

Penerapan *Content-Based Filtering* dan *K-Nearest Neighbor* pada Sistem Rekomendasi Wisata Probolinggo

DOI: <http://dx.doi.org/10.35889/progresif.v22i1.3508>

Creative Commons License 4.0 (CC BY –NC)



Syamsul Arifin¹, Rina Dewi Indahsari^{2*}

Teknik Informatika, Institut Teknologi Dan Bisnis Asia Malang, Malang, Indonesia

*e-mail Corresponding Author: rinadewi@asia.ac.id

Abstract

Probolinggo Regency has a wide range of tourism destinations; however, the limited dissemination of tourism information makes it difficult for tourists to determine suitable travel destinations. This study aims to develop a tourism destination recommendation system to assist users in obtaining recommendations based on their preferences. The system is developed using the Content-Based Filtering (CBF) method with TF-IDF weighting and cosine similarity, as well as the K-Nearest Neighbor (KNN) algorithm based on cosine distance. The system process includes tourism data collection, text feature extraction and weighting, similarity calculation, and the presentation of main destination recommendations and similar destinations. System evaluation using precision and recall metrics shows that the Content-Based Filtering (CBF) and K-Nearest Neighbor (KNN) methods achieve precision values of 93.75% and 96.00%, respectively, with both methods obtaining a recall value of 100%, indicating that the system is able to provide relevant tourism destination recommendations.

Keywords: Recommendation System; Probolinggo Tourism; Content-Based Filtering; TF-IDF; K-Nearest Neighbor; Cosine Similarity.

Abstrak

Kabupaten Probolinggo memiliki potensi destinasi wisata yang beragam, namun penyebaran informasi wisata yang belum optimal menyulitkan wisatawan dalam menentukan tujuan perjalanan. Penelitian ini bertujuan mengembangkan sistem rekomendasi destinasi wisata untuk membantu pengguna memperoleh rekomendasi sesuai preferensi. Sistem dikembangkan menggunakan metode *Content-Based Filtering* (CBF) dengan pembobotan TF-IDF dan *cosine similarity* serta algoritma *K-Nearest Neighbor* (KNN) berbasis *cosine distance*. Proses sistem meliputi pengumpulan data destinasi wisata, ekstraksi dan pembobotan fitur teks, perhitungan tingkat kemiripan, serta penyajian rekomendasi destinasi utama dan destinasi serupa. Pengujian sistem menggunakan metrik precision dan recall menunjukkan bahwa metode *content-based filtering* (CBF) dan *K-Nearest Neighbor* (KNN) menghasilkan nilai *precision* masing-masing sebesar 93,75% dan 96,00%, dengan nilai *recall* keduanya mencapai 100%, sehingga sistem mampu memberikan rekomendasi destinasi wisata yang relevan.

Kata kunci: Sistem Rekomendasi; Pariwisata Probolinggo; Content-Based Filtering; TF-IDF; K-Nearest Neighbor; Cosine Similarity.

1. Pendahuluan

Sektor pariwisata merupakan salah satu penggerak utama pertumbuhan ekonomi daerah karena berkontribusi terhadap peningkatan pendapatan, penciptaan lapangan kerja, serta pelestarian potensi alam dan budaya lokal [1]. Perkembangan teknologi informasi turut mendorong perubahan dalam pengelolaan dan promosi destinasi wisata, di mana pemanfaatan platform digital seperti sistem e-tourism dan strategi pemasaran digital menjadi semakin penting dalam menjangkau wisatawan secara lebih luas dan efektif [2][3]. Transformasi digital dalam

sektor pariwisata menuntut pengelola destinasi dan pemerintah daerah untuk mampu menyajikan informasi wisata yang cepat, akurat, dan terintegrasi agar promosi destinasi dapat bersaing secara efektif dengan daerah lain [4].

Kabupaten Probolinggo merupakan salah satu wilayah di Provinsi Jawa Timur yang memiliki beragam potensi wisata berbasis alam dan budaya, seperti Gunung Bromo, Pantai Bantar, serta berbagai destinasi wisata religi dan edukatif. Berdasarkan data Badan Pusat Statistik Kabupaten Probolinggo (2024), jumlah kunjungan wisatawan pada tahun 2023 mencapai 892.050 orang dan mengalami peningkatan sebesar 6,7% dibandingkan tahun sebelumnya [5]. Meskipun demikian, pengelolaan informasi destinasi wisata di Kabupaten Probolinggo masih belum terintegrasi secara optimal, sehingga wisatawan cenderung hanya mengunjungi destinasi populer sementara destinasi lain dengan potensi serupa kurang terekspos. Upaya digitalisasi yang dilakukan oleh Dinas Kepemudaan, Olahraga dan Pariwisata (Disporapar) masih berada pada tahap pengembangan dan belum mampu menyediakan sistem yang memberikan rekomendasi destinasi wisata secara personal berbasis karakteristik konten dan preferensi pengguna [6]. Kondisi tersebut menyebabkan proses pemilihan destinasi wisata masih bergantung pada informasi umum dan tingkat popularitas, yang berpotensi menghambat pemerataan kunjungan wisata di Kabupaten Probolinggo.

