

Optimasi Rekomendasi Destinasi Wisata Menggunakan Metode CBF Dan GIS Dengan Pendekatan Iterative

Ahmad Faidh Rifai¹, Hadi Zakaria^{*2}

^{1,2} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang

e-mail: ¹ahmadfaidh2003@gmail.com, ²dosen00274@unpam.ac.id

^{*}Correspondence author

Abstrak

Dinas Pariwisata dan Ekonomi Kreatif Provinsi DKI Jakarta memiliki peran dalam pengelolaan dan pengembangan destinasi wisata, termasuk penyediaan informasi melalui Tourist Information Center (TIC). Namun, rekomendasi destinasi yang masih dilakukan secara manual belum mampu memberikan rekomendasi sesuai preferensi wisatawan, sementara umpan balik wisatawan belum terdokumentasi secara terpusat. Penelitian ini bertujuan mengembangkan sistem rekomendasi destinasi wisata berbasis web yang menyediakan rekomendasi sesuai preferensi, informasi geografis, dan fitur umpan balik untuk mendukung evaluasi destinasi. Metode Content-Based Filtering (CBF) digunakan berdasarkan kemiripan preferensi pengguna dengan karakteristik destinasi melalui semantic embedding dan cosine similarity. Geographic Information System (GIS) digunakan untuk menyajikan informasi lokasi dan rute, sedangkan pengembangan sistem menerapkan pendekatan Iterative Development. Kinerja sistem dievaluasi menggunakan Precision dan Recall pada lima skenario query dengan $K=1$, $K=3$, $K=5$, dan $K=10$. Hasil evaluasi menunjukkan rata-rata precision sebesar 1,00 pada $K=1$, $K=3$, dan $K=5$, serta 0,86 pada $K=10$. Rata-rata recall meningkat dari 0,09 pada $K=1$ menjadi 0,69 pada $K=10$. Evaluasi pengguna terhadap 20 responden memperoleh rata-rata 83,4% dengan kategori sangat kuat. Sistem dapat membantu wisatawan menemukan destinasi sesuai preferensi serta menyediakan informasi lokasi, rute, dan umpan balik.

Kata kunci—Content Based Filtering, Geographic Information System, Iterative Development, Sistem Rekomendasi Destinasi Wisata

1. PENDAHULUAN

Perkembangan teknologi informasi telah mengubah cara wisatawan memperoleh informasi dan menentukan destinasi wisata. Semakin banyak dan beragamnya destinasi yang tersedia memberikan lebih banyak pilihan bagi wisatawan, tetapi pada saat yang sama dapat menyulitkan wisatawan dalam menentukan destinasi yang paling sesuai dengan kebutuhan dan preferensi masing-masing. Informasi umum mengenai destinasi wisata belum tentu cukup untuk mendukung pengambilan keputusan yang bersifat personal. Oleh karena itu, sistem rekomendasi semakin banyak dimanfaatkan untuk menyaring informasi yang tersedia dan memberikan alternatif destinasi berdasarkan karakteristik serta preferensi pengguna [1].

Dinas Pariwisata dan Ekonomi Kreatif Provinsi DKI Jakarta merupakan instansi pemerintah yang memiliki tanggung jawab dalam pengelolaan dan pengembangan destinasi wisata di wilayah DKI Jakarta. Berdasarkan hasil observasi terhadap proses pelayanan informasi pada Tourist Information Center (TIC), rekomendasi destinasi masih diberikan secara manual oleh petugas berdasarkan pengetahuan dan informasi yang tersedia. Proses tersebut belum memiliki mekanisme yang mampu menyesuaikan rekomendasi secara dinamis dengan preferensi masing-masing wisatawan. Wisatawan juga masih harus menentukan sendiri relevansi suatu destinasi berdasarkan informasi umum yang diberikan. Kondisi tersebut menunjukkan perlunya suatu sistem yang dapat membantu wisatawan dalam mencari dan memilih destinasi sesuai dengan preferensi pribadi.

