

## Penerapan Metode *Content-Based Filtering* dalam Sistem Rekomendasi Objek Wisata di Aceh Tamiang

Dinda Pratiwi<sup>\*1</sup>, Asrianda<sup>2</sup>, Lidya Rosnita<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Fakultas Teknik, Universitas Malikussaleh, Indonesia

Email: <sup>1</sup>dinda.200170142@mhs.unimal.ac.id, <sup>2</sup>asrianda@unimal.ac.id, <sup>3</sup>lidyarosnita@unimal.ac.id

### Abstrak

Pariwisata berbasis teknologi informasi menjadi semakin penting dalam era digital untuk mempermudah wisatawan menemukan destinasi yang sesuai dengan preferensi mereka. Penelitian ini bertujuan mengembangkan sistem rekomendasi objek wisata berbasis *web* di Aceh Tamiang menggunakan metode *content-based filtering*. Sistem ini dirancang untuk menganalisis ulasan dari Google Maps agar memberikan rekomendasi yang relevan dengan preferensi pengguna. Data ulasan dikumpulkan melalui SerpApi, kemudian diproses melalui tahapan *preprocessing*, perhitungan bobot TF, IDF, dan TF-IDF, serta analisis kesamaan menggunakan *cosine similarity*. Hasil penelitian menunjukkan bahwa sistem ini mampu memberikan rekomendasi yang sesuai dengan preferensi wisatawan, khususnya untuk atribut seperti "pantai", "piknik", "sepi", dan "sungai". Sistem ini diharapkan dapat membantu pengambilan keputusan wisatawan secara lebih efisien dan memberikan kontribusi positif terhadap pengembangan pariwisata di Aceh Tamiang.

**Kata Kunci:** *Content-Based Filtering, Cosine Similarity, Rekomendasi Wisata, TF-IDF*

### Abstract

Information technology-based tourism is becoming increasingly important in the digital era to assist travelers in finding destinations that match their preferences. This study aims to develop a web-based tourist attraction recommendation system in Aceh Tamiang using the content-based filtering method. The system is designed to analyze reviews from Google Maps to provide recommendations relevant to user preferences. Review data was collected through SerpApi and processed through stages of preprocessing, TF, IDF, and TF-IDF weighting calculations, as well as similarity analysis using cosine similarity. The results indicate that this system can provide recommendations that align with travelers' preferences, particularly for attributes such as "beach," "picnic," "quiet," and "river." This system is expected to help travelers make decisions more efficiently and contribute positively to the development of tourism in Aceh Tamiang.

**Keywords:** *Content-Based Filtering, Cosine Similarity, Tourism Recommendations, TF-IDF*

## 1. PENDAHULUAN

Kabupaten Aceh Tamiang, yang terbentuk dari pemekaran Kabupaten Aceh Timur, terletak di perbatasan antara Aceh dan Sumatera Utara. Wisata bahari dan pesisir di daerah ini berkembang sangat pesat. Potensi yang melimpah menjadikan objek wisata tersebut bisa dijadikan produk unggulan daerah [1]. Namun, pariwisata di Aceh Tamiang masih menghadapi beberapa kendala utama, seperti keterbatasan informasi di *platform online* yang sering kali seragam, sulit diakses, dan tidak mencerminkan preferensi wisatawan secara individual. Akibatnya, banyak objek wisata yang belum terdokumentasi dengan baik atau dikenal secara luas.

Selain itu, wisatawan sering mengalami *information overload*, yaitu kesulitan memilih destinasi dari banyaknya informasi yang tersedia [2]. Kondisi ini menuntut adanya sistem yang mampu menyaring informasi dan memberikan rekomendasi personal sesuai preferensi pengguna. Sistem rekomendasi menjadi salah satu solusi untuk mengatasi permasalahan tersebut dengan membantu wisatawan menemukan destinasi yang relevan dan memperkenalkan objek wisata yang kurang dikenal agar mendapatkan perhatian lebih luas.

Sistem rekomendasi adalah alat untuk membantu pengguna dalam pengambilan keputusan, sistem rekomendasi telah diterapkan dalam berbagai bidang, seperti layanan *streaming* (Netflix), *e-commerce* (Amazon), dan pariwisata [3].

Ada tiga pendekatan utama yang tersedia untuk sistem rekomendasi: *content-based filtering*, *collaborative filtering*, dan *hybrid recommender*. Metode *content-based filtering* merekomendasikan *item* berdasarkan kesamaan dengan *item* yang dipilih [4], *collaborative filtering* menawarkan rekomendasi berdasarkan preferensi pengguna lain [5]. Sedangkan *hybrid recommender* adalah kombinasi dari kedua metode tersebut.

