

**IMPLEMENTASI METODE CONTENT-BASED FILTERING
MENGUNAKAN ALGORITMA COSINE SIMILARITY
PADA SISTEM REKOMENDASI WISATA**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi *Sistem Informasi*



disusun oleh

ARDIAN PRIMADITYA

22.12.2408

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**IMPLEMENTASI METODE CONTENT-BASED FILTERING
MENGUNAKAN ALGORITMA COSINE SIMILARITY
PADA SISTEM REKOMENDASI WISATA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh
ARDIAN PRIMADITYA
22.12.2408

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**IMPLEMENTASI METODE CONTENT-BASED FILTERING
MENGUNAKAN ALGORITMA COSINE SIMILARITY PADA SISTEM
REKOMENDASI WISATA**

yang disusun dan diajukan oleh

Ardian Primaditya

22.12.2408

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 Februari 2026

Dosen Pembimbing,



Afrig Aminuddin, S.Kom., M.Eng., Ph.D.
NIK. 190302351

HALAMAN PENGESAHAN

SKRIPSI

IMPLEMENTASI METODE CONTENT-BASED FILTERING
MENGUNAKAN ALGORITMA COSINE SIMILARITY PADA SISTEM
REKOMENDASI WISATA

yang disusun dan diajukan oleh

Ardian Primaditya

22.12.2408

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anggit Dwi Hartanto, S.Kom., M.Kom.
NIK. 190302163

Dr. Wiji Nurastuti, S.E., M.T., CRP.
NIK. 190302258

Afrig Aminuddin, S.Kom., M.Eng., Ph.D.
NIK. 190302351



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ardian Primaditya
NIM : 22.12.2408

Menyatakan bahwa Skripsi dengan judul berikut:

Implementasi Metode Content Based Filtering menggunakan Algoritma Cosine Similarity pada Sistem Rekomendasi Wisata

Dosen Pembimbing : Afrig Aminuddin, S.Kom., M.Eng., Ph.D.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 23 Februari 2026

Yang Menyatakan,



Ardian Primaditya

KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi dengan judul “Implementasi Metode Content-Based Filtering Menggunakan Algoritma TF-IDF dan Cosine Similarity pada Sistem Rekomendasi Wisata”.

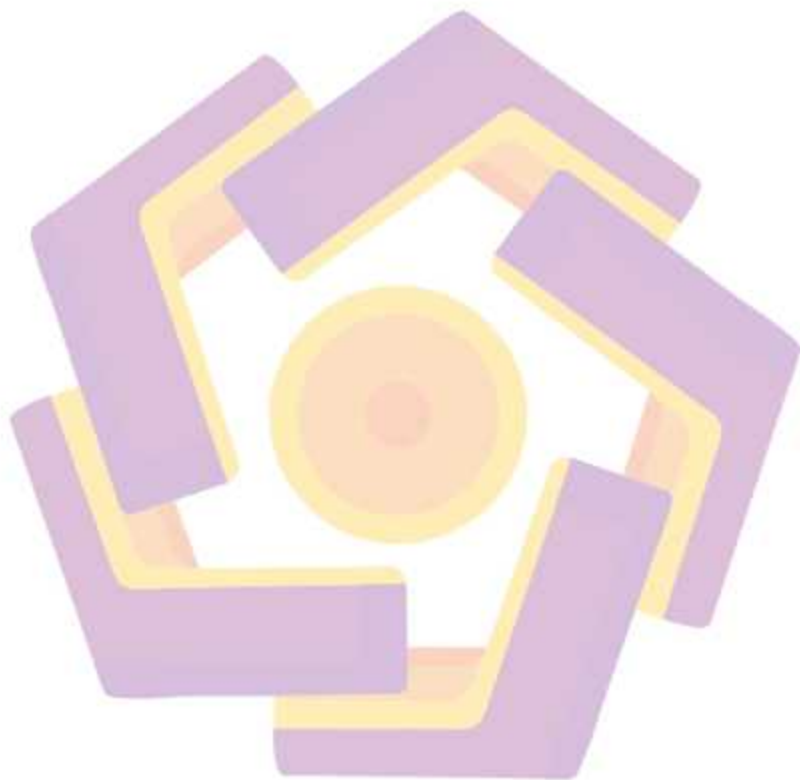
Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta. Penyusunan skripsi ini tidak lepas dari bimbingan, bantuan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta.
 2. Ibu Prof. Dr. Kusrini, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
 3. Bapak Anggit Dwi Hartanto, M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas AMIKOM Yogyakarta.
 4. Bapak Afrig Aminuddin, S.Kom., M.Eng., Ph.D., selaku Dosen Pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan serta arahan kepada penulis hingga terselesaikannya skripsi ini.
 5. Bapak Anggit Dwi Hartanto, M.Kom., dan Ibu Dr. Wiji Nurastuti, S.E., M.T., CRP., selaku Dewan Penguji yang telah meluangkan waktu untuk menguji serta memberikan kritik dan masukan yang sangat berharga bagi kesempurnaan skripsi ini.
 6. Kedua orang tua dan keluarga tercinta yang senantiasa memberikan doa tulus, kasih sayang, serta dukungan moral dan materiil yang tak terhingga.
- Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, penulis terbuka akan kritik dan saran yang membangun dari berbagai