Permasalahan lain berkaitan dengan pengelolaan umpan balik wisatawan. Informasi mengenai pengalaman pengunjung di suatu destinasi wisata dapat menjadi masukan yang berharga bagi pengelola destinasi untuk memahami persepsi dan tingkat kepuasan wisatawan. Namun, apabila umpan balik tidak dikumpulkan dan dikelola melalui suatu platform yang terpusat, informasi tersebut menjadi lebih sulit dimanfaatkan untuk keperluan evaluasi dan pengambilan keputusan [2]. Pada Dinas Pariwisata dan Ekonomi Kreatif Provinsi DKI Jakarta,

umpan balik wisatawan belum terdokumentasi secara sistematis dalam suatu platform terpusat sehingga pemanfaatannya sebagai sumber evaluasi destinasi wisata masih terbatas. Oleh karena itu, diperlukan suatu sistem yang tidak hanya membantu wisatawan memperoleh rekomendasi yang sesuai dengan preferensi, tetapi juga menyediakan media digital untuk mengumpulkan umpan balik secara terstruktur.

Sistem rekomendasi dapat digunakan sebagai salah satu pendekatan untuk membantu wisatawan memperoleh alternatif destinasi yang lebih relevan dengan preferensi mereka. Salah satu metode yang dapat diterapkan adalah *Content-Based Filtering* (CBF). Metode ini menghasilkan rekomendasi dengan menentukan relevansi antara profil atau preferensi pengguna dengan karakteristik item yang tersedia [3]. Dalam bidang pariwisata, karakteristik destinasi dapat direpresentasikan melalui berbagai atribut, seperti kategori, deskripsi, aktivitas, dan informasi lain yang menggambarkan suatu destinasi. Beberapa penelitian telah menerapkan CBF pada sistem rekomendasi pariwisata dan menunjukkan bahwa pendekatan berbasis konten dapat digunakan untuk menghasilkan rekomendasi destinasi berdasarkan preferensi pengguna [4] [5].

Representasi informasi destinasi merupakan salah satu aspek penting dalam CBF karena kualitas representasi dapat memengaruhi kemampuan sistem dalam menentukan kemiripan antar konten. Representasi berbasis teks seperti *Term Frequency-Inverse Document Frequency* (TF-IDF) telah banyak digunakan dalam sistem rekomendasi karena memberikan bobot pada kata berdasarkan kemunculannya dalam dokumen. Namun, pendekatan tersebut lebih berfokus pada hubungan pada tingkat kata sehingga belum sepenuhnya mampu menangkap kesamaan semantik dalam teks [6]. Pendekatan berbasis *embedding* dapat merepresentasikan informasi tekstual ke dalam bentuk vektor sehingga hubungan semantik antar teks dapat dipertimbangkan dalam proses pengukuran kemiripan. Pada penelitian ini, karakteristik destinasi dan preferensi pengguna direpresentasikan menggunakan model *multilingual-e5-small*, sedangkan *cosine similarity* digunakan untuk mengukur tingkat kemiripan antara representasi yang dihasilkan.

Selain rekomendasi berdasarkan preferensi, informasi mengenai lokasi dan rute menuju destinasi juga penting dalam membantu wisatawan menentukan destinasi yang akan dikunjungi. *Geographic Information System* (GIS) dapat dimanfaatkan untuk menyajikan informasi spasial, lokasi destinasi, dan rute perjalanan sehingga wisatawan dapat memperoleh informasi geografis yang lebih kontekstual. Penelitian mengenai sistem rekomendasi pariwisata berbasis lokasi menunjukkan bahwa informasi geografis dapat menjadi salah satu faktor penting dalam meningkatkan personalisasi dan relevansi rekomendasi destinasi [7]. Oleh karena itu, penelitian ini mengintegrasikan GIS dengan sistem rekomendasi sehingga wisatawan tidak hanya memperoleh rekomendasi destinasi berdasarkan preferensi, tetapi juga informasi lokasi dan rute menuju destinasi yang dipilih.