Dalam konteks pariwisata, sistem rekomendasi telah banyak diterapkan untuk membantu wisatawan menemukan destinasi yang sesuai dengan minat dan kebutuhan mereka [7][8]. Nurpila et al. [9] mengembangkan sistem rekomendasi wisata berbasis web dengan menerapkan metode *Content-Based Filtering* (CBF) yang memanfaatkan riwayat interaksi pengguna untuk merepresentasikan preferensi wisatawan, sehingga rekomendasi yang dihasilkan lebih sesuai dengan minat individu. Selanjutnya, Pradana et al. [10] membandingkan metode *Content-Based Filtering* dan *Collaborative Filtering* pada sistem rekomendasi dan menunjukkan bahwa *Content-Based Filtering* memiliki kinerja yang lebih baik dalam menghasilkan rekomendasi berbasis preferensi pengguna. Hasil penelitian tersebut menegaskan bahwa metode CBF efektif dalam menangkap karakteristik dan minat individu pengguna, sehingga relevan untuk diterapkan pada sistem rekomendasi destinasi wisata yang menuntut personalisasi. Di sisi lain, pendekatan berbasis kemiripan juga banyak digunakan dalam pengembangan sistem rekomendasi. Akhmad et al. [11] mengimplementasikan *K-Nearest Neighbor* Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi. Sementara itu, Wiguna et al. [12] menerapkan algoritma KNN pada sistem rekomendasi game dengan menganalisis kesamaan atribut item yang pernah dipilih pengguna guna menghasilkan rekomendasi yang relevan. Meskipun berbagai penelitian telah menerapkan metode *Content-Based Filtering* dan algoritma *K-Nearest Neighbor* dalam sistem rekomendasi, penelitian yang memanfaatkan kedua pendekatan tersebut secara fungsional dan saling melengkapi dalam satu sistem rekomendasi destinasi wisata masih terbatas, khususnya pada konteks pariwisata daerah. Selain itu, penelitian yang secara spesifik mengkaji penerapan sistem rekomendasi destinasi wisata berbasis preferensi pengguna pada wilayah Kabupaten Probolinggo juga masih sangat terbatas.

Berdasarkan celah penelitian tersebut, penelitian ini mengusulkan pengembangan sistem rekomendasi destinasi wisata di Kabupaten Probolinggo dengan memanfaatkan metode *Content-Based Filtering* (CBF) dan algoritma *K-Nearest Neighbor* (KNN). Metode *Content-Based Filtering* digunakan untuk menghasilkan rekomendasi utama destinasi wisata berdasarkan preferensi pengguna yang direpresentasikan melalui karakteristik konten, sedangkan algoritma *K-Nearest Neighbor* dimanfaatkan untuk menampilkan rekomendasi destinasi wisata serupa ketika pengguna mengakses detail suatu destinasi berdasarkan tingkat kemiripan atribut konten. Pendekatan ini memungkinkan sistem memberikan rekomendasi yang bersifat personal sekaligus kontekstual sesuai dengan interaksi pengguna. Kebaruan penelitian ini terletak pada pemanfaatan CBF sebagai mekanisme rekomendasi berbasis preferensi pengguna dan KNN sebagai rekomendasi lanjutan berbasis kemiripan destinasi pada konteks pariwisata daerah, khususnya Kabupaten Probolinggo.

2. Metodologi

Metodologi penelitian ini menjelaskan tahapan pengolahan data destinasi wisata dan pemodelan sistem rekomendasi menggunakan *Content-Based Filtering* (CBF) dan *K-Nearest Neighbor* (KNN). Tahapan penelitian meliputi pengumpulan data, *preprocessing* dan ekstraksi fitur teks menggunakan TF-IDF, perhitungan kemiripan menggunakan *cosine similarity*, serta

pembentukan rekomendasi berdasarkan pemeringkatan kemiripan. Evaluasi kinerja sistem dilakukan menggunakan metrik *precision* dan *recall* untuk mengukur relevansi rekomendasi.

2.1. Pengumpulan Data Pariwisata Probolinggo

Data penelitian berupa data destinasi wisata Kabupaten Probolinggo yang diperoleh dari berbagai sumber publik, seperti *Google Maps*, situs resmi Disporapar Kabupaten Probolinggo, dan portal pariwisata daring. Pengumpulan data dilakukan pada periode September hingga Oktober 2025. Dataset terdiri atas 35 destinasi wisata yang mencakup kategori alam, budaya, religi, dan edukatif. Struktur dataset yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Struktur Dataset Destinasi Wisata Kabupaten Probolinggo

No	Nama Atribut	Keterangan
1	Nama Destinasi	Identitas objek wisata
2	Kategori	Jenis wisata (alam, budaya, religi, edukatif)
3	Deskripsi	Karakteristik dan informasi destinasi
4	Fasilitas	Fasilitas pendukung destinasi wisata
5	Rating	Informasi penilaian pengguna
6	Ulasan	Komentar atau pengalaman pengunjung
7	Jam Operasional	Informasi waktu kunjungan
8	Gambar	Dokumentasi visual destinasi
9	Koordinat Lokasi	Informasi geografis destinasi

Berdasarkan struktur dataset pada Tabel 1, atribut yang digunakan sebagai fitur utama dalam sistem rekomendasi meliputi nama destinasi, kategori, deskripsi, dan fasilitas, sedangkan atribut lainnya digunakan sebagai informasi pendukung. Seluruh data destinasi wisata telah melalui proses validasi dari minimal dua sumber publik dan digunakan sebagai dataset referensi dalam pemodelan *Content-Based Filtering* dan *K-Nearest Neighbor*. Dataset destinasi wisata digunakan sebagai basis data referensi dalam proses pengukuran kemiripan konten antar item, sehingga penelitian ini tidak melibatkan tahapan pembelajaran model prediktif.

2.2. Preprocessing Data

Tahap preprocessing data merupakan tahap awal yang dilakukan untuk menyiapkan data teks destinasi wisata sebelum proses ekstraksi fitur dan perhitungan kemiripan. Data teks yang diproses berasal dari atribut konten destinasi wisata, yaitu nama destinasi, kategori, deskripsi, dan fasilitas, yang digabungkan menjadi satu dokumen teks untuk setiap destinasi wisata agar representasi konten mencakup seluruh karakteristik destinasi. Selanjutnya dilakukan proses pembersihan teks (*text cleaning*) untuk meningkatkan kualitas representasi data teks. Mengacu pada Pratiwi et al. [7], tahapan pembersihan teks yang dilakukan dalam penelitian ini meliputi:

- 1) *Case Folding* Adalah proses mengubah semua huruf dalam sebuah kalimat menjadi huruf kecil (*lowercase*), sehingga memiliki format penulisan yang seragam.
- 2) *Tokenizing* Yaitu proses membagi teks kalimat atau dokumen menjadi potongan-potongan kata (*tokens*) yang berdiri sendiri.
- 3) *Filtering (Stopword Removal)* Adalah proses menyeleksi kata penting dan menghilangkan kata-kata yang tidak deskriptif seperti tanda baca, kata penghubung, dan kata ganti agar data menjadi lebih padat informasi.