pihak. Semoga skripsi ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan dan pihak-pihak yang membutuhkan.

Yogyakarta, 23 Februari 2026

Penulis

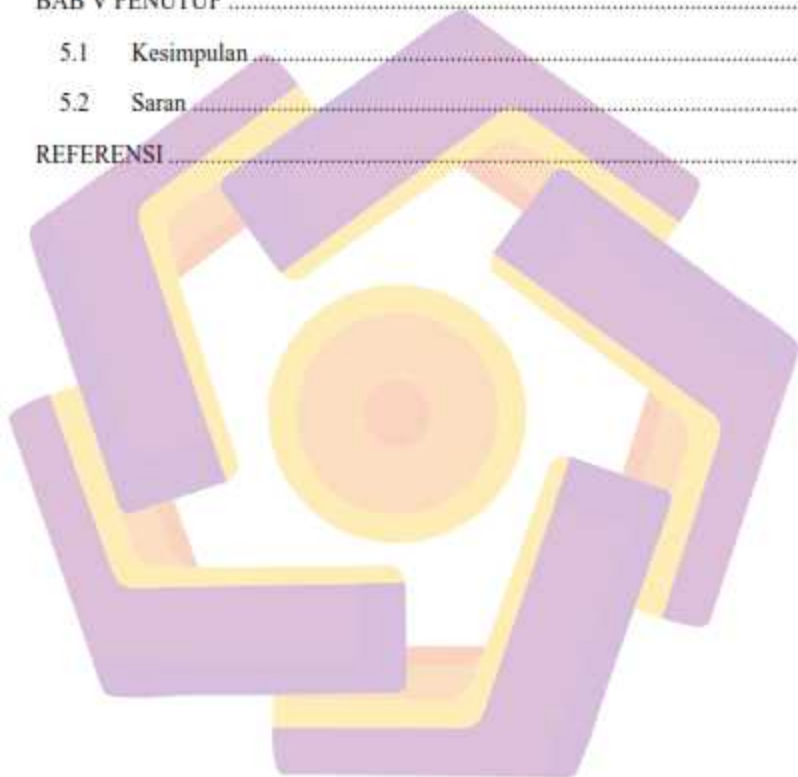


DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xvii
<i>ABSTRACT</i>	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.1.1 Penelitian Terdahulu	6
2.1.2 Keaslian Penelitian.....	9

