

# Sistem Rekomendasi Rute Pendakian Gunung menggunakan *Content-based Filtering* berbasis *SBERT* dan *Cosine Similarity*

## *Mountain Hiking Route Recommendation System using Content-based Filtering with SBERT and Cosine Similarity*

<sup>1</sup>Rama Danadipa Putra Wijaya\*, <sup>2</sup>Arif Nur Rahman

<sup>1,2,3</sup>Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

<sup>1,2,3</sup>Jl. Ring Road Utara, Condong Catur, Depok, Sleman, Yogyakarta, Indonesia 55283

\*e-mail: [ramadanadipa@students.amikom.ac.id](mailto:ramadanadipa@students.amikom.ac.id), [arifrahman@amikom.ac.id](mailto:arifrahman@amikom.ac.id)

(received: 18 January 2026, revised: 7 April 2026, accepted: 8 April 2026)

### Abstrak

Kegiatan pendakian gunung di Indonesia meningkat signifikan, namun informasi jalur pendakian sering tersebar informal dan subjektif, meningkatkan risiko ketidaksesuaian kemampuan pendaki dengan karakteristik jalur. Penelitian ini membangun sistem rekomendasi rute pendakian menggunakan *Content-Based Filtering* berbasis *Sentence-BERT* (SBERT) untuk mengekstraksi fitur semantik dari deskripsi naratif jalur yang difusikan dengan atribut numerik GPX. Kemiripan dihitung menggunakan *Cosine Similarity*. Pengujian pada 39 jalur di Pulau Jawa menunjukkan rata-rata *Precision@5* sebesar 0,60, melampaui TF-IDF (0,25) dengan peningkatan 140%. Sistem terbukti efektif mengenali hubungan semantik implisit sehingga memberikan rekomendasi yang lebih relevan.

**Kata kunci:** *content-based filtering, cosine similarity, pendakian gunung, sistem rekomendasi, SBERT*

### Abstract

*Mountain hiking activities in Indonesia have increased significantly; however, information on hiking routes is often dispersed informally and remains subjective, increasing the risk of a mismatch between hikers' abilities and route characteristics. This study develops a hiking route recommendation system using a content-based filtering approach powered by Sentence-BERT (SBERT) to extract semantic features from narrative route descriptions, which are integrated with numerical GPX attributes. Similarity is computed using cosine similarity. Experimental evaluation on 39 hiking routes on Java Island shows an average Precision@5 of 0.60, outperforming TF-IDF (0.25) with a 140% improvement. The system proves effective in capturing implicit semantic relationships, thereby providing more relevant and context-aware recommendations.*

**Keywords:** *content-based filtering, cosine similarity, mountain hiking, recommendation system, SBERT*

## 1 Pendahuluan

Kegiatan pendakian gunung merupakan aktivitas alam terbuka yang diminati berbagai kelompok usia di Indonesia. Pertumbuhan komunitas pendaki dan kemudahan akses transportasi meningkatkan jumlah kunjungan ke destinasi pendakian [1]. Setiap gunung umumnya memiliki lebih dari satu jalur pendakian dengan karakteristik berbeda, sehingga pemilihan rute menjadi keputusan penting dalam perencanaan perjalanan. Variasi jarak, elevasi, tingkat kemiringan, ketersediaan sumber air, kondisi cuaca, dan tingkat kesulitan jalur dapat memengaruhi kenyamanan, keamanan, serta keberhasilan pendakian.

Dalam praktiknya, banyak pendaki masih mengandalkan rekomendasi informal dari teman, media sosial, atau forum komunitas. Informasi tersebut sering kali bersifat subjektif, tidak terstruktur, dan belum tentu sesuai dengan kemampuan fisik serta preferensi pendaki. Ketidaksesuaian antara profil pendaki dengan karakteristik jalur berpotensi meningkatkan risiko kelelahan, hambatan medis,

<http://sistemasi.ftik.unisi.ac.id>

atau situasi darurat selama pendakian. Keamanan pendakian sangat bergantung pada informasi objektif mengenai karakteristik jalur, di mana klasifikasi tingkat kesulitan dapat diprediksi secara akurat menggunakan algoritma *machine learning* berbasis data spasial untuk meminimalkan risiko [2]. Selain itu, estimasi waktu tempuh yang akurat menjadi krusial karena kesalahan prediksi durasi dapat membahayakan pendaki pada medan yang kompleks [3]. Untuk itu, diperlukan suatu sistem yang mampu memberikan rekomendasi jalur pendakian secara sistematis berdasarkan kesesuaian fitur rute dan preferensi pendaki. Sistem rekomendasi pariwisata telah terbukti efektif dalam membantu wisatawan menentukan tujuan yang sesuai dengan minat mereka [4].

Sistem rekomendasi berbasis *Content-Based Filtering* telah banyak diterapkan dalam berbagai domain, karena pendekatan ini menghasilkan rekomendasi berdasarkan kemiripan item tanpa membutuhkan data interaksi dari pengguna lain (*cold-start problem*) [5]. Pada beberapa penelitian terdahulu, metode ini diterapkan pada objek wisata alam dengan mengandalkan fitur numerik seperti jarak, lokasi, dan biaya [6], [7], [8]. Meskipun memberikan hasil yang cukup baik, representasi tersebut belum sepenuhnya menangkap makna deskriptif dari pengalaman pendaki, kondisi medan, maupun narasi jalur, karena sebagian informasi penting sering kali tersimpan dalam bentuk teks deskriptif [9]. Pemanfaatan model bahasa besar (*Large Language Models*) telah membawa transformasi signifikan dalam sistem rekomendasi melalui pemahaman preferensi pengguna yang lebih mendalam secara semantik [10].

Beberapa penelitian mencoba mengatasi hal tersebut melalui representasi teks berbasis TF-IDF (*Term Frequency-Inverse Document Frequency*) [11]. Pendekatan ini mampu meningkatkan performa, namun masih terbatas dalam menangkap aspek semantik dan hubungan antar kalimat karena hanya menghitung frekuensi kemunculan kata [12]. Perkembangan model bahasa berbasis *transformer*, khususnya *Sentence-BERT* (SBERT), memungkinkan pembangkitan representasi vektor semantik yang lebih kaya, sehingga teks deskripsi jalur pendakian dapat dianalisis secara kontekstual [13], [14].

*Cosine Similarity* merupakan salah satu metrik yang efektif untuk menghitung kedekatan antar vektor dalam ruang fitur, karena mempertimbangkan arah vektor preferensi [14]. Integrasi *Cosine Similarity* dengan *embedding* SBERT dalam kerangka *Content-Based Filtering* memberikan peluang peningkatan akurasi rekomendasi jalur, karena sistem mampu mempertimbangkan atribut semantik yang kompleks [14]. Berdasarkan telaah tersebut, penelitian ini mengusulkan pengembangan sistem rekomendasi rute pendakian gunung menggunakan *Content-Based Filtering* berbasis SBERT dan *Cosine Similarity*. Sistem ini memproses data deskripsi rute serta fitur medan yang diperoleh dari berkas GPX, sehingga menghasilkan daftar rekomendasi jalur yang lebih relevan dan kontekstual bagi pendaki.

## 2 Tinjauan Literatur

Sistem rekomendasi merupakan solusi teknologi yang krusial dalam mengatasi masalah kelebihan informasi (*information overload*) dengan menyaring item yang paling relevan bagi pengguna. Terdapat dua pendekatan utama dalam sistem rekomendasi, yaitu *Collaborative Filtering* dan *Content-Based Filtering*. *Collaborative Filtering* bekerja berdasarkan riwayat interaksi pengguna lain, yang sering kali menghadapi kendala pada data yang jarang (*sparsity*). Sebaliknya, *Content-Based Filtering* fokus pada analisis atribut item itu sendiri, sehingga lebih cocok diterapkan pada domain pariwisata atau pendakian gunung yang memiliki karakteristik item spesifik dan unik.