2.3. Ekstraksi Fitur dan Pembobotan TF-IDF

Setelah tahap preprocessing, dokumen teks destinasi wisata direpresentasikan ke dalam bentuk numerik melalui proses ekstraksi fitur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini digunakan untuk mengukur tingkat kepentingan suatu term dalam sebuah dokumen relatif terhadap keseluruhan dokumen dalam dataset [13].

Term Frequency (TF) menunjukkan frekuensi kemunculan suatu term pada sebuah dokumen. Semakin sering suatu term muncul dalam dokumen, maka semakin besar nilai TF yang dimiliki. Perhitungan nilai TF dinyatakan pada Persamaan (1).

$$TF_{t,d} = \frac{f_{t,d}}{\sum f_d} \quad (1)$$

Keterangan:

$TF_{t,d}$ = nilai *Term Frequency term t* pada dokumen d

$f_{t,d}$ = jumlah kemunculan *term t* pada dokumen d

$\sum f_d$ = jumlah seluruh *term* dalam dokumen d

Inverse Document Frequency (IDF) digunakan untuk mengukur tingkat kepentingan suatu *term* terhadap seluruh dokumen dalam dataset [14]. *Term* yang muncul pada banyak dokumen akan memiliki nilai IDF yang lebih kecil. Perhitungan nilai IDF dinyatakan pada Persamaan (2).

$$IDF_t = \log \left(\frac{N}{df_t} \right) \quad (2)$$

Keterangan:

IDF_t = nilai *Inverse Document Frequency term t*

N = jumlah total dokumen dalam dataset

df_t = jumlah dokumen yang mengandung *term t*

Bobot akhir TF-IDF diperoleh dari hasil perkalian nilai *Term Frequency* dan *Inverse Document Frequency*. Perhitungan bobot TF-IDF dinyatakan pada Persamaan (3).

$$TF - IDF_{t,d} = TF_{t,d} \times IDF_t \quad (3)$$

Keterangan:

$TF-IDF_{t,d}$: bobot TF-IDF *term t* pada dokumen d

$TF_{t,d}$: nilai *Term Frequency term t* pada dokumen d

IDF_t : nilai *Inverse Document Frequency term t*

Hasil dari proses pembobotan TF-IDF berupa vektor numerik yang merepresentasikan dokumen teks destinasi wisata. Vektor TF-IDF ini selanjutnya digunakan sebagai dasar dalam perhitungan tingkat kemiripan pada tahap pemodelan *Content-Based Filtering* dan algoritma *K-Nearest Neighbor*.

2.4. Pemodelan *Content-Based Filtering* (CBF)

Metode *Content-Based Filtering* digunakan untuk menghasilkan rekomendasi destinasi wisata berdasarkan tingkat kemiripan konten antara data destinasi wisata dan preferensi pengguna. Dalam proses rekomendasi, sistem menerima input berupa teks pencarian pengguna. Teks inputan tersebut diproses menggunakan tahapan preprocessing dan pembobotan TF-IDF yang sama dengan data destinasi wisata, sehingga baik data destinasi maupun input pengguna berada pada ruang vektor yang sama dan dapat dibandingkan secara langsung.

Tingkat kemiripan antara vektor input pengguna dan vektor destinasi wisata dihitung menggunakan metode *cosine similarity*. Metode ini mengukur kesamaan berdasarkan sudut antara dua vektor dalam ruang vektor berdimensi banyak. Semakin besar nilai *cosine similarity*, maka semakin tinggi tingkat kemiripan konten antara input pengguna dan destinasi wisata [15]. Perhitungan *cosine similarity* dinyatakan pada Persamaan (4).

$$\text{Cosine Similarity}(q, d_i) = \frac{q \cdot d_i}{\|q\| \times \|d_i\|} \quad (4)$$

Keterangan:

q = vektor TF-IDF masukan pengguna

d_i = vektor TF-IDF destinasi wisata ke- i

$q \cdot d_i$ = hasil perkalian dot product antara dua vektor

$\|q\|$ = panjang (magnitudo) vektor masukan pengguna

$\|d_i\|$ = panjang (magnitudo) vektor destinasi wisata ke- i

Nilai *cosine similarity* berada pada rentang 0 hingga 1, di mana nilai yang lebih besar menunjukkan tingkat kemiripan konten yang lebih tinggi. Berdasarkan nilai tersebut, sistem melakukan pemeringkatan destinasi wisata dari nilai kemiripan tertinggi hingga terendah, kemudian menetapkan destinasi dengan nilai tertinggi sebagai rekomendasi utama.

2.5. Pemodelan K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor* (KNN) digunakan untuk menghasilkan rekomendasi destinasi wisata serupa berdasarkan tingkat kemiripan konten antar destinasi. Seluruh data destinasi wisata direpresentasikan dalam bentuk vektor TF-IDF dan digunakan sebagai basis data referensi dalam proses pencarian tetangga terdekat. Ketika pengguna mengakses halaman detail suatu destinasi, vektor TF-IDF destinasi tersebut dijadikan sebagai titik acuan untuk menghitung tingkat kemiripan dengan destinasi wisata lainnya dalam dataset.