Beberapa penelitian terdahulu telah menerapkan metode *Content-Based Filtering* (CBF) untuk membangun sistem rekomendasi destinasi wisata. Penelitian [8] menerapkan CBF dengan algoritma TF-IDF dan *cosine similarity* untuk memberikan rekomendasi wisata di Kabupaten Pekalongan berdasarkan kategori destinasi. Namun, representasi konten yang digunakan dalam penelitian tersebut berfokus pada informasi kategori sehingga karakteristik tekstual lain dari destinasi belum dimanfaatkan secara langsung dalam proses pengukuran kemiripan. Penelitian lain [9] mengembangkan sistem rekomendasi wisata di Boyolali menggunakan CBF dengan TF-IDF dan *cosine similarity*. Penelitian tersebut menggabungkan nama destinasi, kategori, fasilitas, dan deskripsi sebagai representasi konten untuk memperkaya informasi yang digunakan dalam proses rekomendasi. Namun, representasi konten masih dibangun menggunakan TF-IDF yang merepresentasikan teks berdasarkan bobot kata dalam dokumen. Sementara itu, penelitian [10] mengembangkan sistem rekomendasi destinasi wisata yang menggabungkan CBF dengan analisis fitur geospasial menggunakan TF-IDF, *cosine similarity*, dan *haversine* untuk mempertimbangkan konteks geografis dalam rekomendasi. Meskipun demikian, pendekatan tersebut menempatkan analisis geografis sebagai bagian dari proses rekomendasi, sedangkan penelitian ini memisahkan fungsi rekomendasi dan informasi geografis, yaitu menggunakan semantic embedding dan cosine similarity untuk menentukan pemeringkatan rekomendasi,

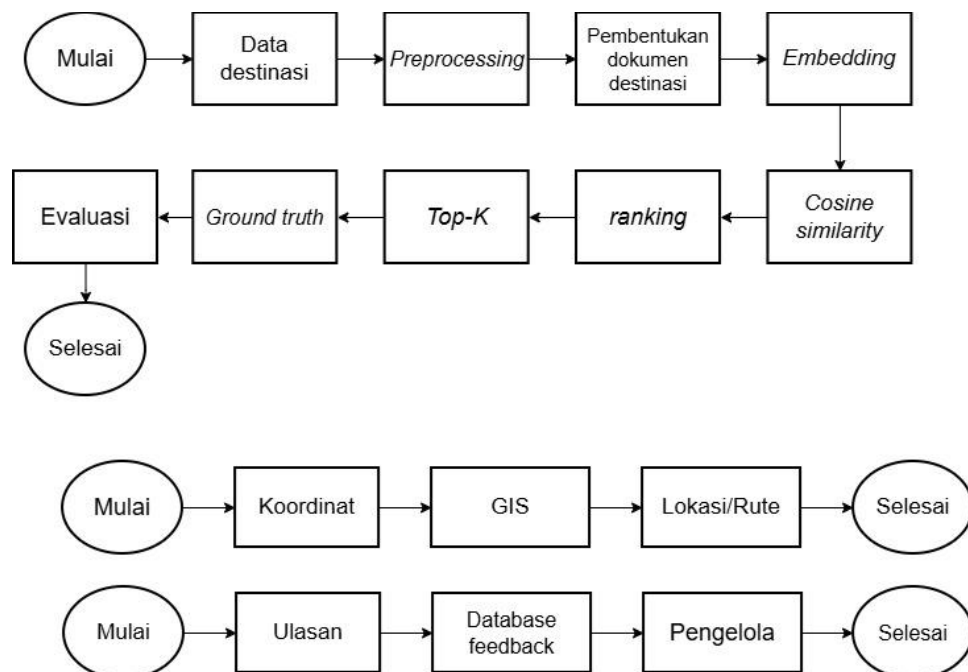
sementara GIS digunakan untuk menyediakan informasi lokasi dan rute menuju destinasi yang telah direkomendasikan.

Berdasarkan penelitian-penelitian tersebut, masih terdapat ruang untuk mengembangkan sistem rekomendasi destinasi wisata menggunakan pendekatan representasi yang berbeda, yaitu semantic embedding, untuk menangkap hubungan semantik antara preferensi wisatawan dan karakteristik destinasi. Oleh karena itu, penelitian ini mengembangkan sistem rekomendasi menggunakan CBF dengan model *multilingual-e5-small* untuk merepresentasikan nama, kategori, kota, aktivitas, dan deskripsi destinasi ke dalam ruang vektor, kemudian menggunakan *cosine similarity* untuk menentukan tingkat kemiripan dan pemeringkatan rekomendasi. Selain itu, penelitian ini mengintegrasikan GIS sebagai komponen pendukung untuk menyediakan informasi lokasi dan rute serta menyediakan fitur ulasan sebagai media pengumpulan umpan balik wisatawan dalam satu sistem berbasis web. Sistem dikembangkan menggunakan pendekatan *Iterative Development* yang memungkinkan proses pengembangan dilakukan secara bertahap melalui beberapa iterasi dan disesuaikan dengan kebutuhan sistem.[11]

