangunan, Kec. Dlingo, Kabupaten Bantul, Daerah Istimewa Yogyakarta	4.6	Toilet, Mula area parkir, panggung makan, taman tematik, tempat istirahat, free wifi, rumah seribu kayu kerdus, panggung taman bunga spot foto, hammock, terasgel Hamack ganda area outdoor
Saribu Batu Sanga Langit	Tempat wisata ini cukup unik. Dulu hanya Indonesia, serang langit berarti penyanga langit. Nama tersebut diberikan karena kemampuannya yang terlihat seolah-olah menyerupai langit, menurut Visiting Jogja. Selain pemandangannya, kerucut lain tempat wisata ini adalah spot foto berupa rumah hobbit, rumah seribu kayu, dan panggung keron.	wisata hutan	Jl. Hutan Pinus Ngajir, Sukorame, Mangunan, Kec. Dlingo, Kabupaten Bantul, Daerah Istimewa Yogyakarta	4.6	toilet, jeep wisata, aula, mushola, warung aneka kuliner, panggung alam, flying fox, camping ground serta spot foto menarik.
Bukit Wisata Palsapayang	Dari puncak bukit, Anda akan melihat Bukit Sereno yang sangat luas dan hijau. Selain itu, Anda juga bisa melihat Gunung Merapi dan Gunung Merbabu yang mendominasi di kejauhan. Daya tarik dari Bukit Wisata Palsapayang yang membuat banyak wisatawan berkunjung adalah spot fotonya yang instagramable. Beberapa spot foto yang paling populer di bukit ini adalah ayunan raksasa, gerbu pandang, dan sumber air.	wisata kulon praga	Soropati RT.007/RW.003, Clapper 2, Hargiliryo, Kec. Kibau, Kabupaten Kulon Progo, Daerah Istimewa Yogyakarta	4.5	mushola, gerabu untuk beristirahat, toilet, serta pendopo amerta
Mudak Sereno	Tidak jauh dari Kalibiryo, Anda bisa mengunjungi Mudak Sereno yang juga menawarkan pengalaman wisata yang menarik. Bukit ini berfungsi sebagai landan air untuk pertanian sekitar. Di sini, Anda dapat menikmati taman bambu air yang memiliki tempat duduk, spot foto, dan rakit tradisional.	wisata kulon praga	Srewo Tengah, Hargaililo, Kec. Kibau, Kabupaten Kulon Progo, Daerah Istimewa Yogyakarta	4.6	Panoraman ikan "setar merah" sebagai ikon produsen utama di bukit ini. Perahu wisata, camping ground, Villa, terasgel Hamack, toilet umum, Bengkel, Rumah jamar.

Gambar 9. Hasil penginputan kata "spot foto"

Dari hasil penginputan kata “spot foto” menampilkan hasil 4 wisata, yaitu hutan pinus mangunan, seribu batu songgo langit, bukit wisata pulepayung, dan waduk sermo. Berikut hasil nya ditampilkan pada Gambar 10:

Evaluasi Berdasarkan Input Pengguna (spot foto):

Akurasi: 50.00%

Presisi: 25.00%

Recall: 50.00%

F1-Score: 33.33%

Gambar 10. Nilai hasil penginputan kata “spot foto”

Dari hasil penginputan kata “spot foto” menunjukkan nilai akurasi 50%, presisi 25%, recall 50% dan f1-score 33.33%. Kemudian untuk hasil penginputan kata “pemandangan alam” mendapatkan hasil yang disajikan pada Gambar 11 berikut:

Nama Wisata	Deskripsi	Kategori	Alamat	Rating	Fasilitas
Pantai Wediombo	Pantai Wediombo merupakan bekas kawah gunung berapi yang tenggelam ke bawah permukaan laut. Bentuk teluk disebut menyerupai setengah lingkaran, serta pantainya menghadap ke arah barat. Jika ingin berkunjung, waktu terbaik adalah pada sore hari, karena kamu akan mendapatkan pemandangan alam menawan dari sunset yang bisa disaksikan dari bibir pantai.	wisata gunung kidul	Jepitu, Girisubo, Gunung Kidul, Yogyakarta.	4.5	Toilet Parkir Luas Mushola Tempat Makan Surfing Camping Oleh-oleh Spot Foto
Wisata Kali Biru	Salah satu destinasi wisata yang paling menarik adalah Kalibiru yang telah meraih popularitas di dalam dan luar negeri. Daya tarik utamanya adalah pemandangan alam yang memesona dan segar. Pohon raksasa yang menjadi daya tarik utama membuat pengunjung untuk berfoto di atasnya dengan latar belakang pegunungan dan waduk Sermo.	wisata kulon progo	Hargowilis, Kokap, kabupaten Kulon Progo	4.4	Gardu Pandang/ Rumah Pohon Spot Flower Spot Panggung Kano Terbang Sky Bike atau Sepeda Langit, Tempat makan (warung kecil) Tempat pertemuan Masjid Tempat parkir Penginapan (home stay)

Gambar 11. Hasil penginputan kata “pemandangan alam”.