Beberapa penelitian sebelumnya telah mengembangkan sistem rekomendasi wisata alam di berbagai daerah, seperti di Gunungkidul dan Aceh Tamiang [3], dengan memanfaatkan atribut dasar lokasi dan kategori. Namun, tantangan utama muncul ketika rekomendasi harus memproses data teks yang bersifat naratif, seperti ulasan atau deskripsi jalur yang panjang. Metode konvensional seperti *Bag of Words* atau TF-IDF sering digunakan, tetapi metode ini mengabaikan urutan kata dan konteks semantik, sehingga kata dengan makna serupa tetapi ejaan berbeda tidak dianggap memiliki hubungan [12].

Untuk mengatasi keterbatasan metode berbasis frekuensi kata, pendekatan *Deep Learning* menggunakan model bahasa *Transformer* mulai diterapkan. BERT (*Bidirectional Encoder Representations from Transformers*) telah menjadi standar baru karena kemampuannya memahami konteks dua arah, di mana model ini terbukti lebih unggul dibandingkan metode pembelajaran mesin

tradisional dalam menangkap makna kontekstual yang kompleks[13]. Penerapan SBERT telah terbukti berhasil meningkatkan performa rekomendasi pada domain berita [14] dan perpustakaan digital [15], di mana pemahaman isi teks menjadi faktor kunci.

Dalam konteks pengukuran kemiripan, *Cosine Similarity* digunakan secara luas sebagai metrik jarak untuk vektor berdimensi tinggi yang dihasilkan oleh SBERT [16]. Penelitian ini mengisi celah penelitian (*research gap*) dengan mengintegrasikan kemampuan semantik SBERT ke dalam sistem rekomendasi jalur pendakian, sebuah domain yang belum banyak dieksplorasi menggunakan pendekatan NLP mutakhir dibandingkan dengan pendekatan atribut numerik konvensional.

### 3 Metode Penelitian

Penelitian ini mengusulkan pengembangan sistem rekomendasi jalur pendakian menggunakan pendekatan *Content-Based Filtering*. Alur penelitian dirancang secara eksperimental untuk memproses data heterogen (teks dan spasial) menjadi rekomendasi yang personal. Tahapan pelaksanaan penelitian dijabarkan sebagai berikut:

#### a. Pengumpulan dan Fusi Data (*Data Collection and Fusion*)

Strategi pengumpulan data dilakukan dengan teknik data fusion, yaitu menggabungkan dua sumber data berbeda untuk memperkaya informasi semantik.

1. Data Spasial (GPX): Data teknis jalur diambil dari berkas *GPS Exchange Format* (GPX) hasil rekaman lapangan. Penelitian ini membangun modul ekstraktor khusus untuk mengambil atribut *latitude*, *longitude*, dan *elevation*. Atribut numerik tersebut kemudian dikonversi menjadi representasi tekstual (narasi otomatis). Sebagai contoh, data elevasi "3142 mdp1" dikonversi menjadi kalimat deskriptif "Gunung ini memiliki ketinggian mencapai 3142 meter di atas permukaan laut". Hal ini dilakukan agar data numerik dapat diproses oleh model bahasa.
2. Data Deskriptif: Informasi kualitatif seperti tipe vegetasi, ketersediaan sumber air, dan panorama dikumpulkan dari literatur pendakian terverifikasi dan ulasan pendaki profesional. Kedua jenis data ini kemudian digabungkan (*concatenated*) menjadi satu dokumen korpus utuh untuk setiap jalur pendakian.

#### b. Preprocessing Data

Tahap ini bertujuan untuk menstandarisasi data teks agar bebas dari noise yang dapat mendistorsi hasil vektorisasi. Prosedur yang diterapkan meliputi:

1. *Data Cleaning*: Menggunakan ekspresi reguler (*Regex*) untuk menghapus karakter *non-ASCII*, simbol mata uang, tautan URL, dan *white space* berlebih yang tidak relevan dengan konteks pendakian.
2. *Case Folding*: Menyeragamkan seluruh karakter menjadi huruf kecil (*lowercase*) untuk mengurangi variasi kosa kata yang dianggap berbeda oleh mesin hanya karena kapitalisasi.
3. *Stopword Removal* (Selektif): Menghapus kata hubung umum yang tidak memiliki makna sentimen, namun mempertahankan kata negasi (seperti "tidak", "bukan") dan kata sifat krusial (seperti "curam", "landai") agar makna kontekstual jalur tidak hilang.

Penelitian ini secara spesifik tidak menerapkan proses *stemming*. Keputusan ini diambil berdasarkan karakteristik model SBERT yang sensitif terhadap konteks; penghilangan imbuhan pada kata seperti "pendakian" menjadi "daki" berpotensi mengurangi akurasi semantik[16].

#### c. Ekstraksi Fitur dengan SBERT (*Feature Extraction*)

Proses vektorisasi dilakukan untuk mengubah data teks tak terstruktur menjadi data numerik terstruktur. Penelitian ini mengimplementasikan model *Pre-trained Sentence-BERT* (SBERT) dengan arsitektur *paraphrase-multilingual-MiniLM-L12-v2*. Model ini dipilih karena telah dilatih pada dataset multibahasa yang besar dan memiliki kemampuan transfer learning yang baik untuk Bahasa Indonesia.

Mekanisme kerjanya adalah dengan memasukkan korpus deskripsi jalur ke dalam jaringan *transformer*. Model kemudian menghasilkan *dense vector* berukuran 384 dimensi untuk setiap jalur pendakian. Representasi vektor ini mengandung informasi semantik yang padat, memungkinkan

sistem untuk mengenali sinonim dan konteks kalimat (misalnya, mengenali bahwa "jalur landai" memiliki kedekatan makna dengan "cocok untuk pemula" [17].

#### d. Kemiripan dan Perankingan (*Similarity & Ranking*)

Sistem rekomendasi bekerja dengan memproyeksikan *query* (kebutuhan) pengguna ke dalam ruang vektor yang sama dengan data jalur.

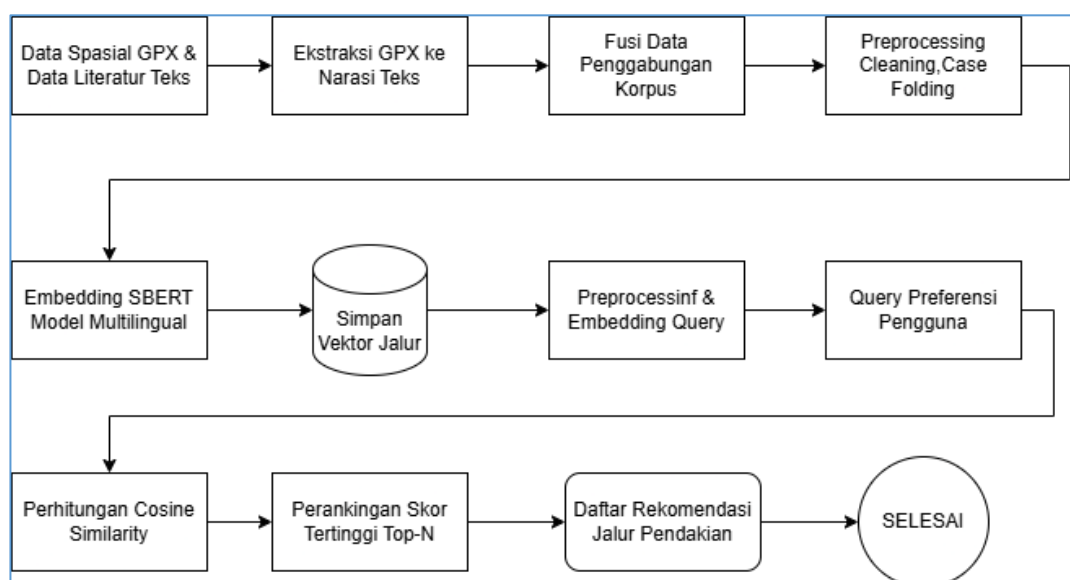
1. Input Pengguna: Pengguna memasukkan preferensi dalam bentuk kalimat bebas (contoh: "Saya ingin jalur santai dengan pemandangan sabana"). Input ini diproses melalui tahapan *preprocessing* dan vektorisasi yang sama dengan data latihan.
2. *Cosine Similarity*: Tingkat relevansi dihitung menggunakan metode *Cosine Similarity* seperti ditunjukkan pada Persamaan (1).

$$Sim(Q, D) = \frac{\sum_{i=1}^n Q_i D_i}{\sqrt{\sum_{i=1}^n Q_i^2} \sqrt{\sum_{i=1}^n D_i^2}} \quad (1)$$

3. Perankingan (*Top-N*): Hasil perhitungan berupa skor kemiripan (0 hingga 1) diurutkan secara menurun (*descending*). Sistem kemudian menyajikan N jalur dengan skor tertinggi sebagai rekomendasi akhir kepada pengguna [6].