Perhitungan jarak pada algoritma KNN dilakukan menggunakan *cosine distance*, yang merupakan transformasi dari *cosine similarity*. Pendekatan ini digunakan untuk mengukur perbedaan arah antar vektor dalam ruang berdimensi tinggi dan sesuai untuk representasi teks berbasis TF-IDF [16]. Dalam implementasinya sebagai metrik jarak, nilai *cosine similarity* ditransformasikan menjadi *cosine distance* agar dapat digunakan dalam mekanisme pencarian tetangga terdekat. Perhitungan *cosine distance* dinyatakan pada Persamaan (5).

$$\text{Cosine Distance}(d_q, d_i) = 1 - \text{Cosine Similarity}(d_q, d_i) \quad (5)$$

Keterangan:

d_q = vektor TF-IDF destinasi wisata yang sedang diakses pengguna

d_i = vektor TF-IDF destinasi wisata ke- i

Cosine Similarity(d_q, d_i) = tingkat kemiripan konten antara dua destinasi wisata

Nilai jarak yang lebih kecil menunjukkan tingkat kemiripan konten yang lebih tinggi. Berdasarkan nilai tersebut, algoritma KNN memilih sejumlah destinasi dengan jarak terkecil sebagai tetangga terdekat, dengan nilai K ditetapkan sebesar 5.

2.6. Evaluasi Kinerja Sistem Rekomendasi

Evaluasi kinerja sistem rekomendasi dilakukan untuk mengukur tingkat relevansi hasil rekomendasi destinasi wisata berdasarkan kesesuaian konten antara destinasi wisata acuan dan hasil rekomendasi. Metode evaluasi yang digunakan dalam penelitian ini adalah *precision* dan *recall*. *Precision* digunakan untuk mengukur proporsi rekomendasi relevan terhadap seluruh rekomendasi yang dihasilkan sistem, sedangkan *recall* digunakan untuk mengukur kemampuan sistem dalam menampilkan seluruh destinasi wisata yang relevan. Perhitungan *precision* dan *recall* dilakukan berdasarkan lima rekomendasi teratas yang ditampilkan oleh sistem untuk setiap destinasi wisata acuan, dengan rumus sebagai berikut:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

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

Keterangan:

TP (*True Positive*) = jumlah destinasi wisata yang direkomendasikan dan dinilai relevan.

FP (*False Positive*) = jumlah destinasi wisata yang direkomendasikan tetapi dinilai tidak relevan.

FN (*False Negative*) = jumlah destinasi wisata yang relevan tetapi tidak direkomendasikan sistem.

Penilaian relevansi dilakukan dengan membandingkan kesesuaian konten antara destinasi wisata acuan dan hasil rekomendasi. Nilai *precision* menunjukkan tingkat ketepatan rekomendasi yang dihasilkan sistem, sedangkan nilai *recall* menunjukkan kemampuan sistem dalam menampilkan seluruh destinasi wisata yang relevan.

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil implementasi sistem rekomendasi destinasi wisata serta analisis kinerja metode yang digunakan. Hasil yang ditampilkan meliputi proses pengolahan data, pembentukan model rekomendasi menggunakan *Content-Based Filtering* (CBF) dan *K-Nearest Neighbor* (KNN), serta evaluasi kinerja sistem berdasarkan tingkat relevansi rekomendasi yang dihasilkan.

3.1 Hasil

1) *Preprocessing* Data Teks

Tahap *preprocessing* menghasilkan data teks destinasi wisata yang telah dibersihkan dan diseragamkan. Teks yang semula mengandung variasi penulisan, tanda baca, dan kata-kata umum diolah menjadi kumpulan kata yang lebih terstruktur dan relevan. Hasil *preprocessing* ditunjukkan pada **Tabel 2**.

Table 2. Hasil <i>Preprocessing</i> Data Teks Destinasi Wisata	
Sebelum	Sesudah
"Air Terjun Madakaripura. Air terjun ini sering disebut juga sebagai Air Terjun Abadi karena debit airnya yang melimpah. Dengan ketinggian sekitar 200 meter, air terjun ini dikelilingi oleh tebing curam. Fasilitas: Area Parkir, Toilet/MCK, Warung Makan, Mushola."	"air terjun madakaripura air terjun abadi debit airnya melimpah ketinggian meter air terjun dikelilingi tebing curam fasilitas area parkir toilet mck warung makan mushola"

Hasil ini menunjukkan bahwa *preprocessing* mampu meningkatkan kualitas data teks sehingga mendukung proses ekstraksi fitur pada tahap selanjutnya.

2) Pembobotan Fitur Menggunakan TF-IDF

Proses pembobotan TF-IDF menghasilkan representasi numerik teks destinasi wisata dalam bentuk vektor. Setiap kata memiliki bobot yang mencerminkan tingkat kepentingannya terhadap suatu destinasi dibandingkan dengan keseluruhan dataset.

Table 3. Hasil Pembobotan TF-IDF Destinasi		
No	Term	Bobot TF-IDF
1	terjun	0.345611
2	air terjun	0.345611
3	air	0.266106
4	tebing	0.115204
...		
92	toilet	0.028636

Berdasarkan hasil pembobotan tersebut, term *terjun* dan *air terjun* memiliki bobot tertinggi sehingga menjadi representasi utama karakteristik destinasi wisata. Hal ini menunjukkan bahwa kedua term tersebut memiliki frekuensi kemunculan yang tinggi serta tingkat keunikan yang lebih dominan dibandingkan term lainnya. Vektor TF-IDF yang dihasilkan selanjutnya digunakan sebagai input pada proses *Content-Based Filtering* dan *K-Nearest Neighbor* (KNN) untuk menghitung tingkat kemiripan antar destinasi wisata dan menghasilkan rekomendasi yang relevan bagi pengguna.

3) Pencarian Rekomendasi Menggunakan *Content-Based Filtering* (CBF)

Pada tahap *Content-Based Filtering* (CBF), sistem menghasilkan rekomendasi destinasi wisata berdasarkan tingkat kemiripan antara masukan pengguna (*query*) dan data destinasi

wisata yang tersimpan dalam dataset. Query pengguna diproses melalui tahapan preprocessing yang sama dengan data destinasi, kemudian direpresentasikan dalam bentuk vektor TF-IDF.

Selanjutnya, sistem menghitung nilai *cosine similarity* antara vektor query pengguna dan vektor TF-IDF setiap destinasi wisata. Nilai kemiripan ini menunjukkan tingkat kedekatan konten antara kebutuhan pengguna dan karakteristik destinasi wisata. Destinasi dengan nilai *cosine similarity* tertinggi dipilih sebagai hasil rekomendasi dan diurutkan berdasarkan tingkat kemiripan.