Penelitian sebelumnya menunjukkan bahwa metode *collaborative filtering* sering menghadapi masalah *cold start*, yaitu sulit memberikan rekomendasi bagi pengguna baru yang belum memiliki riwayat aktivitas[6]. Sebagai contoh, penelitian “Pengembangan Aplikasi Sistem Rekomendasi Tempat Wisata dengan *Collaborative Filtering*” menunjukkan akurasi sistem rekomendasi sebesar 91,60% menggunakan Mean Absolute Error (MAE) [7], namun tidak membahas solusi atas masalah *cold start*. Di sisi lain, *content-based filtering* menawarkan keunggulan dalam waktu eksekusi yang lebih cepat, sebagaimana dijelaskan dalam penelitian “Penerapan Metode *Content-Based Filtering* pada Sistem Rekomendasi” [8].

Studi lain, seperti “*Tour Spot Recommendation System via Content-Based Filtering*,” berhasil memanfaatkan ulasan pengguna untuk membantu wisatawan menemukan destinasi sesuai anggaran dan minat, sehingga meningkatkan pengalaman pengguna [9]. Penelitian tersebut berfokus pada rekomendasi destinasi wisata di berbagai negara dan wilayah berdasarkan dataset yang dirancang oleh penulis. Meskipun demikian, pendekatan ini lebih diarahkan pada pengelompokan destinasi secara umum tanpa mengeksplorasi konteks spesifik suatu wilayah tertentu, seperti Aceh Tamiang. Penelitian ini diharapkan dapat mengisi celah tersebut dengan menerapkan pendekatan serupa dalam konteks lokal, menggabungkan analisis ulasan berbasis TF-IDF dan *cosine similarity* untuk menghasilkan rekomendasi yang lebih personal dan relevan. Selain itu, metode *content-based filtering* memiliki keuntungan yang bersifat *user independence*, yang berarti tidak tergantung pada apakah objek itu baru atau telah digunakan oleh orang lain [5].

Penelitian ini bertujuan mengembangkan sistem rekomendasi wisata di Aceh Tamiang menggunakan metode *content-based filtering* dengan memanfaatkan ulasan Google Maps. Sistem ini menganalisis ulasan menggunakan TF-IDF dan *cosine similarity* untuk memberikan rekomendasi personal sesuai preferensi wisatawan, seperti pantai, sungai, dan sepi. Pendekatan ini diharapkan mengatasi masalah *cold start*, memperkenalkan destinasi kurang dikenal, serta meningkatkan keberagaman informasi wisata. Hasil penelitian diharapkan membantu wisatawan menemukan destinasi sesuai preferensi sekaligus mendukung pengembangan pariwisata lokal di Aceh Tamiang.

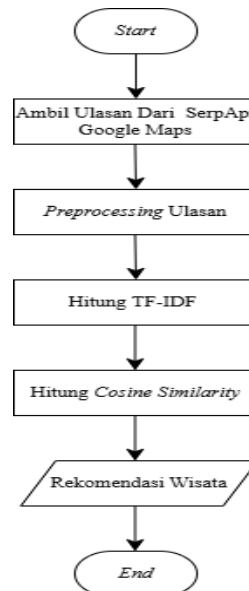
## 2. METODE PENELITIAN

Data ulasan wisata dikumpulkan secara efisien dengan menggunakan API SerpApi, yang memungkinkan akses *real-time* ke informasi dari Google Maps. Penulis memanfaatkan pustaka Python seperti *serpapi* untuk mengakses data ulasan yang relevan berdasarkan lokasi dan jenis tempat wisata. Setelah mendapatkan data, penulis memproses ulasan menggunakan teknik *preprocessing* dengan bantuan pustaka NLTK untuk tokenisasi, penghapusan *stopword*, serta *stemming* menggunakan Sastrawi. Semua proses ini memungkinkan penulis untuk menganalisis dan membangun sistem rekomendasi wisata yang lebih akurat dan relevan.

Berikut adalah tahapan dalam pengumpulan data ulasan:

- Persiapan API: Penulis mendaftar di *platform* SerpApi untuk mendapatkan akses token API yang diperlukan untuk mengakses data ulasan dari Google Maps.
- Pengambilan Data: Menggunakan API SerpApi, penulis melakukan *query* untuk mendapatkan informasi tentang objek wisata yang relevan. Data yang diambil meliputi nama tempat dan ulasan, yang dapat digunakan untuk analisis.
- Proses Pengolahan Data: Setelah data ulasan diperoleh, data tersebut akan diproses menggunakan teknik *preprocessing*.