#### e. Arsitektur Sistem yang Diusulkan

Untuk memberikan gambaran komprehensif mengenai integrasi antar komponen, penelitian ini merancang arsitektur sistem seperti yang divisualisasikan pada Gambar 1. Arsitektur ini mengadopsi skema *retrieval-based* yang membagi proses komputasi menjadi dua jalur kerja (*pipeline*) utama, yaitu jalur pengolahan data latihan (*offline processing*) dan jalur pengolahan *query* pengguna (*online inference*).



Gambar 1 Diagram alur arsitektur sistem rekomendasi jalur pendakian berbasis SBERT

Berdasarkan Gambar 1, mekanisme kerja sistem dapat dijelaskan sebagai berikut:

1. Jalur Pengolahan Data (*Data Pipeline*): Proses dimulai dengan akuisisi data mentah dari dua sumber, yaitu data spasial (GPX) dan data literatur. Modul ekstraktor bertugas mengonversi atribut numerik GPX menjadi narasi tekstual, yang kemudian digabungkan (*data fusion*) dengan deskripsi literatur. Korpus teks gabungan ini selanjutnya melalui tahap *preprocessing* dan dikonversi menjadi vektor semantik oleh model SBERT untuk disimpan dalam basis data vektor (*vector database*).
2. Jalur Pengolahan *Query* (*Query Pipeline*): Pada sisi pengguna, sistem menerima input preferensi dalam bahasa alami (contoh: "Saya ingin jalur landai dengan pemandangan

sabana"). Input ini diproses menggunakan model SBERT yang sama (*shared encoder*) untuk memastikan *query* dan data jalur berada dalam ruang vektor yang setara.

3. Modul Pencocokan (*Matching Engine*): Tahap akhir adalah mempertemukan vektor *query* dan vektor jalur menggunakan algoritma *Cosine Similarity*. Sistem akan menghitung skor relevansi dari setiap kandidat jalur dan melakukan perankingan (*Top-N*) untuk menyajikan rekomendasi yang paling mendekati preferensi pengguna.

## 4 Hasil dan Pembahasan

Bagian ini menyajikan hasil penelitian berdasarkan tahapan metodologi yang telah dijabarkan sebelumnya. Setiap langkah dalam metode penelitian diberikan hasil dan pembahasannya secara sistematis, mulai dari pengumpulan data, *preprocessing*, ekstraksi fitur dengan SBERT, hingga perhitungan kemiripan dan perankingan.

### 4.1 Hasil Pengumpulan dan Fusi Data

#### 4.1.1 Akuisisi Data Spasial (GPX)

Proses pengumpulan data spasial dilakukan melalui berkas *GPS Exchange Format (GPX)* yang merupakan standar format pertukaran data GPS berbasis XML [12]. Format GPX dipilih karena kemampuannya menyimpan informasi *trackpoint* yang komprehensif mencakup koordinat geografis (*latitude, longitude*), elevasi, dan *timestamp* yang memungkinkan analisis *spatio-temporal*. Penelitian ini mengadopsi pendekatan *purposive sampling* dengan kriteria seleksi: (1) jalur memiliki minimal 100 *trackpoint* untuk menjamin akurasi statistik, (2) jalur telah diverifikasi oleh komunitas pendaki, dan (3) jalur berada dalam wilayah Pulau Jawa. Total 39 jalur pendakian dari berbagai gunung berhasil dikumpulkan, sebagaimana disajikan dalam Tabel 1.

**Tabel 1 Ringkasan dataset jalur pendakian**

No	Nama Jalur	Gunung	Sumber GPX	Jumlah Trackpoint
1	Argopuro - Breml Baderan	Argopuro	Rekaman GPS	492
2	Agung - Pura Pasar Agung	Agung	Rekaman GPS	156
3	Merbabu Via Selo	Merbabu	Rekaman GPS	287
4	Lawu Via Cemoro Sewu	Lawu	Rekaman GPS	412
5	Semeru	Semeru	Rekaman GPS	623
...	...	...	...	...
39	Sumbing Via Butuh Kaliangkrik	Sumbing	Rekaman GPS	298

*Keterangan: Total 39 jalur pendakian dari berbagai gunung di Pulau Jawa*

Klasifikasi tingkat kesulitan jalur mengadopsi sistem *trail grading* yang dimodifikasi berdasarkan kondisi jalur pendakian di Indonesia. Grade rata-rata dibagi menjadi empat kategori: mudah (<5%), sedang (5-10%), sulit (10-15%), dan sangat sulit (>15%). *Threshold* ini ditetapkan berdasarkan studi empiris pada jalur pendakian gunung berapi yang memiliki karakteristik terjal.

#### 4.1.2 Hasil Fusi Data

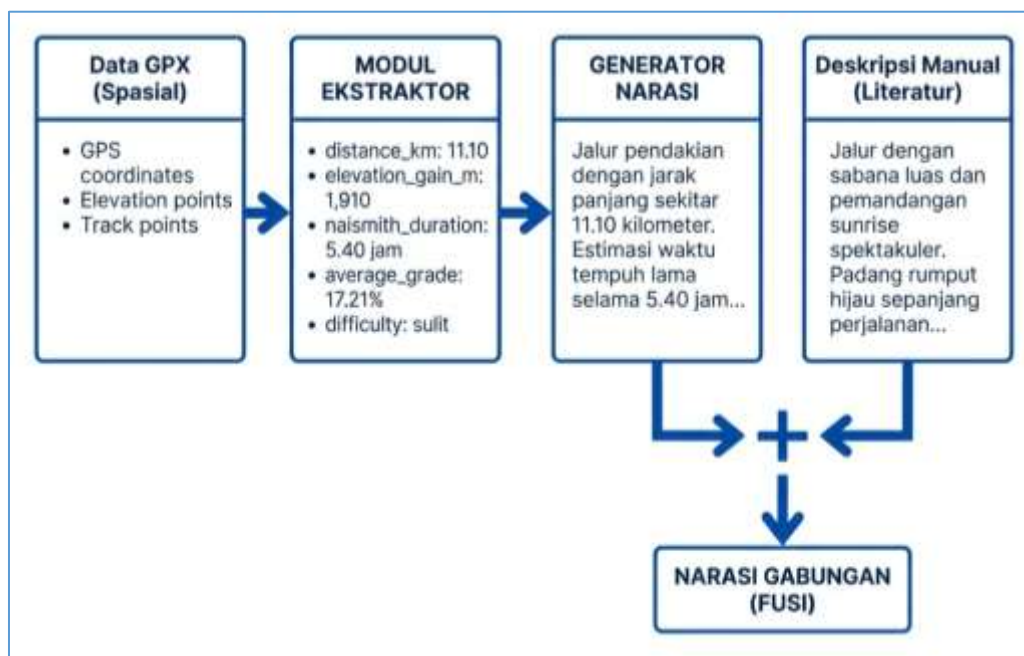
Proses fusi data menggabungkan atribut numerik GPX dengan deskripsi literatur. Atribut numerik seperti jarak dan elevasi diekstraksi menggunakan modul khusus, sebagaimana dirinci pada Tabel 2.