Table 4. Hasil Perhitungan *Cosine Similarity* dengan query pencarian: "wisata air terjun alam"

Peringkat	Nama Destinasi	Skor Similarity
1	Air Terjun Umbulan	0.4492
2	Air Terjun Madakaripura	0.4195
3	Air Terjun Purba Tirai	0.3762
4	Air Terjun Kali Pedati	0.3515
5	Air Terjun Jaran Goyang	0.2573
6	Kampung Wisata Rabunan	0.1117
7	Bukit Betitang	0.1046
8	Bermi Eco Park	0.1031
9	Sungai Pekalen	0.0942
10	Ranu Segaran	0.0775

Berdasarkan data pada Tabel 4, Air Terjun Umbulan menempati peringkat tertinggi dengan nilai similarity sebesar 0.4492, yang menunjukkan bahwa konten destinasi tersebut paling relevan dengan masukan pengguna.

4) Rekomendasi Wisata Serupa Menggunakan *K-Nearest Neighbor* (KNN)

Pada tahap ini, sistem menghasilkan rekomendasi wisata serupa berdasarkan kemiripan konten antar destinasi menggunakan algoritma *K-Nearest Neighbor* (KNN). Proses ini dilakukan ketika pengguna mengakses halaman detail suatu destinasi wisata. Vektor TF-IDF destinasi yang dipilih pengguna digunakan sebagai titik acuan untuk menghitung jarak terhadap vektor TF-IDF destinasi wisata lainnya dalam dataset.

Pengukuran jarak dilakukan menggunakan *cosine distance*, di mana destinasi dengan nilai jarak terkecil dianggap memiliki tingkat kemiripan tertinggi. Berdasarkan hasil pengujian, nilai K ditetapkan sebesar 5, sehingga sistem menampilkan lima destinasi wisata dengan tingkat kemiripan tertinggi sebagai rekomendasi wisata serupa.

Table 5. Rekomendasi Destinasi Serupa dengan target wisata : "Kampung Wisata Rabunan"

Peringkat	Nama Destinasi	Kategori	Skor Similarity
1	Sungai Pekalen	Sungai & Arung Jeram	0.1551
2	Air Terjun Purba Tirai	Air Terjun	0.0820
3	Bermi Eco Park	Taman & Rekreasi	0.0785
4	Air Terjun Kali Pedati	Air Terjun	0.0675
5	Pantai Bahak	Pantai	0.0655

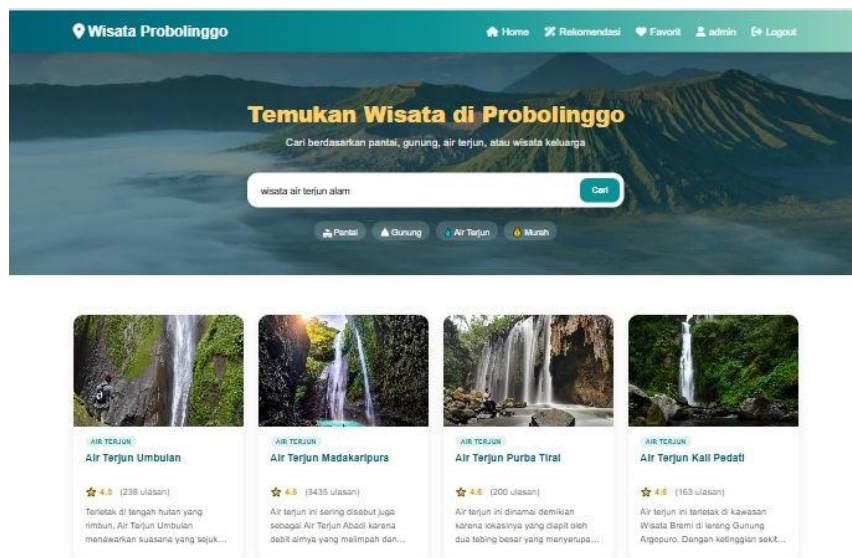
Berdasarkan hasil tersebut, destinasi wisata yang direkomendasikan memiliki karakteristik konten yang serupa dengan destinasi acuan, baik dari sisi deskripsi maupun fasilitas pendukung, sehingga dapat menjadi alternatif pilihan wisata bagi pengguna.

5) Output Sistem Rekomendasi Wisata

Hasil implementasi sistem rekomendasi wisata disajikan melalui antarmuka web yang memungkinkan pengguna berinteraksi langsung dengan sistem. Output sistem ditampilkan dalam dua bagian utama, yaitu halaman pencarian destinasi wisata dan halaman detail destinasi wisata.

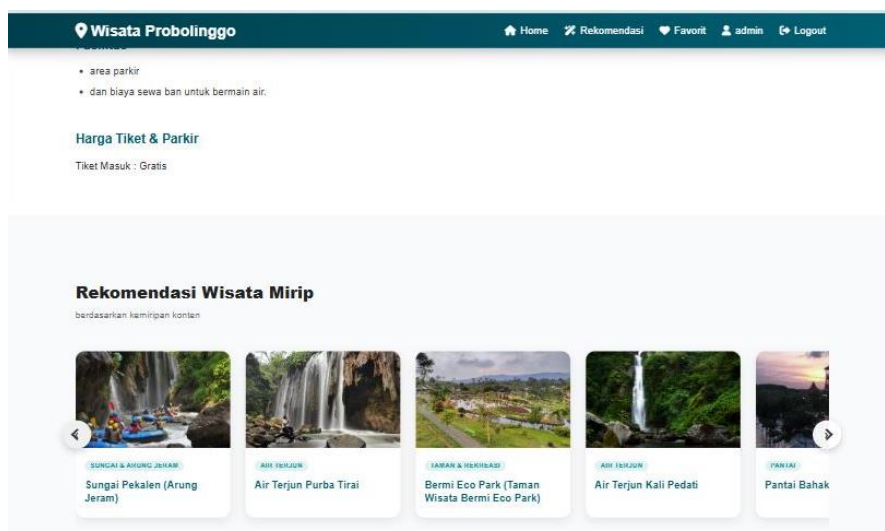
a) Halaman Pencarian

Hasil rekomendasi destinasi wisata berdasarkan input pengguna ditampilkan pada halaman pencarian. Tampilan halaman pencarian destinasi wisata dapat dilihat pada Gambar 1, yang menunjukkan daftar destinasi hasil rekomendasi beserta informasi ringkas masing-masing destinasi.