Penelitian ini bertujuan membantu wisatawan menemukan destinasi yang sesuai dengan preferensi pribadi sekaligus menyediakan informasi lokasi dan rute dalam satu sistem. Selain itu, sistem dirancang untuk menyediakan mekanisme umpan balik yang dapat membantu Dinas Pariwisata dan Ekonomi Kreatif Provinsi DKI Jakarta memperoleh informasi mengenai pengalaman wisatawan terhadap destinasi. Kinerja metode rekomendasi dievaluasi menggunakan *Precision@K* dan *Recall@K* untuk mengukur ketepatan dan cakupan rekomendasi yang dihasilkan. Dengan demikian, penelitian ini diharapkan dapat menghasilkan sistem yang mendukung pencarian destinasi wisata secara personal sekaligus menyediakan informasi pendukung bagi evaluasi dan pengembangan destinasi wisata di DKI Jakarta.

2. METODE PENELITIAN

Penelitian ini dilaksanakan melalui serangkaian tahapan yang sistematis, yaitu dimulai dengan pengumpulan data destinasi, *preprocessing*, pembentukan dokumen, *embedding*, *cosine similarity*, *ranking*, *top-k*, *ground truth* dan evaluasi hasil yang diberikan. Alur proses penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur proses penelitian

Pengumpulan Data

Data yang digunakan dalam penelitian ini berjumlah 178 destinasi wisata di wilayah DKI Jakarta yang dikumpulkan pada tahun 2025. Data destinasi diperoleh dari situs resmi Dinas Pariwisata dan Ekonomi Kreatif DKI Jakarta, *Google Maps*, serta sumber informasi daring yang relevan. Data yang dikumpulkan meliputi nama destinasi, deskripsi, alamat, kota, kategori, aktivitas, latitude, dan longitude, yang digunakan sebagai data dalam proses rekomendasi menggunakan metode *Content-Based Filtering* (CBF). Selain pengumpulan data destinasi, dilakukan observasi dan wawancara pada proses pelayanan informasi di *Tourist Information Center* (TIC) untuk memperoleh informasi mengenai proses pemberian rekomendasi dan kebutuhan sistem. Studi literatur dilakukan untuk memperoleh landasan teoritis dan penelitian terdahulu yang relevan dengan sistem rekomendasi, *semantic embedding*, CBF, dan GIS. Data destinasi yang didapat terdiri atas beberapa atribut yang menggambarkan karakteristik destinasi, meliputi nama destinasi, deskripsi, alamat, kota, kategori, aktivitas, serta koordinat geografis berupa latitude dan longitude. Selain itu, studi literatur dilakukan dengan mempelajari penelitian terdahulu dan sumber ilmiah yang relevan dengan sistem rekomendasi pariwisata, *Content-Based Filtering*, representasi *embedding*, *cosine similarity*, dan *Geographic Information System* (GIS). Data destinasi yang diperoleh selanjutnya digunakan sebagai masukan pada tahap *preprocessing* sebelum dilakukan proses *embedding* dan perhitungan kemiripan.

Preprocessing Data

Tahap *preprocessing* dilakukan untuk membersihkan dan menyeragamkan data tekstual sehingga dapat digunakan sebagai masukan pada proses vektorisasi. *Preprocessing* diperlukan untuk menyiapkan data teks sebelum digunakan pada tahap pengolahan atau ekstraksi fitur lebih lanjut [12]. Tahapan ini meliputi *case folding*, *remove special*, dan *normalize space*. *Case folding* dilakukan dengan mengubah seluruh huruf kapital menjadi huruf kecil agar format penulisan teks menjadi seragam. Selanjutnya, *remove special* digunakan untuk menghapus karakter khusus yang tidak memiliki makna semantik dalam pengolahan teks, seperti tanda baca, simbol, dan karakter non-alfanumerik. Sementara itu, *normalize space* dilakukan untuk menormalkan penggunaan spasi dengan menghapus spasi berlebih sehingga teks memiliki struktur yang lebih konsisten dan siap digunakan pada tahap pengolahan berikutnya.