2.2	Dasar Teori.....	12
2.2.1	Sistem Rekomendasi.....	12
2.2.2	Content Based Filtering	17
2.2.3	TF-IDF	19
2.2.4	Cosine Similarity	22
2.2.5	Confusion Matrix.....	26
BAB III METODE PENELITIAN		31
3.1	Objek Penelitian.....	31
3.2	Alur Penelitian	31
3.2.1	Pengumpulan Data	31
3.2.2	Preprocessing Data.....	32
3.2.3	TF-IDF	32
3.2.4	Cosine Similarity	33
3.2.5	Hasil Rekomendasi	34
3.2.6	Evaluasi.....	35
3.3	Alat dan Bahan.....	35
3.3.1	Perangkat Keras (Hardware).....	35
3.3.2	Perangkat Lunak (Software).....	36
3.3.3	Bahan Penelitian.....	36
BAB IV HASIL DAN PEMBAHASAN		37
4.1	Pengumpulan Data	37
4.2	Preprocessing Data.....	38
A.	Data Cleaning.....	39
B.	Text Preprocessing.....	44
4.3	Implementasi TF-IDF	56

4.4	Implementasi Cosine Similarity.....	58
4.5	Hasil Rekomendasi	61
4.6	Evaluasi.....	62
4.7	Implementasi Antarmuka.....	68
BAB V PENUTUP		72
5.1	Kesimpulan.....	72
5.2	Saran.....	73
REFERENSI		75

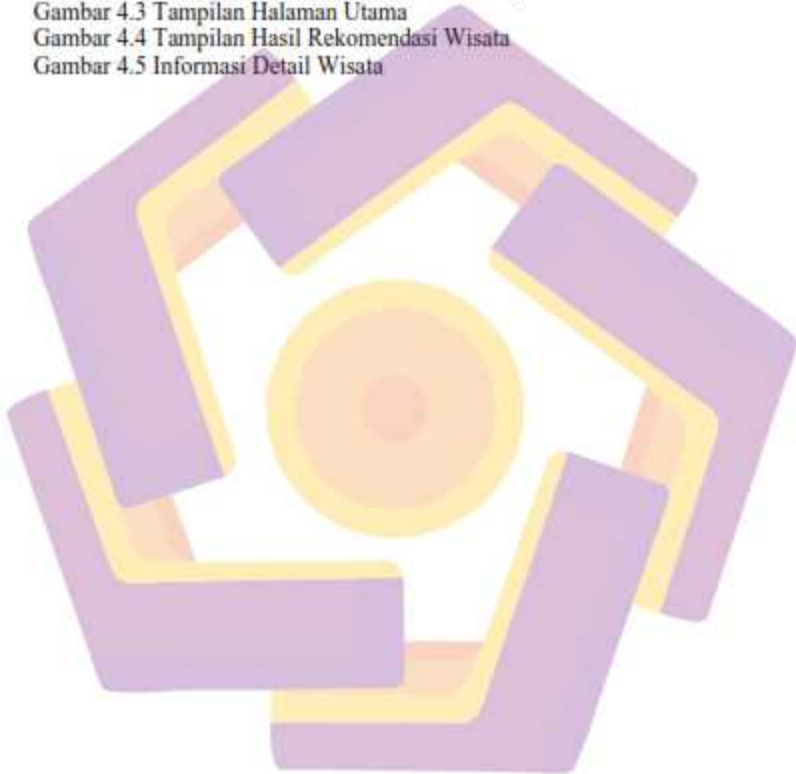


DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	10
Tabel 2.2. Confusion Matrix	27
Tabel 4.1 Atribut Data	39
Tabel 4.2 Perubahan Data	43
Tabel 4.3 Transformasi Penamaan Atribut Data	44
Tabel 4.4 Penggabungan Kolom	45
Tabel 4.5 Penjelasan Alur Pemrosesan	47
Tabel 4.6 Hasil Case Folding	48
Tabel 4.7 Hasil Punctuation Removal	50
Tabel 4.8 Tokenization	52
Tabel 4.9 Hasil Stopword Removal	54
Tabel 4.10 Stemming	57
Tabel 4.11 Matriks TF-IDF	59
Tabel 4.12 Perhitungan Cosine Similarity	61
Tabel 4.13 Hasil Rekomendasi	63
Tabel 4.14 Daftar Kategori Wisata Alam	64
Tabel 4.15 Hasil Rekomendasi Air Terjun Jumog	66
Tabel 4.16 Hasil Keseluruhan	68