Misalnya, dalam rekomendasi wisata, jika seorang pengguna memilih atribut "piknik", "pantai", "sungai", dan "sepi", sistem akan mencari ulasan tempat wisata yang memiliki atribut yang sama dan merekomendasikannya. Dengan pendekatan ini, sistem dapat memberikan rekomendasi yang lebih relevan, dan dapat lebih efektif dalam menyesuaikan preferensi pengguna. Prosedur rekomendasi dilakukan dengan mengambil ulasan dari tahun 2022-2024, kemudian melakukan *text preprocessing*, pembobotan TF-IDF, dan perhitungan *cosine similarity*, seperti gambar 1 dibawah ini.



Gambar 1. Flowchart Metode Content-Based Filtering

## 2.1. Text Preprocessing

Tahapan *text preprocessing* dilakukan untuk membersihkan kalimat ulasan dari elemen-elemen yang tidak relevan seperti simbol, angka, dan kata-kata yang kurang signifikan serta berguna untuk mengoptimalkan representasi data yang mendukung sistem rekomendasi agar lebih akurat dan efektif. Proses *preprocessing* melibatkan beberapa tahapan, yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*.

- Case folding* adalah proses mengubah semua huruf dalam sebuah kalimat menjadi huruf kecil, sehingga memiliki penulisan format yang sama [10]. Contohnya, kalimat "Pantai ini sangat Indah!" akan diubah menjadi "pantai ini sangat indah!".
- Tokenizing* yaitu proses membagi teks, seperti kalimat, paragraf, atau dokumen, menjadi token-token atau bagian-bagian yang lebih kecil dan lebih khusus. Misalnya, kalimat "Pantai ini sangat indah" akan diubah menjadi token-token: ["Pantai", "ini", "sangat", "indah"].
- Filtering (stopword removal)* adalah proses memilih kata penting untuk dokumen (ulasan wisata) dan menghilangkan kata-kata yang tidak deskriptif, seperti tanda baca, kata penghubung, dan kata ganti sehingga ulasan wisata menjadi lebih padat informasi [11]. Misalnya, kalimat "Pantai ini sangat indah dan bersih" setelah *filtering* akan menjadi "Pantai indah bersih", menghapus kata "ini" dan "dan" yang tidak deskriptif.
- Stemming* adalah proses untuk mendapatkan kata dasar dari kata-kata yang telah melalui tahap penyaringan. Proses ini menghapus imbuhan, termasuk awalan (*prefix*), akhiran (*suffix*), dan sisipan (*infix*), dan gabungan antara awalan dan akhiran (*confix*) guna menghindari redundansi dalam analisis [12]. Misalnya, kata "bermain" dan "bermainan" akan diproses menjadi "main", menghilangkan imbuhan seperti "ber-" dan "-an" untuk mengurangi redundansi dalam analisis.

## 2.2. Hitung TF-IDF

TF-IDF adalah metode pembobotan kata yang melibatkan proses tokenisasi, penghapusan kata-kata umum (*stopwords*), dan *stemming*. Frekuensi kemunculan kata dalam sebuah dokumen menunjukkan seberapa penting kata tersebut dalam dokumen tersebut. Metode ini mengombinasikan nilai TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*) untuk memberikan bobot pada kata-kata berdasarkan hasil *query*. TF menunjukkan frekuensi kemunculan suatu kata dalam dokumen tertentu, yang dihitung dengan membandingkan jumlah kemunculan kata tersebut dengan total jumlah kata dalam dokumen. IDF menekankan pentingnya sebuah kata dalam keseluruhan koleksi dokumen, di mana kata yang jarang muncul akan memiliki nilai IDF lebih tinggi. IDF dihitung dengan membagi total jumlah dokumen dalam koleksi dengan jumlah dokumen yang mengandung kata tersebut [13].

Penggunaan TF-IDF mempermudah dalam mengenali kata kunci yang memiliki relevansi lebih tinggi dalam suatu dokumen dibandingkan dengan dokumen lainnya dalam kumpulan data. Oleh karena itu, TF-IDF sering diterapkan dalam berbagai aplikasi seperti pencarian informasi dan analisis teks untuk meningkatkan ketepatan dalam menemukan informasi yang sesuai. Dengan demikian, dapat disimpulkan bahwa TF-IDF sangat efektif dalam mengukur kesamaan antara kata kunci dan dokumen.

Persamaan (1) digunakan melakukan penghitungan *term frequency*:

$$tf_{i,j} = \frac{f_{i,j}}{d_j} \quad (1)$$

Keterangan:

$tf_{i,j}$  = *term frequency* dari kata I yang terdapat dalam dokumen j

$f_{i,j}$  = jumlah kemunculan kata i dalam dokumen j

$d_j$  = total keseluruhan kata pada dokumen j

Persamaan (1) dapat membantu untuk memahami seberapa signifikan kata tersebut dalam konteks *document* yang sedang dianalisis, sedangkan, *Inverse Document Frequency* (IDF) memberikan bobot yang lebih besar pada kata-kata yang jarang muncul pada seluruh koleksi *document*, dengan formula:

$$idf_i = \log\left(\frac{n+1}{df_i+1}\right) + 1 \quad (2)$$

Persamaan (2) digunakan untuk menghitung IDF.