Preprocessing Data pada penelitian ini dilakukan secara terbatas untuk mempertahankan informasi semantik yang terdapat pada data teks. *Stopword removal* dan *stemming* tidak dilakukan karena model yang digunakan adalah *multilingual-e5-small* berbasis *Transformer*, yang memanfaatkan konteks kata dalam proses pembentukan representasi *embedding*. Tokenisasi dilakukan oleh *tokenizer* bawaan model *multilingual-e5-small* pada tahap vektorisasi. Informasi lokasi seperti “Jakarta Selatan” serta atribut kategori dan aktivitas tetap dipertahankan karena dapat memberikan konteks penting dalam merepresentasikan karakteristik destinasi. Dengan demikian, *preprocessing* pada penelitian ini tidak menerapkan tahapan pengolahan teks secara agresif seperti yang umum digunakan pada pendekatan berbasis TF-IDF.

Pembentukan Dokumen Destinasi

Setelah tahap *preprocessing*, data destinasi dibentuk menjadi dokumen teks yang merepresentasikan karakteristik setiap destinasi. Dokumen destinasi disusun dengan menggabungkan beberapa atribut yang relevan, yaitu nama destinasi, kategori, kota, aktivitas, dan deskripsi. Penggabungan atribut dilakukan agar informasi yang digunakan dalam proses rekomendasi tidak hanya bergantung pada deskripsi destinasi, tetapi juga mempertimbangkan karakteristik dan aktivitas yang tersedia pada destinasi tersebut. Setiap atribut yang telah melalui *preprocessing* kemudian digabungkan ke dalam satu teks dengan format yang konsisten. Dokumen tersebut selanjutnya digunakan sebagai masukan pada model *embedding* untuk menghasilkan representasi vektor dari setiap destinasi. Dengan demikian, setiap destinasi memiliki representasi semantik yang mencakup informasi mengenai nama, kategori, lokasi kota, aktivitas, dan deskripsinya.

Embedding

Data teks yang telah melalui tahap preprocessing selanjutnya diubah menjadi representasi numerik menggunakan model *intfloat/multilingual-e5-small*. Model E5 dirancang untuk menghasilkan representasi vektor dari teks yang dapat digunakan pada berbagai tugas seperti *retrieval*, pengelompokan, dan klasifikasi, sehingga teks dapat dibandingkan berdasarkan representasi numeriknya [13]. Pada penelitian ini, proses *embedding* diterapkan pada teks preferensi pengguna dan teks karakteristik destinasi wisata menggunakan model *embedding* yang sama, tetapi diberikan *instruction prefix* yang berbeda sesuai dengan fungsinya. Teks preferensi pengguna diberikan prefix *query*:, sedangkan teks karakteristik destinasi diberikan prefix *passage*: sebelum diproses oleh model. Penggunaan prefix tersebut membedakan teks yang berperan sebagai permintaan pencarian (*query*) dan teks yang berperan sebagai dokumen yang dicari (*passage*). Dengan demikian, format input yang digunakan adalah *query*: [preferensi pengguna] untuk teks preferensi pengguna dan *passage*: [informasi destinasi] untuk teks destinasi. Representasi dari kedua jenis teks tersebut kemudian dihasilkan dalam ruang vektor yang sama sehingga dapat dibandingkan menggunakan *cosine similarity*. Model *multilingual-e5-small* menghasilkan representasi *embedding* berdimensi 384 [14]. Vektor preferensi pengguna kemudian dibandingkan dengan vektor masing-masing destinasi pada tahap perhitungan *cosine similarity* untuk menentukan tingkat kemiripan antara preferensi pengguna dan karakteristik destinasi.

Cosine Similarity

Setelah proses vektorisasi menghasilkan representasi *embedding* untuk preferensi pengguna dan setiap destinasi wisata, tahap selanjutnya adalah menghitung tingkat kemiripan antara kedua representasi tersebut menggunakan *cosine similarity*. Metode ini digunakan untuk mengukur kesamaan arah antara vektor preferensi pengguna dan vektor karakteristik destinasi. Nilai *cosine similarity* yang semakin mendekati 1 menunjukkan tingkat kemiripan yang semakin tinggi, sehingga destinasi tersebut memiliki tingkat relevansi yang lebih besar terhadap preferensi pengguna [15]. Nilai *cosine similarity* dihitung menggunakan persamaan (1)

$$COS_{(A,B)} = \frac{A.B}{|A|.|B|} = \frac{\sum_{i=1}^n U_i \times V_i}{\sqrt{\sum_{i=1}^n (U_i)^2} \times \sqrt{\sum_{i=1}^n (V_i)^2}} \quad (1)$$