DAFTAR GAMBAR

Gambar 2.1 Collaborative Filtering	14
Gambar 2.2 Content Based Filtering	15
Gambar 2.3 Cosine Similarity	25
Gambar 3.1 Alur Penelitian	31
Gambar 4.1 Alur Preprocessing Data	41
Gambar 4.2 Visualisasi Heatmap Cosine Similarity	62
Gambar 4.3 Tampilan Halaman Utama	70
Gambar 4.4 Tampilan Hasil Rekomendasi Wisata	70
Gambar 4.5 Informasi Detail Wisata	72



DAFTAR LAMBANG DAN SINGKATAN

A	Vektor bobot untuk dokumen.
B	Vektor bobot untuk kueri.
d	Dokumen.
t	Kata (term).
$df(t)$	Document Frequency.
$f_{t,d}$	Jumlah kemunculan kata t dalam dokumen d .
$tf_{t,d}$	Nilai Term Frequency untuk kata t dalam dokumen d .
$idf(t)$	Nilai Inverse Document Frequency untuk kata t .
$W_{t,d}$	Bobot akhir kata t terhadap dokumen d .
N	Total jumlah dokumen dalam korpus.
n_d	Total jumlah seluruh kata dalam dokumen d .
$Sim(A, B)$	Nilai kemiripan antara vektor A dan B .
θ	Sudut yang terbentuk antara dua vector.
$ A $	Panjang dari vektor A .
CBF	Content-Based Filtering.
CF	Collaborative Filtering.
FN	False Negative.
FP	False Positive.
IDE	Integrated Development Environment.
IDF	Inverse Document Frequency.
TF	Term Frequency.
TF-IDF	Term Frequency - Inverse Document Frequency.
TN	True Negative.
TP	True Positive.
VSM	Vector Space Model.

DAFTAR ISTILAH



Ground Truth	Data acuan yang merepresentasikan kondisi atau kebenaran yang sesungguhnya, digunakan sebagai pembanding dalam evaluasi.
Tokenization	Proses memecah kalimat atau kumpulan teks menjadi potongan unit terkecil.
Collaborative Filtering	Metode sistem rekomendasi yang memprediksi preferensi pengguna berdasarkan kesamaan perilaku dari pengguna lain.
Missing Values	Kondisi data atau nilai yang kosong.
Term Frequency (TF)	Metrik yang mengukur seberapa sering sebuah kata muncul dalam suatu dokumen tertentu.
Preprocessing	Tahap awal pengolahan data mentah untuk mempersiapkan data agar terstruktur sebelum diproses algoritma.
Content-Based Filtering (CBF)	Metode sistem rekomendasi yang memberikan saran berdasarkan kesesuaian konten dengan preferensi pengguna.
False Negative (FN)	Kesalahan prediksi di mana sistem gagal merekomendasikan item yang sebenarnya relevan.
Recall	Metrik yang mengukur rasio keberhasilan sistem dalam menemukan kembali informasi yang relevan.