Keterangan:

$idf_i$  = *inverse document frequency*

$df_i$  = banyaknya dokumen yang mengandung kata i

$n$  = total dokumen  $df_i$

TF-IDF menggabungkan kedua metrik ini untuk memberikan bobot pada setiap *term* dalam *document*, dengan tujuan meningkatkan akurasi dengan mengurangi pengaruh kata-kata yang sering muncul namun tidak terlalu signifikan. Algoritma TF-IDF sangat berguna untuk berbagai aplikasi seperti mesin pencari, pengelompokan dokumen, dan sistem rekomendasi, karena dapat menangkap esensi penting dari konten yang dianalisis. Persamaan (3) digunakan untuk menghitung bobot *term frequency – inverse document frequency*.

$$w = tf \times idf \quad (3)$$

Keterangan:

$w$  = bobot TF-IDF

$tf$  = *term frequency*

$idf$  = *inverse document frequency*

### 2.3. Cosine Similarity

*Cosine similarity* merupakan metode umum yang digunakan untuk mengukur kesamaan. Metode ini berperan sebagai dasar untuk menghitung tingkat relevansi antara *query* dan dokumen, serta di antara berbagai dokumen lainnya. *Cosine similarity* dapat dihitung berdasarkan sudut  $\theta$  yang terbentuk antara vektor-vektor [14]. Dengan menggunakan TF-IDF (*Term Frequency - Inverse Document Frequency*), yang memberikan bobot pada setiap *term* dalam dokumen, proses perhitungan *cosine similarity* menjadi lebih akurat dan dapat lebih efektif digunakan untuk menentukan kemiripan antara preferensi pengguna (*query*) dengan vektor atribut objek wisata (*document*) [15]. Algoritma *cosine similarity* sangat berguna ketika membuat sistem rekomendasi karena dapat menghitung sejauh mana dua dokumen memiliki kesamaan dalam konteks kata-kata yang penting, memungkinkan untuk rekomendasinya yang lebih relevan dan personal. Sebagaimana dijelaskan dalam persamaan (4). Dengan pendekatan ini, sistem dapat menilai dengan akurat kemiripan antara *query* dengan *document*, sehingga memberikan rekomendasi yang sesuai dengan preferensi pengguna.

$$\text{sim}(q, d_j) = \frac{q \cdot d_j}{|q| \cdot |d_j|} \quad (4)$$

Keterangan:

$\text{sim}(q, d_j)$  = *similarity* antara vektor *query* dan vektor dokumen ke-j

$q$  = vektor *query*

$d_j$  = vektor dokumen ke-j

$|q|$  = norma *euclidean* dari vektor *query*

$|d_j|$  = norma *euclidean* vektor dokumen ke-j

## 3. HASIL DAN PEMBAHASAN

Di Aceh Tamiang terdapat 45 objek wisata yang terdaftar di Google Maps, mencakup berbagai jenis destinasi yang menawarkan keindahan alam, sejarah, aktivitas rekreasi yang menarik bagi para wisatawan, dan lain-lainnya. Tabel 1 menunjukkan daftar wisata yang terdaftar di Google Maps.

Tabel 1. Daftar Wisata di Aceh Tamiang

Doc	Nama Wisata
D1	Aceh Tamiang <i>Recreation Park</i>
D2	Air Terjun Gunung Pandan
D3	Air Terjun Keramat
D4	Air Terjun Sangkapane
D5	Air Terjun Tamsar 27
D6	Batu Bukit Kapur Kaloy
D7	Hutan Kota Aceh Tamiang
D8	Hutan Pribadi Tamiang <i>Camp</i>
D9	Istana Benua Raja
D10	Istana Kerajaan Karang
.....	.....
D41	TPI Tanjung Keramat
D42	Wisata Dam Sungai Rengas
D43	Wisata Pantai Kaloy
D44	Wisata Sungai Bampo
D45	Wisata Pendidikan Mangrove Ujung Tamiang

### 3.1. Preprocessing

Setelah melakukan *preprocessing*, penulis telah menyaring ulasan wisata yang mengandung atribut yang telah dipilih dan mendapatkan hasil 23 wisata dari jumlah keseluruhannya yaitu sebanyak 45 wisata. Contoh penerapan *text preprocessing* seperti Tabel 2.