Gambar 1. Halaman pencarian dan daftar rekomendasi (CBF)

b) Halaman Detail & Rekomendasi Serupa



Gambar 2. Halaman Detail & Rekomendasi Destinasi Serupa (KNN)

Ketika pengguna memilih salah satu destinasi, sistem menampilkan halaman detail destinasi yang dilengkapi dengan rekomendasi wisata serupa. Rekomendasi wisata serupa

ditampilkan pada bagian bawah halaman detail destinasi sebagai alternatif pilihan bagi pengguna, sebagaimana ditunjukkan pada Gambar 2.

3.2. Evaluasi Kinerja Sistem

1) Hasil Evaluasi *Content-Based Filtering* (CBF)

Pengujian metode *Content-Based Filtering* (CBF) dilakukan terhadap lima kata kunci dengan mengevaluasi kesesuaian hasil rekomendasi berdasarkan kategori, deskripsi, dan fasilitas destinasi. Tabel 6 menyajikan sampel rekomendasi yang dikelompokkan menjadi relevan (*True Positive*) dan tidak relevan (*False Positive*).

Table 6. Sampel Hasil Pencarian CBF

Kata Kunci	Relevan	Tidak Relevan
Wisata Pantai	Greenthing Beach & Mangrove forest, Pantai Bentar, Pantai Duta	Kampung Wisata Rabunan
Wisata Pegunungan	Bukit Dami, Danau Ranu Merah, Kampung Wisata Rabunan	-
Air Terjun	Air Terjun Madakaripura, Air Terjun Umbulan, Air Terjun Kali Pedati	-
Wisata Keluarga	Wisata Ronggojalu, Pantai Bentar, Pantai Mayangan	Bukit Betitang
Wisata Alam Terbuka	Ranu Segaran, Kampung Wisata Rabunan, Bermi Eco Park (Taman Wisata Bermi Eco Park)	-

Prosedur evaluasi tersebut diterapkan secara konsisten pada seluruh kata kunci untuk memperoleh nilai *True Positive* (TP), *False Positive* (FP), dan *False Negative* (FN). Rekapitulasi hasil evaluasi disajikan pada Tabel 7.

Table 7. Evaluasi Kinerja Sistem Rekomendasi (CBF)

Kata Kunci	TP	FP	FN	Total
wisata pantai	9	1	0	10
wisata pegunungan	5	0	0	5
air terjun	6	0	0	6
wisata keluarga	6	1	0	7
wisata alam terbuka	4	0	0	4
Total	30	2	0	32

Nilai *precision* dan *recall* sistem dihitung berdasarkan akumulasi keseluruhan hasil pencarian, sehingga merepresentasikan kinerja sistem secara global dengan mengacu pada Persamaan (6) dan (7). Berdasarkan substitusi nilai pada baris total diperoleh:

$$Precision = \frac{30}{30 + 2} = \frac{30}{32} = 93,75\%$$

$$Recall = \frac{30}{30 + 0} = \frac{30}{30} = 100\%$$

Hasil tersebut menunjukkan bahwa metode CBF memiliki tingkat ketepatan rekomendasi yang tinggi dan mampu merekomendasikan seluruh destinasi relevan pada proses pengujian.

2) Hasil Evaluasi *K-Nearest Neighbor* (KNN)

Pengujian algoritma *K-Nearest Neighbor* (KNN) dilakukan menggunakan skema *Precision* dan *Recall* dengan merekomendasikan sejumlah destinasi teratas berdasarkan tingkat kemiripan terhadap destinasi target. Evaluasi dilakukan dengan menilai kesesuaian hasil rekomendasi berdasarkan kategori dan karakteristik konten. Tabel 8 menyajikan sampel rekomendasi yang dikelompokkan menjadi relevan (*True Positive*) dan tidak relevan (*False Positive*).

Table 8. Sampel Hasil Rekomendasi Serupa KNN

Kata Kunci	Relevan	Tidak Relevan
Air Terjun Madakaripura	Air Terjun Umbulan, Air Terjun Kali Pedati, Air Terjun Purba Tirai	-
Gunung Bromo	Bukit Teletubbies Bromo, Puncak Seruni Point, Gunung Argopuro	-
Pantai Bentar	Pantai Duta, BeeJayBakauResort (BJBR)	Papua Waterpark
Kampung Wisata Rabunan	Sungai Pekalen (Arung Jeram), Air Terjun Purba Tirai, Bermi Eco Park	-
Ranu Segaran	Danau Ranu Merah, Danau Ranu Agung, Danau Taman Hidup	-

Prosedur evaluasi yang sama diterapkan pada seluruh destinasi target untuk memperoleh nilai *True Positive* (TP), *False Positive* (FP), dan *False Negative* (FN). Rekapitulasi hasil evaluasi disajikan pada Tabel 9.

Table 9. Evaluasi Kinerja Sistem Rekomendasi (KNN)

Target Destinasi	TP	FP	FN	Total
Air Terjun Madakaripura	5	0	0	5
Gunung Bromo	5	0	0	5
Pantai Bentar	4	1	0	5
Kampung Wisata Rabunan	5	0	0	5
Ranu Segaran	5	0	0	5
Total	24	1	0	25

Nilai *precision* dan *recall* sistem dihitung berdasarkan akumulasi keseluruhan hasil rekomendasi dengan mengacu pada Persamaan (6) dan (7). Berdasarkan substitusi nilai pada baris total diperoleh:

$$Precision = \frac{24}{24 + 1} = \frac{24}{25} = 96,00\%$$

$$Recall = \frac{24}{24 + 0} = \frac{24}{24} = 100\%$$

Hasil tersebut menunjukkan bahwa KNN mampu menghasilkan rekomendasi dengan tingkat ketepatan tinggi serta mencakup seluruh destinasi relevan pada skenario pengujian.