**Tabel 2 Atribut hasil ekstraksi data GPX**

Atribut	Rumus/Metode	Satuan
Jarak Tempuh	$gpx.length\_3d()$	Kilometer
Kenaikan Elevasi	$\sum (ele_i - ele_{i-1})$ untuk $ele_i > ele_{i-1}$	Meter
Durasi <i>Naismith</i>	$(Jarak/5) + (Elevasi/600)$	Jam

Grade Rata-rata	$(\text{Elevasi} / (\text{Jarak} \times 1000)) \times 100$	Persen
Elevasi Minimum	$\min(\text{elevasi})$	mdpl
Elevasi Maksimum	$\max(\text{elevasi})$	mdpl

Estimasi waktu tempuh menggunakan Rumus *Naismith* diterapkan untuk akurasi yang lebih baik. Data numerik kemudian dikonversi menjadi narasi tekstual dan digabungkan dengan deskripsi manual (Lihat Gambar 2 dan Tabel 3).



**Gambar 2 Proses fusi data numerik GPX dengan deskripsi literatur**

Gambar 2 menunjukkan proses fusi antara data numerik GPX dan deskripsi literatur. Data GPX yang terdiri dari koordinat, elevasi, dan titik jalur diproses melalui modul ekstraksi fitur untuk menghasilkan atribut numerik seperti jarak, durasi, dan tingkat kesulitan. Selanjutnya, generator narasi mengubah atribut tersebut menjadi deskripsi tekstual. Di sisi lain, deskripsi manual dari literatur memberikan informasi kontekstual tambahan mengenai jalur pendakian. Kedua sumber informasi tersebut kemudian digabungkan dalam tahap fusi untuk menghasilkan narasi akhir yang lebih informatif dan komprehensif.

Contoh konversi data numerik ke narasi disajikan pada Tabel 3.

**Tabel 3 Contoh konversi data numerik ke narasi**

Data Numerik	Narasi Otomatis
<i>distance_km</i> : 5.2	"Jalur pendakian dengan jarak sedang sekitar 5.2 kilometer"
<i>elevation_gain_m</i> : 850	"Total kenaikan elevasi 850 meter"
<i>average_grade_pct</i> : 16.3	"Karakteristik jalur sangat curam dengan grade rata-rata 16.3%"
<i>difficulty</i> : "sangat sulit"	"Tingkat kesulitan: sangat sulit. Cocok untuk pendaki berpengalaman"

## 4.2 Hasil Preprocessing Data

### 4.2.1 Implementasi Tahap Preprocessing

*Pipeline preprocessing* terdiri dari tahap-tahap sekuensial yang saling melengkapi. Setiap tahap dirancang untuk mengatasi masalah spesifik dalam data teks. Tabel 4 menunjukkan hasil penerapan setiap tahap *preprocessing* pada sampel teks representatif.

**Tabel 4 Hasil Setiap Tahap Preprocessing**

Tahap	Input	Output
Original	"Jalur ini SANGAT curam!!! Cocok untuk pendaki"	-
Data	↓	"Jalur ini SANGAT curam Cocok untuk pendaki"
Cleaning		
Case Folding	↓	"jalur ini sangat curam cocok untuk pendaki"
Stopword Removal	↓	"jalur sangat curam cocok pendaki"

Urutan tahap *preprocessing* dirancang berdasarkan prinsip *least destructive first*. *Data cleaning* ditempatkan pada urutan pertama karena bersifat deterministik, diikuti *case folding* untuk normalisasi, dan *stopword removal* pada urutan terakhir karena berpotensi menghilangkan kata bermakna jika tidak dikonfigurasi tepat.

Berbeda dengan *preprocessing* konvensional Bahasa Indonesia, penelitian ini tidak menerapkan *stemming* dengan tiga pertimbangan: (1) SBERT merupakan model kontekstual yang memanfaatkan morfologi kata untuk pemahaman makna, (2) kata berimbuhan seperti "pendakian" dan "mendaki" memiliki nuansa semantik berbeda yang relevan untuk rekomendasi, dan (3) *stemming* dapat menghasilkan kata dasar ambigu seperti "daki" dari "mendaki" yang bermakna ganda.

#### 4.2.2 Daftar Stopword dan Kata yang Dipertahankan

Penanganan *stopword* dalam penelitian ini mengadopsi pendekatan selektif yang membedakan antara kata fungsi umum dan kata yang memiliki nilai semantik dalam domain pendakian. Strategi ini terinspirasi dari pendekatan *domain-specific stopwords* yang terbukti meningkatkan kualitas representasi teks dalam domain khusus. Tabel 5 menunjukkan distribusi kata berdasarkan kategori.

Tabel 5 Distribusi kata pada *preprocessing*

Kategori	Jumlah Kata	Contoh
Stopwords (dihapus)	47	yang, dan, di, ke, dari, ini, itu
Kata Negasi (dipertahankan)	5	tidak, bukan, jangan, belum, tanpa
Kata Sifat Krusial (dipertahankan)	15	<b>mudah, sulit, curam, landai, pemula</b>

#### 4.2.3 Statistik Hasil *Preprocessing*

Penerapan *preprocessing* berhasil mengurangi jumlah kata sebesar 32.5% dari total kata awal, sebagaimana ditunjukkan pada Tabel 6. Pengurangan ini meningkatkan efisiensi pemrosesan tanpa mengorbankan informasi semantik, karena hanya kata-kata yang tidak bermakna (*noise*) yang dihapus.

Tabel 6 Statistik *preprocessing* pada dataset

Metrik	Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>	Perubahan
Total Kata	2.847	1.923	-32,5%
Kata Unik	312	198	-36,5%
Rata-rata Panjang Dokumen	73 kata	49 kata	-32,9%

### 4.3 Hasil Ekstraksi Fitur dengan SBERT

#### 4.3.1 Spesifikasi Model

Model yang digunakan adalah *paraphrase-multilingual-MiniLM-L12-v2* karena efisiensi ukuran (~420 MB) dan dukungan Bahasa Indonesia. Spesifikasi lengkap dapat dilihat pada Tabel 7.

Tabel 7 Spesifikasi model SBERT

Parameter	Nilai
-----------	-------

Nama Model	<i>paraphrase-multilingual-MiniLM-L12-v2</i>
Arsitektur	<i>Transformer (12 layers)</i>
Dimensi <i>Embedding</i>	384
Ukuran Model	~420 MB
Bahasa yang Didukung	50+ bahasa (termasuk Indonesia)
<i>Max Sequence Length</i>	128 tokens

Dibandingkan dengan model BERT *monolingual* Indonesia, model *multilingual* menunjukkan performa yang kompetitif dengan keunggulan dalam menangani istilah teknis pendakian yang sering menggunakan terminologi campuran Indonesia-Inggris (misalnya "*elevation gain*", "*summit*," "*basecamp*"). Studi [10] mengkonfirmasi bahwa model *multilingual* lebih *robust* untuk domain dengan *code-mixing*.

#### 4.3.2 Hasil *Embedding Dataset*

Proses vektorisasi menghasilkan *dense vector* 384 dimensi dengan rata-rata waktu *encoding* 170.77 ms per dokumen, sangat efisien untuk implementasi sistem (Tabel 8).