Library Sastrawi	Pustaka dalam bahasa pemrograman Python untuk memproses tata bahasa Indonesia,
Cosine Similarity	Metode untuk mengukur tingkat kemiripan.
Data Sekunder	Data yang diperoleh tidak secara langsung dari objek penelitian.
True Positive (TP)	Kondisi keberhasilan sistem secara tepat memprediksi item yang relevan.
Top-N Recommendation	Daftar sejumlah N item teratas.
Text Mining	Proses penambangan data yang berfokus pada pengolahan data teks untuk menemukan pola.
Data Cleaning	Proses pembersihan data yang tidak valid, tidak lengkap, atau tidak sesuai.
Case Folding	Tahap dalam <i>preprocessing</i> untuk menyeragamkan seluruh huruf dalam dokumen teks menjadi huruf kecil.
Hybrid Filtering	Pendekatan yang mengombinasikan dua atau lebih metode rekomendasi.
False Positive (FP)	Kesalahan prediksi di mana sistem merekomendasikan item yang sebenarnya tidak relevan bagi pengguna.
Inverse Document Frequency (IDF)	Perhitungan bobot yang mengukur seberapa umum atau langka sebuah kata.

Confusion Matrix	Metode evaluasi berbentuk tabel matriks untuk mengukur kinerja algoritma klasifikasi dengan membandingkan hasil prediksi sistem dengan kondisi sebenarnya.
Snake_case	Format penulisan variabel di mana spasi digantikan dengan garis bawah.
Vector Space Model (VSM)	Representasi matematis di mana dokumen teks diubah menjadi bentuk vektor numerik agar dapat dihitung jarak kemiripannya.
Stemming	Proses transformasi kata berimbuhan kembali menjadi bentuk kata dasarnya.
Stopword Removal	Proses eliminasi kata-kata umum yang sering muncul namun memiliki kandungan informasi rendah.
Data Mining	Proses pengumpulan dan pengolahan data.
Corpus	Sekumpulan dokumen atau data teks yang digunakan sebagai bahan analisis.
Information Retrieval	Temu kembali informasi, dokumen yang relevan dari koleksi dokumen yang besar.
Dot Product:	Operasi perkalian titik.
Noise	Data gangguan yang tidak relevan.
Feature Engineering	Proses teknis penggabungan atau modifikasi atribut data untuk menciptakan fitur baru yang lebih representatif bagi model.

Precision

Metrik evaluasi yang mengukur tingkat ketepatan antara informasi yang diminta pengguna dengan jawaban yang diberikan oleh sistem.

Term Frequency - Inverse Document Frequency (TF-IDF)

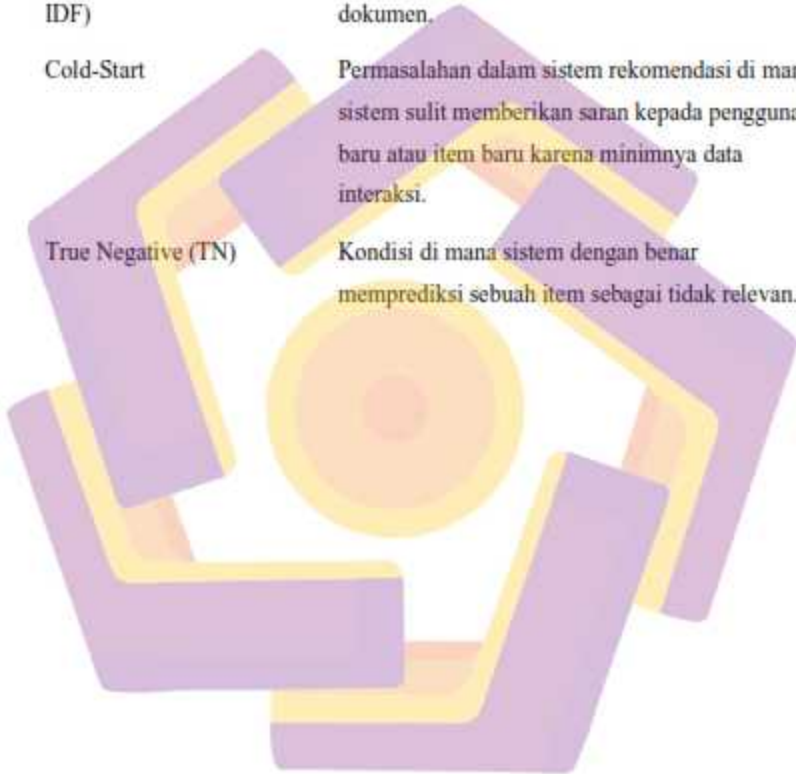
Metode pembobotan untuk merepresentasikan tingkat kepentingan sebuah kata terhadap dokumen.