3.3. Pembahasan

Hasil evaluasi menunjukkan bahwa metode *Content-Based Filtering* (CBF) memperoleh nilai precision sebesar 93,75%, sedangkan *K-Nearest Neighbor* (KNN) menghasilkan precision sebesar 96,00%, dengan nilai recall keduanya mencapai 100%. Capaian ini menunjukkan bahwa pembobotan TF-IDF dan perhitungan *cosine similarity* maupun *cosine distance* mampu merepresentasikan karakteristik destinasi secara efektif serta mengukur kedekatan antar destinasi secara konsisten. Kombinasi tersebut memungkinkan sistem menyaring dan mengurutkan destinasi yang paling sesuai dengan kata kunci maupun destinasi target sehingga rekomendasi yang dihasilkan tetap relevan.

Temuan ini sejalan dengan penelitian Nurpila et al. [9] dan Pradana et al. [10] yang menyatakan bahwa pendekatan *Content-Based Filtering* efektif dalam menghasilkan rekomendasi berdasarkan kesamaan atribut konten. Dalam penelitian ini, efektivitas tersebut terlihat dari kemampuan sistem mempertahankan tingkat ketepatan yang tinggi pada berbagai kata kunci pengujian. Hal ini menunjukkan bahwa pencocokan berbasis karakteristik deskripsi dan fasilitas destinasi masih menjadi pendekatan yang relevan dalam konteks sistem rekomendasi wisata. Dengan demikian, penelitian ini tidak hanya mengonfirmasi temuan terdahulu, tetapi juga memperkuat bukti bahwa metode CBF tetap efektif diterapkan pada pengelolaan informasi destinasi wisata daerah. Selain itu, hasil evaluasi metode *K-Nearest Neighbor* (KNN) juga mendukung penelitian Dharmawan et al. [11] dan Wiguna et al. [12] yang menunjukkan bahwa algoritma KNN efektif dalam menemukan item dengan tingkat kemiripan atribut yang tinggi. Dalam penelitian ini, nilai precision KNN yang tinggi menunjukkan bahwa pendekatan berbasis kedekatan antar destinasi mampu mempertahankan konsistensi relevansi rekomendasi. Jika CBF berperan dalam memproses pencarian berbasis kata kunci, maka KNN berfungsi dalam mengidentifikasi destinasi yang memiliki karakteristik mendekati destinasi target. Dengan demikian, kedua metode tidak bersifat saling menggantikan, melainkan saling melengkapi dalam sistem rekomendasi yang dikembangkan, sehingga memperkuat kualitas hasil rekomendasi secara keseluruhan.

Secara konseptual, penelitian ini memberikan kontribusi pada pengembangan sistem rekomendasi berbasis konten dengan menunjukkan bahwa integrasi CBF dan KNN dalam satu sistem dapat meningkatkan fleksibilitas mekanisme rekomendasi. Selama ini, banyak penelitian menerapkan kedua metode tersebut secara terpisah, baik hanya menggunakan CBF untuk pencarian berbasis preferensi maupun KNN untuk pencarian item serupa. Penelitian ini menunjukkan bahwa penggabungan keduanya memungkinkan sistem melayani dua kebutuhan sekaligus, yaitu pencarian berdasarkan kata kunci dan eksplorasi destinasi yang memiliki karakteristik mendekati destinasi tertentu. Dengan demikian, penelitian ini memberikan model implementasi yang lebih komprehensif dalam pengembangan sistem rekomendasi destinasi wisata.

Dalam pengembangan sistem rekomendasi, penelitian ini menegaskan bahwa kualitas sistem rekomendasi tidak hanya ditentukan oleh pendekatan yang digunakan, tetapi juga oleh representasi fitur dan ukuran kemiripan. Pembobotan TF-IDF yang dikombinasikan dengan *cosine similarity* dan *cosine distance* terbukti mampu mendukung proses rekomendasi secara efektif, khususnya pada kondisi keterbatasan data interaksi pengguna. Namun demikian, penelitian ini masih memiliki keterbatasan pada jumlah dataset serta evaluasi yang belum melibatkan pengguna secara langsung. Oleh karena itu, penelitian selanjutnya dapat menambahkan data interaksi pengguna, memperluas jumlah destinasi, serta melakukan pengujian pada skenario yang lebih beragam untuk meningkatkan generalisasi sistem.

4. Simpulan

Sistem rekomendasi wisata pada penelitian ini memanfaatkan metode *Content-Based Filtering* (CBF) menggunakan pembobotan TF-IDF dan *cosine similarity* serta algoritma *K Nearest Neighbor* (KNN) berbasis *cosine distance*. Hasil evaluasi menunjukkan bahwa metode CBF menghasilkan nilai precision rata-rata sebesar 93,75%, sedangkan KNN menghasilkan *precision* rata-rata sebesar 96,00%, dengan nilai *recall* keduanya mencapai 100%. Nilai tersebut menunjukkan bahwa representasi fitur teks menggunakan TF-IDF mampu menggambarkan karakteristik destinasi wisata secara efektif sehingga sistem dapat menghasilkan rekomendasi yang relevan dengan preferensi pengguna.

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan metode CBF dan KNN dalam satu sistem rekomendasi mampu menghasilkan sistem

rekomendasi destinasi wisata yang efektif dalam memberikan rekomendasi sesuai preferensi pengguna serta kemiripan karakteristik destinasi. Sistem yang dikembangkan berpotensi mendukung promosi destinasi wisata daerah dan membantu pemerataan kunjungan wisatawan melalui penyajian rekomendasi destinasi yang lebih personal dan relevan.