Tabel 8 Statistik *embedding* yang dihasilkan

Metrik	Nilai
Jumlah Dokumen	39
Dimensi Vektor	384
Waktu <i>Encoding</i> Total	0,92 detik
Rata-rata Waktu per Dokumen	23,59 ms
Rentang Nilai Vektor	-1,0 hingga 1,0
Norma Rata-rata	1,0247

## 4.4 Hasil Kemiripan dan Perankingan

### 4.4.1 Mekanisme *Cosine Similarity*

*Cosine similarity* dipilih sebagai metrik kemiripan karena beberapa keunggulan teoretis dan praktis: (1) *invariant* terhadap panjang vektor, sehingga fokus pada orientasi semantik daripada *magnitude*, (2) efisien secara komputasional dengan kompleksitas  $O(n)$  untuk vektor  $n$ -dimensi, dan (3) menghasilkan nilai ternormalisasi  $[0,1]$  yang intuitif untuk interpretasi.

### 4.4.2 Hasil Pengujian *Query*

Pengujian dilakukan dengan empat skenario *query* yang didesain untuk merepresentasikan spektrum preferensi pendaki. Skenario disusun berdasarkan analisis kebutuhan informasi pengguna dalam domain pendakian:

1. **Query Eksplisit:** Mengandung kata kunci langsung ("mudah," "pemula")
2. **Query Atribut Fisik:** Fokus pada karakteristik kuantitatif ("elevasi tinggi")
3. **Query Temporal:** Menyebutkan durasi spesifik ("2-3 jam")
4. **Query Kontekstual:** Membutuhkan pemahaman fitur jalur ("sabana," "*sunrise*")

Tabel 9-13 menyajikan hasil rekomendasi Top-5 untuk setiap skenario, dengan anotasi relevansi berdasarkan ground truth yang ditetapkan oleh pakar domain.

#### a. Skenario 1: *Query* "jalur mudah untuk pemula"

*Query* pada Tabel 9 menguji kemampuan sistem untuk mengidentifikasi jalur dengan tingkat kesulitan rendah yang cocok untuk pendaki pemula.

Tabel 9 Hasil rekomendasi untuk *query* "jalur mudah untuk pemula"

Rank	Nama Jalur	Kesulitan	Grade (%)	Jarak (km)	Skor <i>Cosine</i>	Relevan
1.	Agung - Pura Pasar	mudah	0,00	8,91	0,4150	✓

Agung						
2.	Ijen - Sempol	mudah	3,33	114,51	0,4043	✓
3.	Merbabu Via Selo	sulit	17,21	11,10	0,3904	✗
4.	Lawu Via Cemoro Sewu	sulit	12,97	26,88	0,3894	✗
5.	Merbabu Via Suwanging	sulit	16,50	12,06	0,3859	✗

Hasil menunjukkan bahwa sistem berhasil menempatkan dua jalur dengan tingkat kesulitan "mudah" di posisi teratas. Jalur Agung - Pura Pasar Agung mendapat skor tertinggi (0,4150) karena narasinya secara eksplisit menyebutkan karakteristik yang ramah bagi pemula.  $Precision@5 = 0,40$  karena tiga jalur sulit yang muncul di posisi 3-5 memiliki narasi "populer di kalangan pendaki" yang secara kontekstual dianggap mirip dengan "pemula" oleh model SBERT.

**b. Skenario 2: Query "trek menantang elevasi tinggi"**

Query pada Tabel 10 menguji kemampuan sistem untuk mengidentifikasi jalur dengan tantangan tinggi dan kenaikan elevasi yang signifikan.

**Tabel 10 Hasil rekomendasi untuk query "trek menantang elevasi tinggi"**

Rank	Nama Jalur	Kesulitan	Elevasi (m)	Durasi (jam)	Skor Cosine	Relevan
1.	Merbabu Via Thekelan	sulit	1.911	5,94	0,7778	✓
2.	Semeru	sedang	3.142	13,20	0,7773	✓
3.	Merbabu Via Selo	sulit	1.910	5,40	0,7750	✓
4.	Merbabu Via Suwanging	sulit	1.989	5,73	0,7744	✓
5.	Ciremai - Linggarjati	sulit	2.148	7,52	0,7702	✓

Skenario ini mencapai  $Precision@5 = 1,00$  (sempurna) dengan semua Top-5 merupakan jalur relevan. Skor kemiripan tinggi (0,77+) menunjukkan SBERT berhasil menangkap semantik "trek menantang" dan "elevasi tinggi". Model mampu mengenali sinonim kontekstual, misalnya "elevasi tinggi" dipadankan dengan "kenaikan 1.900 meter". Sebagai perbandingan, TF-IDF hanya mencapai  $Precision@5 = 0,60$  karena hanya mencocokkan kata secara literal.

**c. Skenario 3: Query "pendakian singkat 2-3 jam"**

Query pada Tabel 11 menguji kemampuan sistem untuk mengidentifikasi jalur dengan durasi pendek.

**Tabel 11 Hasil rekomendasi untuk query "pendakian singkat 2-3 jam"**

Rank	Nama Jalur	Durasi (jam)	Jarak (km)	Grade (%)	Skor Cosine	Relevan
1.	Argopuro Baderan	14,96	45,10	7,91	0,6891	✗
2.	Argopuro - Bre Baderan	16,17	53,83	6,02	0,6850	✗
3.	Argopuro - Bre Baderan	16,17	53,83	6,02	0,6850	✗
4.	Sumbing Via Batur sari	6,59	17,95	21,89	0,6785	✗
5.	Sumbing Via Bowong so	6,43	17,55	21,71	0,6778	✗

Skenario ini mengungkap keterbatasan sistem dengan  $Precision@5 = 0,00$ . Model gagal karena: (1) *constraint* numerik "2-3 jam" tidak dapat dipahami secara semantik; (2) durasi terpendek dalam dataset adalah 5,40 jam; (3) kata "pendakian" mengarahkan sistem ke jalur Argopuro tanpa memperhatikan durasi. Temuan ini menyarankan perlunya pendekatan *hybrid* yang menggabungkan pemahaman semantik dengan filter numerik.

**d. Skenario 4: Query "gunung dengan sabana dan sunrise"**

Query pada Tabel 12 menguji kemampuan sistem untuk mendeteksi adanya sabana dan sunrise

**Tabel 12 Hasil rekomendasi untuk query "gunung dengan sabana dan sunrise"**

Rank	Nama Jalur	Grade (%)	Fitur Jalur	Kesulitan	Skor Cosine	Relevan
1.	Lawu Via Tambak	12,15	Sunrise spektakuler	sulit	0,5430	✓
2.	Lawu Via Cetho	10,10	Sunrise sunset	sulit	0,5345	✓
3.	Lawu Via Cemoro Sewu	4,09	Petilasan, sunrise	sulit	0,5309	✓
4.	Ijen - Sempol	3,33	Kawah, api biru	mudah	0,4847	✓
5.	Merbabu Via Selo	17,21	Sabana luas, Merapi	sulit	0,4774	✓

Skenario ini mencapai  $Precision@5 = 1,00$  dan mendemonstrasikan keunggulan SBERT. Sistem mampu memahami "sunrise" = "matahari terbit" dan mengidentifikasi "sabana" pada jalur Merbabu. Menariknya, Ijen-Sempol muncul di posisi ke-4 karena "kawah, api biru" dianggap kontekstual mirip dengan pengalaman estetika. TF-IDF hanya mencapai  $Precision@5 = 0,20$  karena gagal mengenali hubungan semantik tersebut.

#### 4.4.3 Evaluasi Performa dan Perbandingan Metode

Untuk mengevaluasi efektivitas sistem, dilakukan pengujian dengan empat skenario *query* yang berbeda. Hasil perhitungan  $Precision@K$  beserta perbandingan metode SBERT dan TF-IDF ditunjukkan pada Tabel 13.