Tabel 2. Hasil *Preprocessing*

Doc	Sebelum <i>Preprocessing</i>	Setelah <i>Preprocessing</i>
D2	Perjalanan yang lumayan melelahkan akan terbayar pas tiba di air terjun gunung pandan. Arus sungai yang memberikan suara derusnya amat tenang. Hijaunya hutan memberikan warna yang indah. Sungguh indah tempatnya. Walaupun perjalanannya lumayan jauh sekitar satu jam berjalan kaki. Beautifull view.	jalan lumayan lelah bayar pas air terjun gunung pandan arus sungai suara derus tenang hijau hutan warna indah sungguh indah tempat jalan lumayan jam jalan kaki beautifull view
D3	air terjun yang indah, harus mempersiapkan stamina karena harus mendaki untuk sampai ke air terjun ini. Sepi pengunjung mungkin karena akses yang sulit.cocok untuk hiking	air terjun indah stamina daki air terjun sepi unjung akses sulit cocok hiking
...	.....	.....
D32	Pantai yang cocok untuk piknik bersama keluarga karena bisa menggunakan sepeda motor. Pantainya gratis, sepi pengunjung, banyak pohon-pohon. Disini juga bisa camp bersama keluarga atau teman.	pantai cocok piknik keluarga sepeda motor pantai gratis sepi unjung pohon pohon camp keluarga teman
D37	Bukit Kerang sepi pengunjung seperti tempat wisata lainnya yang menghadirkan pemandangan alam, lokasi tempatnya tidak terlalu mulus. Tapi, situs sejarah ini layak dikunjungi sebagai objek wisata sejarah dan untuk belajar bagi pelajar siswa/siswi.	bukit kerang sepi unjung wisata hadir pandang alam lokasi tempat tidak mulus situs sejarah layak kunjung objek wisata sejarah belajar ajar siswa siswi

### 3.2. Perhitungan TF-IDF

Langkah awal sebelum perhitungan adalah memberi bobot pada ulasan wisata yang mengandung kata *term* (atribut) yang dipilih pengguna, di mana setiap dokumen yang mengandung *term* diberi bobot 1. Tabel 3 menampilkan bobot *term*.

Tabel 3. Bobot *Term*

<i>Term</i>	Bobot			
	piknik	pantai	sungai	sepi
Q	1	1	1	1
D2	0	0	1	0
D3	0	0	0	1
D6	0	0	2	0
D14	0	0	0	1
D19	0	0	0	1
D20	2	0	0	2
D23	1	0	0	0
D24	2	0	0	2
D25	0	0	5	1
D26	1	0	0	0
D27	0	2	0	0
D28	0	0	4	0
D29	0	3	0	0
D30	0	1	0	1
D31	1	4	0	0
D32	1	2	0	1
D37	0	0	0	1
D39	1	0	0	0
D40	1	0	0	0
D41	1	0	0	1
D42	1	0	0	1
D43	0	1	0	1
D44	0	0	3	0
DF	10	6	5	12



Nilai DF (*Document Frequency*) adalah jumlah dari kemunculan suatu term dalam setiap dokumen (wisata). Hitung nilai TF (*Term Frequency*) dengan menggunakan persamaan (1). Perhitungan TF (*Term Frequency*) ditampilkan pada Tabel 4.

Tabel 4. Perhitungan TF

Term	TF			
	piknik	pantai	sungai	sepi
Q	0.250	0.250	0.250	0.250
D2	0.000	0.000	0.036	0.000
D3	0.000	0.000	0.000	0.077
...	...	...	...	...
D32	0.067	0.133	0.067	0.000
D37	0.000	0.000	0.000	0.043
...	...	...	...	...
D43	0.000	0.067	0.000	0.067
D44	0.000	0.000	0.120	0.000

IDF (*Inverse Document Frequency*) berguna untuk menunjukkan seberapa penting suatu *term* dalam keseluruhan kumpulan *document*. Menghitung nilai IDF persamaan (2). Perhitungan IDF dapat dilihat pada Tabel 5.

Tabel 5. Perhitungan IDF

IDF			
piknik	pantai	sungai	sepi
1.6198	1.5528	1.3565	1.4437

Setelah menghitung nilai TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*), langkah selanjutnya adalah menghitung nilai TF-IDF dengan menggunakan persamaan (3). Perhitungan TF-IDF ditampilkan pada Tabel 6.

Tabel 6. Perhitungan TF-IDF

Term	TF-IDF			
	piknik	pantai	sungai	sepi
Q	0.445	0.558	0.597	0.403
D2	0.000	0.000	0.085	0.000
D3	0.000	0.000	0.000	0.124
...	...	...	...	...
D32	0.119	0.298	0.000	0.067
D37	0.000	0.000	0.000	0.070
...	...	...	...	...
D43	0.000	0.149	0.000	0.108
D44	0.000	0.000	0.286	0.000

### 3.3. Perhitungan Cosine Similarity

Kemiripan vektor *query* (preferensi pengguna) dihitung dengan setiap *document* (wisata) yang tersedia menggunakan rumus *cosine similarity*. Hitung hasil perkalian skalar antara Q dan 23 dokumen lainnya. Hasil perkalian dari setiap dokumen dengan q kemudian dijumlahkan, dan hasilnya dapat dilihat seperti Tabel 7.