Daftar Referensi

- [1] C. Salsabilla and D. W. Utomo, "Sistem Rekomendasi Wisata Kabupaten Pekalongan dengan Content Based Filtering," *Sistemasi*, vol. 14, no. 1, pp. 262 - 270, 2025, doi: 10.32520/stmsi.v14i1.4839.
- [2] R. Komalasari, P. Pramesti, and B. Harto, "Teknologi Informasi E-Tourism Sebagai Strategi Digital Marketing Pariwisata," *Altasia: Jurnal Pariwisata Indonesia*, vol. 2, no. 2, pp. 163–170, 2020, doi: 10.37253/altasia.v2i2.559.
- [3] D. Dharmawan, L. K. C. Dewi, H. Purnomo, M. Munizu, and W. Nas, "Analysis of the influence of digital marketing, information technology adoption, and use of social media on the effectiveness of tourism promotion," *Jurnal Sistim Informasi dan Teknologi*, vol. 6, no. 2, pp. 116–120, 2024, doi: 10.60083/jsisfotek.v6i2.374.
- [4] Z. W. Mauliah, H. Yeni, "Implementasi sistem informasi manajemen pariwisata dalam promosi destinasi wisata Kabupaten Mamuju," *JUKONI*, vol. 3, no. 1, pp. 26–31, 2026, doi: 10.70134/jukoni.v3i1.1129.
- [5] B. P. S. K. Probolinggo, "Jumlah Wisatawan Mancanegara dan Domestik di Kabupaten Probolinggo (2016-2023)," Badan Pusat Statistik Kabupaten Probolinggo, Probolinggo, Indonesia, 2024. [Online]. Available: <https://probolinggokab.bps.go.id/id/statistics-table/1/MTM5NyMx/jumlah-wisatawan-mancanegara-dan-domestik-di-kabupaten-probolinggo--2016-2023.html>.
- [6] L. Update, "Disporapar Kabupaten Probolinggo Perkuat Digitalisasi 36 Desa Wisata Lewat Pelatihan DEWI SAE," *Lensa-Update*, Indonesia, 2025. [Online]. Available: https://www.lensa-update.com/2025/11/disporapar-kabupaten-probolinggo_25.html.
- [7] D. Pratiwi, A. Asrianda, and L. Rosnita, "Penerapan Metode *Content-Based Filtering* dalam Sistem Rekomendasi Objek Wisata di Aceh Tamiang," *J. Ilmu Komput. dan Inform.*, vol. 4, no. 2, pp. 85–96, 2025, doi: 10.54082/jiki.169.
- [8] M. Islamiyah, P. Subekti, and T. D. Andini, "Pemanfaatan Metode Item Based Collaborative Filtering Untuk Rekomendasi Wisata Di Kabupaten Malang," *J. Ilm. Teknol. Inf. Asia*, vol. 13, no. 2, pp. 143–150, 2019, doi: 10.32815/jitika.v13i2.70.
- [9] S. Nurpila, H. A. Ismail, and S. A. Prakoso, "Rancang Bangun Rekomendasi Tempat Wisata Di Kabupaten Rembang Berbasis Website Menggunakan Metode Content Based Filtering," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 7, no. 4, pp. 1149-1158, 2024, doi: 10.31539/intecom.v7i4.10718.
- [10] D. S. Pradana, P. Prajoko, and G. P. Hartawan, "Perbandingan Algoritma Content-Based Filtering dan Collaborative Filtering dalam Rekomendasi Kegiatan Ekstrakurikuler Siswa," *Progresif Jurnal Ilmiah Komputet*, vol. 18, no. 2, pp. 151-160, 2022, doi: 10.35889/progresif.v18i2.854.
- [11] M.R. Akhmad, & T.A.Y Siswa, "Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi. *Progresif: Jurnal Ilmiah Komputer*, vol. 18, no. 2, pp. 185-192, 2022. doi:<http://dx.doi.org/10.35889/progresif.v18i2.921>.
- [12] I Putu Marcel Wgunaa and Ida Bagus Gede Dwidasmara, "Sistem Rekomendasi Game dengan Metode K-Nearest Neighbor (KNN)," *J. Nas. Teknol. Inf. dan Apl.*, vol. 2, no. 1, pp. 245–250, 2023, doi: 10.24843/JNATIA.2023.v02.i01.p28.
- [13] R. Al Rasyid and D. H. U. Ningsih, "Penerapan Algoritma TF-IDF dan Cosine Similarity untuk Query Pencarian Pada Dataset Destinasi Wisata", *J. Teknol. Inf. Komun*, vol. 8, no. 1, pp. 170–178, Jan. 2024, doi: 10.35870/jtik.v8i1.1416.
- [14] A. O. Rahmawati, R. Susanto, and H. Hasanah, "Sistem Rekomendasi Bursa Kerja Khusus (Bkk)Menggunakan Metode Content Based Filteringpada SMK Tunas Bangsa", *JINTEKS*, vol. 7, no. 3, pp. 1502–1510, Aug. 2025, doi: 10.51401/jinteks.v7i3.6113.
- [15] S. Oyadila, D. Abdullah, and A. Razi, "Implementasi Content-Based Filtering Dengan Tf-Idf Dan Cosine Similarity Untuk Sistem Rekomendasi Destinasi Wisata Di Aceh Tengah," *rabit*, vol. 10, no. 2, pp. 1329–1339, 2025, doi: 10.36341/rabit.v10i2.6532.
- [16] R. B. Trianto, A. Triyono, and D. M. P. Arum, "Kombinasi Metode K-Nearest Neighbor

- dengan Cosine Similarity untuk Prediksi Serangan Firewall pada Jaringan Komputer," *Jurnal Informatika Universitas Pamulang*, vol. 6, no. 4, 2021, pp. 672-679, doi:10.32493/informatika.v6i4.12680.
- [17] M. Firmansyah, A. Aziz, dan M. Ahsan, "Peningkatan kinerja sistem rekomendasi wisata melalui penerapan algoritma collaborative filtering dan K-nearest neighbors dengan metode klasterisasi K-means," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, pp. 11420–11427, 11 Nov. 2024, doi: 10.36040/jati.v8i6.11488.