Cold-Start

Permasalahan dalam sistem rekomendasi di mana sistem sulit memberikan saran kepada pengguna baru atau item baru karena minimnya data interaksi.

True Negative (TN)

Kondisi di mana sistem dengan benar memprediksi sebuah item sebagai tidak relevan.



INTISARI

Penelitian ini membahas implementasi metode *Content Based Filtering* (CBF) dalam sistem rekomendasi wisata Kabupaten Karanganyar. Tujuan penelitian adalah merancang sistem yang mampu memberikan rekomendasi destinasi sesuai preferensi pengguna berdasarkan deskripsi wisata. Metode yang digunakan adalah algoritma *Term Frequency Inverse Document Frequency* (TF-IDF) untuk menghitung bobot kata pada deskripsi wisata dan *Cosine Similarity* untuk mengukur tingkat kesamaan antar destinasi. Dataset diperoleh dari situs Open Data Kabupaten Karanganyar dan melalui tahap pembersihan data yang meliputi data cleaning dan text preprocessing, sehingga menghasilkan data yang konsisten dan siap dianalisis. Evaluasi kinerja sistem dilakukan menggunakan metode *Confusion Matrix* dengan menghitung nilai *Precision* dan *Recall*. Hasil penelitian menunjukkan bahwa sistem dapat memberikan top-5 rekomendasi wisata berdasarkan kemiripan deskripsi. Evaluasi menghasilkan nilai *Precision* sebesar 0,744 yang menunjukkan sebagian besar rekomendasi sesuai, sedangkan nilai *Recall* sebesar 0,317 masih tergolong rendah. Rendahnya nilai *Recall* dipengaruhi oleh kategori wisata yang digunakan sebagai data pembandingan bersifat umum, sehingga sistem cenderung merekomendasikan wisata dengan deskripsi mirip meskipun berbeda kategori. Selain itu, pengujian hanya terbatas pada dataset yang tersedia sehingga kurang merepresentasikan kondisi lapangan. Secara keseluruhan, sistem ini dapat membantu wisatawan memperoleh rekomendasi destinasi yang relevan dan personal.

Kata kunci: Content-Based Filtering, Cosine Similarity, Sistem Rekomendasi, TF-IDF, Text Mining.

ABSTRACT

This study discusses the implementation of the *Content-Based Filtering* (CBF) method in a tourism recommendation system for Karanganyar Regency. The purpose of this research is to design a system capable of providing destination recommendations that match user preferences based on tourism descriptions. The method applied uses the *Term Frequency-Inverse Document Frequency* (TF-IDF) algorithm to calculate word weights in tourism descriptions and *Cosine Similarity* to measure the level of similarity between destinations. The dataset was obtained from the Karanganyar Regency Open Data website and underwent data cleaning and text preprocessing stages to produce consistent and analysis-ready data. System performance was evaluated using the *Confusion Matrix* method by calculating *Precision* and *Recall* values. The results show that the system is able to generate the top five tourism recommendations based on description similarity. The evaluation produced a *Precision* value of 0.744, indicating that most recommendations are relevant, while the *Recall* value of 0.317 is considered relatively low. The low Recall is influenced by the use of general tourism categories as comparison data, causing the system to recommend destinations with similar descriptions even if they belong to different categories. In addition, the evaluation was limited to the available dataset, which may not fully represent real-world conditions. Overall, the system can assist tourists in obtaining relevant and personalized destination recommendations.

Keyword: Content-Based Filtering, Cosine Similarity, Recommendation System, TF-IDF, Text Mining.