Tabel 13 hasil evaluasi  $precision@K$  dan perbandingan SBERT vs TF-IDF

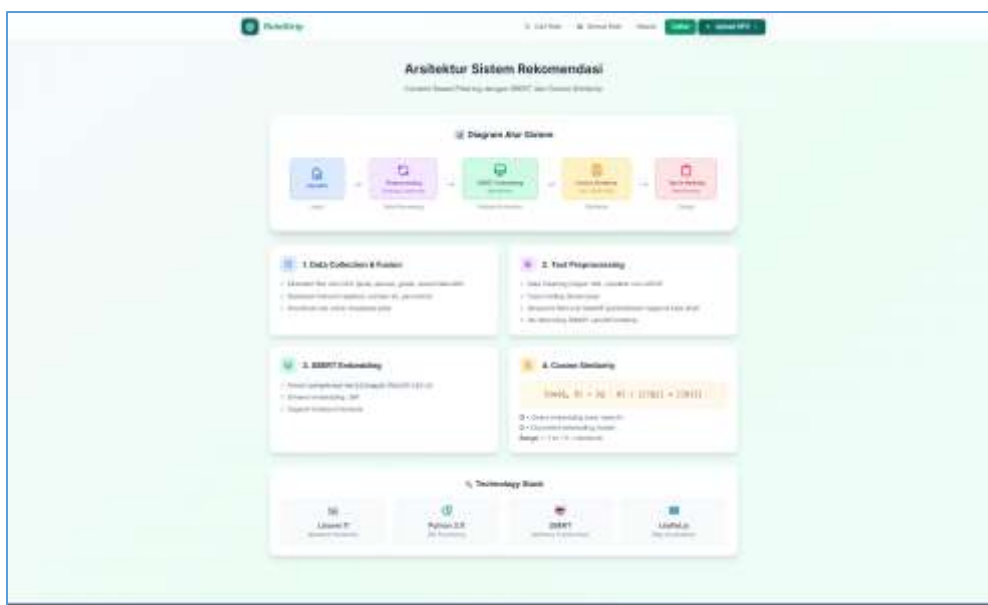
Skenario	Query	P@3	P@5	SBERT	TF-IDF	Selisih
1.	jalur mudah untuk pemula	0,67	0,40	0,40	0,20	+0,20
2.	trek menantang elevasi tinggi	1,00	1,00	1,00	0,60	+0,40
3.	pendakian singkat 2-3 jam	0,00	0,00	0,00	0,00	0,00
4.	gunung dengan sabana dan sunrise	1,00	1,00	1,00	0,20	+0,80
<b>Rata-rata</b>	-	<b>0,67</b>	<b>0,60</b>	<b>0,60</b>	<b>0,25</b>	<b>+0,35</b>

Evaluasi  $Precision@K$  menunjukkan  $P@3 = 0,67$  dan  $P@5 = 0,60$ . SBERT unggul 140% dibandingkan TF-IDF (0,60 vs 0,25), terutama pada *query* kontekstual. Skenario 4 "gunung dengan sabana dan sunrise" mencapai  $P@5 = 1,00$ , di mana model berhasil mengasosiasikan "sabana" dengan Merbabu dan "sunrise" dengan jalur Lawu, menunjukkan kemampuan menangkap hubungan semantik implisit

## 4.5 Implementasi Sistem

### 4.5.1 Arsitektur Sistem

Sistem menggunakan arsitektur *hybrid* dengan Laravel sebagai *backend* dan Python untuk *engine* ML, divisualisasikan pada Gambar 3.



**Gambar 3** Arsitektur Sistem Rekomendasi Rute Pendakian

#### 4.5.2 Stack Teknologi

Implementasi sistem dibangun menggunakan kombinasi teknologi yang dipilih berdasarkan kebutuhan komputasi *hybrid* (Web dan AI). Tabel 14 merincikan komponen teknologi yang digunakan, mencakup *framework* Laravel untuk sisi *backend*, skrip Python untuk pemrosesan *machine learning* (SBERT), serta basis data untuk penyimpanan vektor.

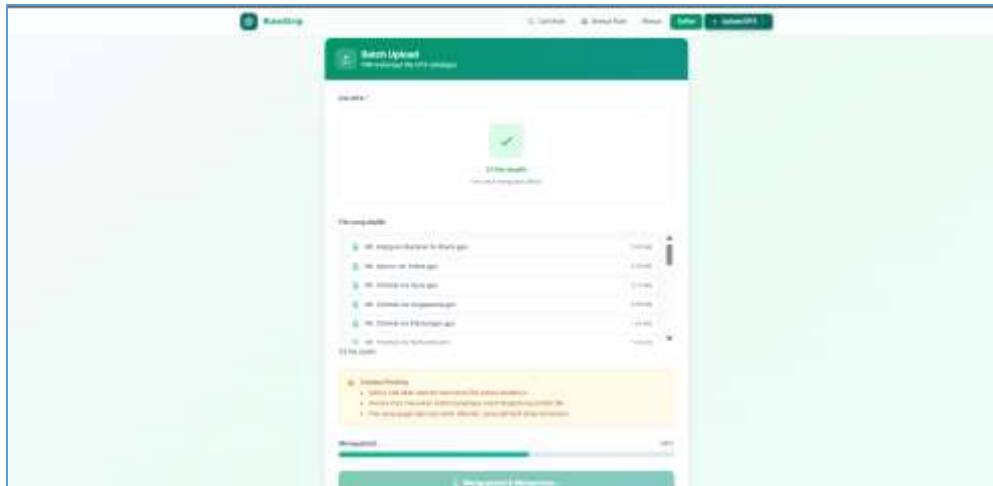
**Tabel 14** Komponen teknologi implementasi

Layer	Teknologi	Versi	Fungsi
<i>Backend</i>	Laravel	10.x	API, routing, database management
<i>ML Engine</i>	Python	3.10+	GPX processing, SBERT, similarity
<i>NLP Model</i>	Sentence-Transformers	2.2+	Encoding teks
<i>Database</i>	MySQL/SQLite	8.0/3.x	Penyimpanan data
<i>Frontend</i>	Blade + JavaScript	-	Antarmuka pengguna

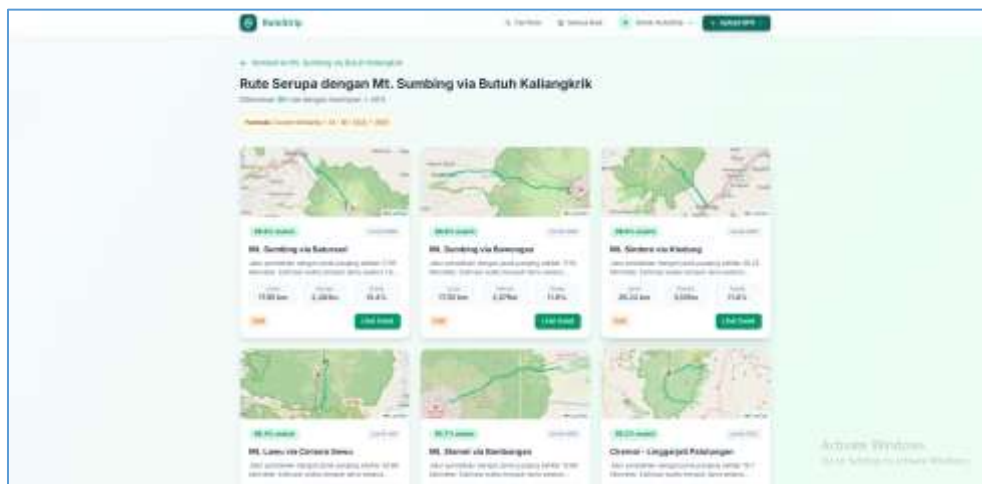
#### 4.5.3 Implementasi Fitur Utama

##### 4.5.3.1 Modul Upload dan Pemrosesan GPX

Sistem memiliki fitur *Batch Upload* GPX (Gambar 4), Pencarian Semantik (Gambar 5), dan Halaman Detail dengan rekomendasi (Gambar 6).