Langkah selanjutnya adalah menghitung panjang vektor (*magnitude*) dari setiap dokumen menggunakan. Panjang vektor digunakan dalam perhitungan kesamaan *cosine* untuk menentukan seberapa mirip satu dokumen dengan dokumen lainnya. Perhitungan panjang vektor seperti yang ditampilkan pada Tabel 8.

Tabel 7. Perkalian Skalar

Term	TF-IDF Query x TF-IDF Document				Total
	piknik	pantai	sungai	sepi	
D2	0.000	0.000	0.051	0.000	0.051
D3	0.000	0.000	0.000	0.050	0.050
...	...	...	...	...	...
D32	0.053	0.166	0.000	0.000	0.262
D37	0.000	0.000	0.000	0.028	0.028
...	...	...	...	...	...
D43	0.000	0.083	0.000	0.043	0.126
D44	0.000	0.000	0.171	0.000	0.171

Tabel 8. Perhitungan Panjang Vektor

Term	Panjang Vektor				Total	Akar
	piknik	pantai	sungai	sepi		
Q	0.311	0.198	0.162	0.356	1.027	1.013
D2	0.000	0.000	0.000	0.007	0.007	0.083
D3	0.000	0.000	0.015	0.000	0.015	0.122
...	...	...	...	...	...	...
D32	0.088	0.014	0.011	0.000	0.113	0.336
D37	0.000	0.000	0.005	0.000	0.005	0.070
...	...	...	...	...	...	...
D43	0.022	0.000	0.011	0.000	0.033	0.183
D44	0.000	0.000	0.000	0.081	0.081	0.284

Terapkan rumus *cosine similarity* dengan menghitung kemiripan antar dokumen dengan menggunakan persamaan (4).

$$\begin{aligned} \text{sim}(Q, D2) &= \frac{0.051}{1.013 \times 0.083} = \frac{0.051}{0.087} = 0.586 \\ \text{sim}(Q, D3) &= \frac{0.050}{1.013 \times 0.122} = \frac{0.050}{0.126} = 0.396 \\ \text{sim}(Q, D32) &= \frac{0.262}{1.013 \times 0.336} = \frac{0.262}{0.342} = 0.766 \\ \text{sim}(Q, D37) &= \frac{0.028}{1.013 \times 0.070} = \frac{0.028}{0.071} = 0.396 \\ \text{sim}(Q, D43) &= \frac{0.126}{1.013 \times 0.183} = \frac{0.126}{0.186} = 0.677 \\ \text{sim}(Q, D44) &= \frac{0.171}{1.013 \times 0.284} = \frac{0.171}{0.291} = 0.587 \end{aligned}$$

Jadi, urutan wisata yang menjadi rekomendasi dari atribut “piknik”, “pantai”, “sungai”, “sepi” dapat dilihat pada Tabel 9.

Tabel 9. Hasil Rekomendasi

Doc	Nama Wisata	Rangking
D32	Pantai Raja Muda	1
D30	Pantai Kupang	2
D43	Wisata Pantai Kaloy	3
D25	Pemandian Sungai Gunung Pandan	4
D31	Pantai Pusong Sium	5
D24	Pemandian Alam Batu Bedulang	6
D20	Kolam Renang Tirta Pertamina Rantau	7
D42	Wisata DAM Sungai Rengas	8
D41	TPI Tanjung Keramat	9
D6	Batu Bukit Kapur Kaloy	10
D28	Pantai Baling Karang	11



D44	Wisata Sungai Bampo	12
D2	Air Terjun Gunung Pandan	13
D27	Pantai Alur Nunang	14
D29	Pantai Ketapang	15
D23	Pemandian Air Panas Kaloy	16
D40	TPI Alur Nunang	17
D26	Pemandian Tualang Niat	18
D39	Taman Wisata Cucu Kakek	19
D14	Kolam Renang Folenda <i>Water Park</i>	20
D3	Air Terjun Keramat	21
D37	Situs Sejarah Bukit Kerang	22
D19	Kolam Renang <i>Palm Garden Park</i>	23

Urutan wisata yang menjadi rekomendasi berdasarkan atribut "piknik", "pantai", "sungai", dan "sepi" ditampilkan pada Tabel 9, dengan peringkat tertinggi diraih oleh Pantai Raja Muda (D32), diikuti oleh Pantai Kupang (D30) dan Wisata Pantai Kaloy (D43). Sistem ini menunjukkan kemampuan metode *content-based filtering* dalam memahami preferensi pengguna dengan menganalisis fitur dari ulasan yang relevan. Hasil ini memperlihatkan bahwa sistem dapat memberikan rekomendasi yang relevan dan sesuai dengan atribut pilihan pengguna.