Dari hasil penginputan kata “pemandangan alam” menunjukkan hasil rekomendasi wisata yaitu wisata Pantai wediombo dan wisata kalibiru. Untuk nilai akurasi, presisi, recall dan f1-scorenya ditampilkan pada Gambar 12:

Evaluasi Berdasarkan Input Pengguna (pemandangan alam):

Akurasi: 100.00%

Presisi: 100.00%

Recall: 100.00%

F1-Score: 100.00%

Gambar 12. Nilai hasil penginputan kata “pemandangan alam”

Hasil inputan kata “pemandangan alam” yang dilakukan oleh user menunjukkan nilai akurasi 100%, presisi 100%, recall 100% dan f1-score 100%.

Pada studi yang dibahas, terdapat perbedaan serta kemajuan jika dibandingkan dengan studi sebelumnya. yang memakai metode hybrid filtering, yang merupakan perpaduan antara collaborative filtering dan content-based filtering. Pada studi ini Uji alpha yang dilakukan dengan 10 responden memberikan 52,6% jawaban "Baik", sementara uji beta dengan 40 responden menunjukkan 77,3% jawaban "Setuju". Hal ini mengindikasikan sistem yang dibuat mampu mengakomodasi kebutuhan pengguna [9]. Sedangkan pada penelitian saat

ini menggunakan Algoritma Naïve Bayes Multinomial dan menghasilkan sistem rekomendasi wisata yang mendapatkan nilai akurasi tertinggi yaitu 100% yang dilakukan dengan uji penginputan kata “pemandangan alam”. Algoritma Naïve Bayes Multinomial dipilih karena kemampuannya dalam klasifikasi teks [20].

4. SIMPULAN

Berdasarkan temuan dari penelitian yang telah dilakukan, Penelitian ini menggunakan 40 data yang terbagi ke dalam 4 kategori yaitu: Wisata Sleman, Wisata Bantul, Wisata Gunung Kidul, dan Wisata Kulon Progo. Algoritma yang digunakan pada penelitian ini yaitu Algoritma Naive Bayes Multinomial karena kemampuannya dalam klasifikasi teks. Hasil dari evaluasi model memperoleh nilai akurasi sebesar sebesar 81.2%, hal ini memperkuat bahwa model memiliki performa yang baik. Hasil penelitian ini menghasilkan sistem rekomendasi wisata menggunakan Algoritma Naïve Bayes Multinomial ini memperoleh nilai akurasi mencapai 100% yang dilakukan dengan penginputan kata “pemandangan alam”. Dengan demikian hasil penelitian menunjukkan bahwa Algoritma Naïve Bayes Multinomial lebih unggul dalam rekomendasi wisata.