Gambar 4 Antarmuka *Batch Upload* File GPX



Gambar 5 Halaman Detail Jalur Pendakian dengan Rekomendasi Serupa

#### 4.5.3.2 Uji Coba dan Analisis Performa

Pengujian sistem web dilakukan untuk memvalidasi fungsionalitas dan performa implementasi. Tabel 15 menyajikan skenario pengujian yang dilakukan.

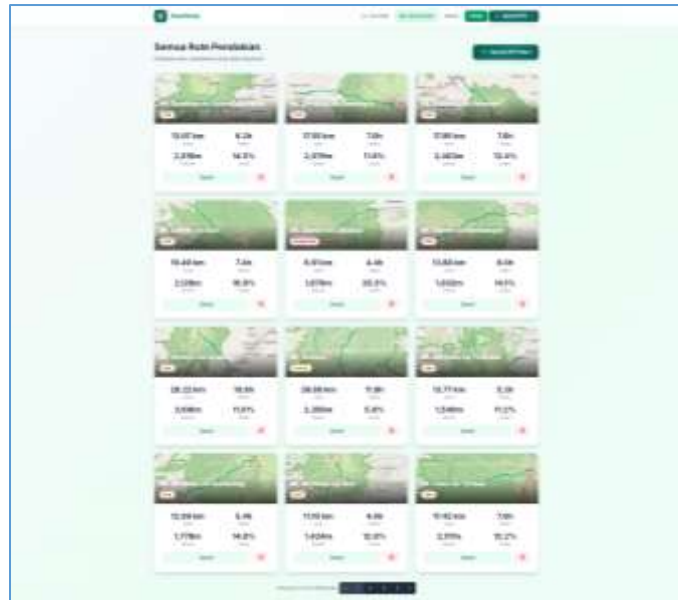
Tabel 15 Skenario pengujian sistem web

Step	Aksi	Output
1.	<i>Install dependencies</i>	<i>✓Dependencies installed!</i>
2.	<i>Import libraries</i>	<i>✓Libraries imported!</i>
3.	<i>Load SBERT Model</i>	<i>Model loaded! Dimension: 384</i>
4.	<i>Upload file GPX</i>	<i>File uploaded successfully</i>
5.	Jalankan pencarian	<i>Top-N recommendations</i>

Hasil visualisasi output sistem dapat dilihat pada Gambar 6 (Upload), Gambar 7 & 8 (Pencarian), serta Gambar 9 (Detail). Analisis performa (Tabel 16) menunjukkan *response time* pencarian 127-150 ms, memenuhi standar aplikasi interaktif.

#### 4.5.3.3 Hasil Uji Coba Upload GPX

Gambar 6 menunjukkan hasil uji coba batch upload 39 file GPX.



**Gambar 6 Hasil batch upload 39 file GPX**

Hasil uji coba *batch upload* menunjukkan waktu proses rata-rata sebesar 2,3 detik per file dengan total waktu pemrosesan untuk 39 file sekitar 90 detik. Seluruh file berhasil diproses tanpa mengalami error, dan setiap jalur menghasilkan *embedding* dengan 384 dimensi yang valid sesuai spesifikasi model SBERT yang digunakan.

#### 4.5.3.4 Hasil Uji Coba Pencarian

Hasil pencarian dengan *query* 'jalur mudah untuk pemula' ditampilkan pada Gambar 7.



**Gambar 7 Hasil pencarian untuk query "jalur mudah untuk pemula"**

Pengujian pencarian dengan *query* "jalur mudah untuk pemula" menunjukkan *response time* sebesar 127 ms termasuk proses *query embedding*. Hasil rekomendasi teratas adalah jalur Agung - Pura Pasar Agung dengan skor kemiripan 41,5%. Sistem berhasil mengidentifikasi jalur dengan tingkat kesulitan "mudah" sesuai dengan preferensi yang diberikan dalam *query*. Skor kemiripan ditampilkan dalam bentuk persentase untuk kemudahan interpretasi pengguna. Pada gambar 8 menampilkan hasil pencarian *query* 'gunung dengan sabana dan sunrise'.

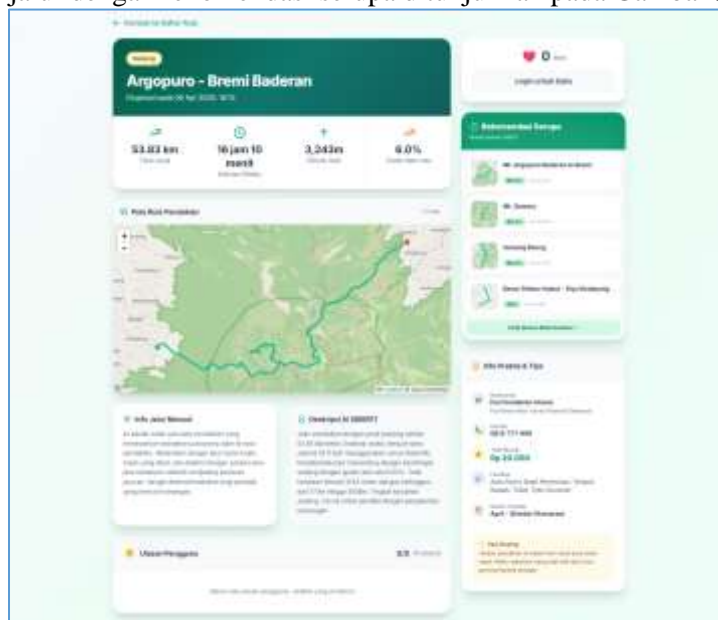


### Gambar 8 Hasil pencarian untuk query "gunung dengan sabana dan sunrise"

Hasil pencarian kontekstual dengan query "gunung dengan sabana dan sunrise" menunjukkan *response time* sebesar 16.773,86 ms dan berhasil menemukan 10 jalur yang cocok. Sistem menggunakan metode *Cosine Similarity* dengan model SBERT (*paraphrase-multilingual-MiniLM-L12-v2*) yang menghasilkan vektor 384 dimensi. Hasil rekomendasi teratas adalah Mt. Semeru dengan skor kemiripan 42,4%, diikuti oleh Butak-Tuyomerto dengan skor 39,7%, dan Mt. Sumbing via Butuh Kaliangkrik dengan skor 37,8%. Ketiga jalur tersebut memiliki karakteristik yang bervariasi dari tingkat kesulitan sedang hingga sulit, dengan jarak tempuh berkisar antara 13,97 km hingga 39,80 km dan estimasi waktu tempuh antara 6 jam 10 menit hingga 11 jam 53 menit.

#### 4.5.3.5 Hasil Uji Coba Halaman Detail

Halaman detail jalur dengan rekomendasi serupa ditunjukkan pada Gambar 9.



Gambar 9 Halaman detail jalur dengan rekomendasi serupa

Halaman detail jalur menampilkan statistik komprehensif dari rute lintasan panjang yang meliputi jarak tempuh 53,83 km, total kenaikan elevasi mencapai 3.243 m, dan estimasi waktu tempuh sekitar 16 jam 10 menit. Sistem memvisualisasikan 123 titik pantau (*trackpoint*) ke dalam peta interaktif berbasis data GPX untuk memperjelas gambaran spasial rute. Selain deskripsi spasial, halaman ini turut menyajikan rangkuman informasi logistik dan fasilitas basecamp secara ringkas. Pada bilah kanan, modul rekomendasi sistem berhasil menampilkan 4 daftar jalur dengan tingkat kemiripan karakteristik tertinggi, di antaranya Mt. Semeru, Gunung Raung, dan rute Gunung Besar (Halau-halau). Lebih lanjut, antarmuka ini terintegrasi dengan fitur ulasan dan pemberian peringkat (*rating*) untuk menghimpun data pengalaman pendaki secara interaktif.