Dengan sistem ini, wisatawan dapat dengan mudah menemukan destinasi yang sesuai dengan minat mereka, yang pada akhirnya meningkatkan pengalaman wisata secara keseluruhan. Sebagai contoh, bagi pengguna yang mencari tempat dengan atribut *hiking* dan *adventure*, sistem ini mampu menyarankan destinasi seperti Air Terjun Sangkapane, yang memenuhi kriteria tersebut. Pendekatan ini tidak hanya meningkatkan kepuasan wisatawan tetapi juga memiliki potensi untuk mendorong mereka merekomendasikan destinasi kepada orang lain, sehingga dapat berdampak positif pada sektor pariwisata daerah.

Dibandingkan dengan penelitian serupa, seperti studi oleh Yohana Imelda Lubis, Dohar Josua Napitupulu, dan Arie Satia Dharma (2020), yang menggunakan metode *hybrid filtering*, pendekatan *content-based filtering* ini memiliki kelebihan dalam hal kesederhanaan penerapan. Penelitian mereka menunjukkan bahwa metode *hybrid filtering*, yang menggabungkan *content-based filtering* dan *collaborative filtering*, memberikan akurasi lebih tinggi dengan nilai MAE terendah (0,3741) dibandingkan kedua metode tersebut secara terpisah. Namun, *hybrid filtering* memerlukan data pengguna yang lebih luas, seperti riwayat kunjungan atau penilaian sebelumnya, yang seringkali sulit tersedia di sektor pariwisata daerah.

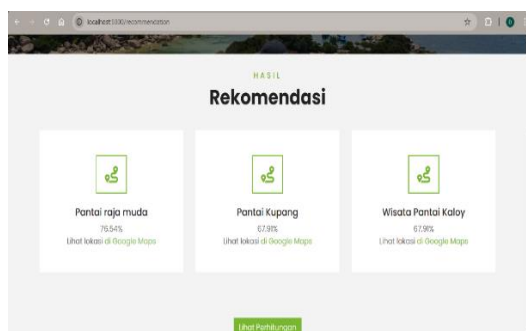
Sebaliknya, sistem rekomendasi yang dikembangkan dalam penelitian ini hanya memerlukan atribut pilihan pengguna untuk memberikan rekomendasi. Hal ini membuat metode *content-based filtering* lebih sederhana untuk diterapkan, terutama di daerah dengan data yang terbatas, namun tetap mampu memberikan hasil yang relevan dan sesuai preferensi pengguna. Dalam konteks sektor pariwisata, metode *content-based filtering* menunjukkan keunggulan dibandingkan *collaborative filtering*, yang sangat bergantung pada keberadaan data riwayat pengguna lainnya.

Sebagai pengembangan lebih lanjut, penelitian ini dapat dibandingkan dengan sistem berbasis *machine learning* atau *deep learning*, yang menggunakan pendekatan pemrosesan data ulasan untuk menghasilkan rekomendasi yang lebih canggih. Namun, pendekatan tersebut cenderung membutuhkan infrastruktur dan sumber daya komputasi yang lebih tinggi. Sistem berbasis *content-based filtering* ini tetap relevan karena menawarkan keseimbangan antara akurasi rekomendasi dan kemudahan implementasi, yang sesuai dengan kebutuhan pariwisata daerah seperti di Aceh Tamiang.

Berikut ini adalah tampilan GUI dari implementasi metode *content-based filtering*:



Gambar 2. Halaman Pilih Atribut



### Gambar 3. Hasil Rekomendasi

#### 4. KESIMPULAN

Sistem rekomendasi wisata berbasis *web* ini menggunakan metode *content-based filtering* dengan pendekatan TF-IDF dan *cosine similarity*, dirancang untuk memberikan rekomendasi objek wisata di Aceh Tamiang berdasarkan atribut ulasan Google Maps yang relevan dengan preferensi pengguna. Penelitian ini menunjukkan bahwa metode tersebut efektif dalam menghasilkan rekomendasi wisata yang relevan, dengan hasil rekomendasi seperti Pantai Raja Muda, Pantai Kupang, dan Wisata Pantai Kaloy untuk atribut "piknik", "pantai", "sungai", dan "sepi". Sistem ini berpotensi meningkatkan pengalaman wisata pengguna sekaligus mendorong promosi pariwisata lokal.

Kontribusi utama penelitian ini adalah penerapan metode yang efisien meski dengan data terbatas. Namun, sistem memiliki keterbatasan seperti bergantung pada satu metode *filtering* dan ulasan yang tidak selalu representatif.