DAFTAR PUSTAKA

- [1] W. Hadi, “Attraction Of The Opak River Flow In The Yogyakarta Region As A Natural And Educational Tourism Destination,” *Jurnal Pariwisata Dan Budaya*, Vol. 11, Pp. 2655–5433, 2020, Doi: 10.31294/Khi.V11i1.7912.
- [2] Badan Pusat Statistik Provinsi Daerah Istimewa Yogyakarta, “Badan Pusat Statistik Provinsi Daerah Istimewa Yogyakarta,” 2024.
- [3] F. Rinjani, “Bukit Paralayang,” 2022.
- [4] D. G. Hutabarat *Et Al.*, “Informasi Artikel Abstrak,” Vol. 5, No. 1, Pp. 1265–1273, 2024, Doi: 10.55338/Jpkmn.V5i1.2890.
- [5] S. Suherman, S. Samsuni, And I. L. Hakim, “Sistem Rekomendasi Wisata Pantai Menggunakan Metode Simple Additive Weighting,” *Ilkom Jurnal Ilmiah*, Vol. 12, No. 1, Pp. 1–7, Apr. 2020, Doi: 10.33096/Ilkom.V12i1.481.1-7.
- [6] I. Putu Gede Hendra Suputra And L. Arida Ayu Rahning Putri, “Sistem Rekomendasi Perjalanan Wisata Berbasis Expert System Dan Algoritma Genetika,” 2022, [Online]. Available: <https://S.Id/Jurnalresistor>
- [7] M. Haris Adiansyah, M. Ahsan, A. E. Budianto, And K. Malang, “Jurnal Terapan Sains & Teknologi Implementasi Metode Weighted Product Sebagai Sistem Rekomendasi Pemilihan Destinasi Wisata Dan Kuliner Favorit Di Malang,” Vol. 2, No. 2, 2020.
- [8] R. F. Alfitroni, U. Chotijah, P. Aisyiyah, And R. Devi, “Sistem Rekomendasi Pemilihan Destinasi Wisata Di Malang Menggunakan Metode Topsis,” *Indexia: Informatic And Computational Intelegent Journal*, Vol. 3, No. 1, Pp. 63–72, 2021.
- [9] A. S. Ningrum, H. C. Rustamaji, And Y. Fauziah, “Content Based Dan Collaborative Filtering Pada Rekomendasi Tujuan Pariwisata Di Daerah Yogyakarta,” 2019. Doi: 10.31315/Telematika.V16i1.3023.
- [10] T. Agustiranti, A. Khalfani Izzati Kurdiana, B. Al Ghiffari, E. Dwi Juniar, And D. Gita Purnama, “Penerapan Naive Bayes Terhadap Sentimen Analisis Media Sosial Twitter Pengguna Kereta Cepat Jakarta-Bandung (Whoosh),” *Jurnal Ilmu Komputer Dan Sistem Informasi (Jikoms)*, Vol. 7, No. 1, Pp. 297–305, 2024.

- [11] Q. Alifa Nanda Prakoso *Et Al.*, “Analisis Sentimen Terhadap Produk Skin Game Di Forum Review Female Daily Menggunakan Metode Multinomial Naïve Bayes Dan Tf-Idf,” *Jurnal Informatik Edisi Ke*, Vol. 18, P. 2022.
- [12] A. Yodi Prayoga, A. Id Hadiana, And F. Rakhmat Umbara, “Deteksi Hoax Pada Berita Online Bahasa Inggris Menggunakan Bernoulli Naïve Bayes Dengan Ekstraksi Fitur Tf-Idf,” *Jurnal Syntax Admiration*, Vol. 2, No. 10, Pp. 1808–1823, Oct. 2021, Doi: 10.46799/Jsa.V2i10.327.
- [13] N. Fibriyanti Arminda, N. Sulistiyowati, And T. Nur Padilah, “Implementasi Algoritma Multinomial Naive Bayes Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Brimo,” 2023.
- [14] I. Wayan *Et Al.*, “Implementasi Metode Tf-Idf Dan Algoritma Naive Bayes Dalam Aplikasi Diabetic Berbasis Android,” *Elektro Dan Ilmu Komputer(Teknik)*, Vol. 3, No. 2, 2023, Doi: 10.55606.
- [15] L. Annisa, A. D. Kalifia, F. Bisnis, D. Humaniora, And U. T. Yogyakarta, “Volume 2 ; Nomor 1,” *Januari*, Pp. 302–307, 2024, Doi: 10.59435/Gjmi.V2i1.249.
- [16] F. Kartika Sari Dewi And T. Purnomo Aji, “Klasifikasi Berita Menggunakan Metode Multinomial Naïve Bayes”.
- [17] L. Mayasari And D. Indarti, “Klasifikasi Topik Tweet Mengenai Covid Menggunakan Metode Multinomial Naïve Bayes Dengan Pembobotan Tf-Idf,” *Jurnal Ilmiah Informatika Komputer*, Vol. 27, No. 1, Pp. 43–53, 2022, Doi: 10.35760/Ik.2022.V27i1.6184.
- [18] Yuyun, Nurul Hidayah, And Supriadi Sahibu, “Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter,” *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, Vol. 5, No. 4, Pp. 820–826, Aug. 2021, Doi: 10.29207/Resti.V5i4.3146.
- [19] M. N. Randhika, J. C. Young, A. Suryadibrata, And H. Mandala, “Implementasi Algoritma Complement Dan Multinomial Naïve Bayes Classifier Pada Klasifikasi Kategori Berita Media Online,” *Ultimatics : Jurnal Teknik Informatika*, Vol. 13, No. 1, 2021.
- [20] I. Budianto, S. N. Anwar, J. T. Lomba, J. Nomor, And K. Semarang, “Analisis Sentiment Pengguna Twitter Mengenai Program Vaksinasi Covid-19 Menggunakan Algoritma Naïve Bayes,” *Jurnal Teknologi Informasi*, Vol. 6, No. 1.