## 4.6 Pembahasan Umum

### 4.6.1 Keunggulan dan Keunikan Penelitian

Untuk memperjelas keunggulan penelitian ini, dilakukan perbandingan dengan penelitian terdahulu sebagaimana disajikan pada Tabel 16, yang menunjukkan perbedaan pada aspek metode, data, dan kinerja sistem.

Tabel 16 Perbandingan dengan penelitian terdahulu

Aspek	Penelitian Ini	[4]	[5]	[6]	[8]
Domain	Jalur Pendakian	Wisata Umum	Wisata Alam	Wisata	Berita

Metode Teks	SBERT	-	-	-	TF-IDF
Data Spasial	GPX	Koordinat	Koordinat	-	-
Pemahaman Semantik	✓	✗	✗	✗	Terbatas
Data Fusion	Numerik + Narasi	-	-	-	-
Bahasa Indonesia	✓ (Multilingual)	✓	✓	✓	✓
Precision@5	0.60	N/A	N/A	N/A	0.25

#### 4.6.2 Kontribusi Penelitian

Penelitian ini memberikan kontribusi orisinal berupa penerapan pertama model *Sentence-BERT* pada sistem rekomendasi jalur pendakian gunung di Indonesia. Pendekatan fusi data yang mengintegrasikan atribut spasial dari data GPX dengan representasi naratif deskriptif menghasilkan vektor semantik yang lebih representatif dibandingkan pendekatan berbasis atribut tunggal. Modifikasi pada tahap *preprocessing* melalui strategi stopword removal selektif terbukti efektif dalam mempertahankan informasi kontekstual yang relevan dengan domain pendakian. Hasil evaluasi menunjukkan superioritas metode yang diusulkan dengan peningkatan *Precision@5* sebesar 140% dibandingkan *baseline* TF-IDF, khususnya pada *query* yang membutuhkan pemahaman hubungan semantik implisit.

## 5 Kesimpulan

Penelitian ini berhasil mengembangkan sistem rekomendasi rute pendakian gunung berbasis *Content-Based Filtering* dengan *Sentence-BERT* (SBERT) dan *Cosine Similarity* yang terbukti efektif mengatasi masalah penyebaran informasi rute yang subjektif dan tidak terstruktur. Hasil pengujian empiris pada 39 jalur pendakian di Pulau Jawa menunjukkan rata-rata *Precision@5* sebesar 0,60, meningkat 140% dibandingkan metode TF-IDF yang hanya mencapai 0,25, dengan waktu respons pencarian 127-150 milidetik yang memenuhi standar aplikasi web interaktif. Kontribusi utama penelitian ini mencakup penerapan pertama model SBERT pada domain rekomendasi jalur pendakian di Indonesia serta strategi fusi data yang mengintegrasikan atribut spasial GPX dengan representasi naratif deskriptif untuk menghasilkan vektor semantik yang lebih representatif. Keterbatasan penelitian meliputi cakupan dataset yang terbatas serta belum terintegrasinya variabel dinamis *real-time* seperti kondisi cuaca, sehingga penelitian selanjutnya disarankan mengimplementasikan pendekatan *Hybrid Filtering* dan *fine-tuning model* SBERT dengan korpus spesifik domain pendakian untuk meningkatkan akurasi.

## Referensi

- [1] C. Salsabilla *et al.*, “Sistemasi: Jurnal Sistem Informasi Sistem Rekomendasi Wisata Kabupaten Pekalongan dengan *Content based Filtering* Pekalongan Regency Tourism Recommendation System with Content based Filtering,” 2025. [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [2] R. Stenzel, M. Lübbering, B. Ulusay, D. Uedelhoven, and R. Sifa, “*Matching Experts to Questions: A Comparison of Recommender Systems*,” 2022. [Online]. Available: <http://ceur-ws.org>
- [3] D. Pratiwi, A. Asrianda, and L. Rosnita, “Penerapan Metode *Content-based Filtering* dalam Sistem Rekomendasi Objek Wisata di Aceh Tamiang,” *Jurnal Ilmu Komputer dan Informatika*, Vol. 4, No. 2, pp. 85–96, Jan. 2025, DOI: 10.54082/jiki.169.
- [4] I. Suryadi, D. Hardan Gutama, D. Danianti, and A. Subhan Yazid, “Sistem Rekomendasi Tempat Wisata Alam di Gunung Kidul menggunakan *Algoritma Content based Filtering* berbasis Web,” 2025.
- [5] R. Faurina, E. Sitanggang, J. W. Supratman, and K. Limun, “Implementasi Metode *Content-based Filtering* dan *Collaborative Filtering* pada Sistem Rekomendasi Wisata di Bali *Implementation of Content-based Filtering and Collaborative Filtering Method in the Tourism Recommendation System in Bali*,” 2023.

- [6] 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.
- [7] I. Yulfiyani and M. Zakariyah, "Optimization of Tourism Destination Recommendations in Batang Regency using Content-based Filtering," 2024. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [8] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification using Sastrawi," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Jul. 2020. DOI: 10.1088/1757-899X/874/1/012017.
- [9] B. Juarto and A. S. Girsang, "International Journal on Informatics Visualization Journal Homepage : [www.joiv.org/index.php/joiv](http://www.joiv.org/index.php/joiv) International Journal on Informatics Visualization Neural Collaborative with Sentence BERT for News Recommender System." [Online]. Available: [www.joiv.org/index.php/joiv](http://www.joiv.org/index.php/joiv)
- [10] K. Sarode and S. Reddy Javaji, "Multi-BERT for Embeddings for Recommendation System."
- [11] Ibrahim Asad and Muhhammad Zakariyah, "Aplikasi Rekomendasi Pemesanan Paket Wisata menggunakan Metode Collaborative Filtering," *METIK JURNAL*, Vol. 7, No. 2, pp. 76–84, Dec. 2023, DOI: 10.47002/metik.v7i2.639.
- [12] M. Asako, Y. Tahara, A. Ohsuga, and Y. Sei, "Deep Learning-based Travel Time Estimation in Hiking with Consideration of Individual Walking Ability," *Cybernetics and Information Technologies*, Vol. 24, No. 4, pp. 3–21, Dec. 2024, DOI: 10.2478/cait-2024-0033.
- [13] H. Hartatik and A. Syafrianto, "Penerapan Model Sentence-BERT untuk Sistem Rekomendasi Buku berbasis Konten di Perpustakaan Digital," *Jurnal Dialektika Informatika (Detika)*, Vol. 6, No. 1, pp. 12–19, Nov. 2025, DOI: 10.24176/detika.v6i1.15916.
- [14] M. Abdul, H. Fathuddin, E. Prakarsa Mandyartha, and A. L. Nurlaili, "Penerapan Sentence-Bert dan Cosine Similarity untuk Pencarian Semantik Dokumen Skripsi dalam Format PDF," *R2J*, Vol. 8, No. 1, 2025, DOI: 10.38035/rj.v8i1.
- [15] F. Fitriaty, S. Amin, M. Musnaini, D. Elliyana, and M. H. Saputra, "Sustainable Strategy Toward Community Life Satisfaction in Heritage Tourism," *Jurnal Kepariwisata Indonesia: Jurnal Penelitian dan Pengembangan Kepariwisata Indonesia*, Vol. 18, No. 2, pp. 257–286, Nov. 2024, DOI: 10.47608/jki.v18i22024.257-286.
- [16] I. Ivanova and M. Wald, "Recommender Systems for Outdoor Adventure Tourism Sports: Hiking, Running and Climbing," *Human-Centric Intelligent Systems*, Vol. 3, No. 3, pp. 344–365, Sep. 2023, DOI: 10.1007/s44230-023-00033-3.
- [17] L. Luo, "RARE: Right Algorithm for the Right Errand; a Multi-Model Machine Learning-based Approach for Tourism Routes and Spots Recommendation," *PeerJ Comput SCI*, Vol. 11, 2025, DOI: 10.7717/peerj-cs.2791.