Penelitian selanjutnya disarankan untuk mengembangkan pendekatan *hybrid filtering*, memperluas data ulasan, dan menambahkan fitur analitik untuk meningkatkan akurasi dan relevansi rekomendasi.

## DAFTAR PUSTAKA

- [1] N. Nurlina, Y. S. Pratama, and P. Andiny, "Strategi Pengembangan Industri Pariwisata (Studi Kasus Objek Wisata Pulau Rukui Kabupaten Aceh Tamiang)," *Jurnal Samudra Ekonomi dan Bisnis*, vol. 13, no. 1, pp. 1–14, Sep. 2021, doi: 10.33059/jsb.v13i1.3195.
- [2] A. I. Putra and R. R. Santika, "Implementasi Machine Learning dalam Penentuan Rekomendasi Musik dengan Metode Content-Based Filtering," *Jurnal Pendidikan Informatika*, vol. 4, no. 1, pp. 121–130, Jun. 2020, doi: 10.29408/edumatic.v4i1.2162.
- [3] M. Alkaff, H. Khatimi, and A. Eriadi, "Sistem Rekomendasi Buku pada Perpustakaan Daerah Provinsi Kalimantan Selatan Menggunakan Metode Content-Based Filtering," *Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 1, pp. 193–202, Sep. 2020, doi: 10.30812/matrik.v20i1.617.

- [4] S. Z. Putri, M. Hariadi, and R. F. Rachmadi, "Rekomendasi Barang pada E-Commerce Menggunakan Content Based Recommender System Berbasis Ethereum Blockchain," *Teknik ITS*, vol. 12, no. 2, pp. 145–152, 2023, doi: 10.12962/j23373539.v12i2.115263.
- [5] K. Hartarto Muliadi and C. Citra Lestari, "Rancang Bangun Sistem Rekomendasi Tempat Makan Menggunakan Algoritma Typicality Based Collaborative Filtering," *Techno.COM*, vol. 18, no. 4, pp. 275–287, Nov. 2019, doi: 10.33633/tc.v18i4.2515.
- [6] M. Rizki and R. Rianto, "Sistem Rekomendasi Hybrid Menggunakan Metode Switching," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 2, pp. 149–160, Aug. 2024, doi: 10.28932/jutisi.v10i2.6220.
- [7] A. Sispianygala and S. Sandino Berutu, "Pengembangan Aplikasi Sistem Rekomendasi Tempat Wisata Dengan Collaborative Filtering," *Progresif: Jurnal Ilmiah Komputer*, vol. 20, no. 20, pp. 828–838, Aug. 2024, doi: 10.35889/progresif.v20i2.2044.
- [8] T. Ridwansyah, B. Subartini, and S. Sylviani, "Penerapan Metode Content-Based Filtering pada Sistem Rekomendasi," *Mathematical Sciences and Applications Journal*, vol. 4, no. 2, pp. 70–77, Apr. 2024, doi: 10.22437/msa.v4i2.32136.
- [9] M. Muneer, U. Rasheed, S. Khalid, and M. Ahmad, "Tour Spot Recommendation System via Content-Based Filtering," in *2022 16th International Conference on Open Source Systems and Technologies, ICOSST 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICOSST57195.2022.10016820.
- [10] L. Aji Andika and P. Amalia Nur Azizah, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indonesian Journal of Applied Statistics*, vol. 2, no. 1, pp. 34–41, May 2019, doi: 10.13057/ijas.v2i1.29998.
- [11] A. E. Budiman and A. Widjaja, "Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 3, pp. 475–488, Dec. 2020, doi: 10.28932/jutisi.v6i3.2892.
- [12] N. Nofiyani and W. Wulandari, "Implementasi Electronic Data Processing Untuk meningkatkan Efektifitas dan Efisiensi Pada Text Mining," *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, pp. 1621–1629, Jul. 2022, doi: 10.30865/mib.v6i3.4332.
- [13] M. A. Rofiqi, Abd. C. Fauzan, A. P. Agustin, and A. A. Saputra, "Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query," *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 1, no. 2, pp. 58–64, Dec. 2019, doi: 10.28926/ilkomnika.v1i2.18.
- [14] R. Al Rasyid, D. Handayani, and U. Ningsih, "Penerapan Algoritma TF-IDF dan Cosine Similarity untuk Query Pencarian Pada Dataset Destinasi Wisata," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, pp. 170–78, Jan. 2024, doi: 10.35870/jti.
- [15] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *Jurnal Bumigora Information Technology*, vol. 3, no. 1, pp. 19–27, Jul. 2021, doi: 10.30812/bite.v3i1.1110.

**Halaman Ini Dikosongkan**