Dengan A atau U merupakan vektor *embedding* dari preferensi pengguna, sedangkan B atau V merupakan vektor *embedding* dari destinasi wisata. Nilai U_i dan V_i menunjukkan nilai *embedding* pada dimensi ke- i , sedangkan merupakan jumlah dimensi *embedding*. Pada penelitian ini, setiap *embedding* memiliki 384 dimensi, sehingga perhitungan dilakukan terhadap seluruh 384 nilai pada masing-masing vektor.

Ranking

Setelah nilai *cosine similarity* diperoleh antara vektor *query* wisatawan dan vektor masing-masing destinasi, setiap destinasi memiliki nilai yang menunjukkan tingkat kemiripannya terhadap *query*. Nilai tersebut kemudian digunakan sebagai dasar dalam proses perankingan hasil rekomendasi. Perankingan dilakukan dengan mengurutkan seluruh destinasi berdasarkan nilai *cosine similarity* secara menurun (*descending*), yaitu dari nilai kemiripan terbesar hingga terkecil. Destinasi dengan nilai kemiripan terbesar ditempatkan pada peringkat teratas karena memiliki representasi semantik yang paling dekat dengan *query* wisatawan. Sebaliknya, destinasi dengan nilai kemiripan yang lebih rendah ditempatkan pada peringkat berikutnya.

Top-K

Setelah seluruh destinasi diurutkan berdasarkan nilai *cosine similarity* secara menurun, sistem memilih sebanyak (K) destinasi dengan peringkat tertinggi sebagai hasil rekomendasi. Nilai (K) menunjukkan jumlah destinasi yang akan ditampilkan kepada wisatawan. Dengan

mekanisme ini, destinasi yang memiliki tingkat kemiripan semantik tertinggi terhadap query akan diprioritaskan untuk ditampilkan.

Ground Truth

Ground truth dalam penelitian ini diperoleh melalui *manual relevance analysis* yang dilakukan oleh peneliti. Analisis dilakukan dengan menilai setiap destinasi berdasarkan kesesuaiannya terhadap kebutuhan atau intent yang terkandung dalam query wisatawan dengan karakteristik destinasi. Setiap destinasi kemudian diklasifikasikan menjadi relevan atau tidak relevan terhadap query. Destinasi dikategorikan relevan apabila karakteristik dan aktivitas yang ditawarkan sesuai dengan kebutuhan yang dinyatakan dalam query, sedangkan destinasi yang tidak memiliki kesesuaian yang memadai dikategorikan tidak relevan. Hasil klasifikasi tersebut digunakan sebagai acuan (*ground truth*) untuk membandingkan hasil rekomendasi sistem.

Evaluasi

Evaluasi dilakukan pada penelitian ini ada 2 yaitu evaluasi hasil rekomendasi dan evaluasi pengguna (*user study*). Evaluasi hasil rekomendasi digunakan untuk mengukur kemampuan sistem dalam menghasilkan rekomendasi destinasi wisata yang relevan dengan preferensi pengguna. Evaluasi menggunakan *Precision@K* dan *Recall@K*. Penilaian dilakukan dengan membandingkan hasil rekomendasi sistem terhadap *ground truth* yang dibentuk melalui *manual relevance analysis*. Penentuan relevansi dilakukan oleh peneliti dengan memeriksa kesesuaian antara preferensi yang terdapat pada query dan atribut destinasi, meliputi kategori, kota, aktivitas, dan deskripsi. Destinasi diberi label relevan apabila karakteristiknya sesuai dengan kebutuhan yang dinyatakan dalam query, sedangkan destinasi yang tidak memenuhi kriteria tersebut diberi label tidak relevan. Hasil penilaian tersebut kemudian digunakan sebagai *ground truth* untuk menghitung *Precision@K* dan *Recall@K*.

Precision digunakan untuk mengukur ketepatan hasil rekomendasi yang diberikan sistem. Nilai *precision* menunjukkan proporsi destinasi wisata yang dinilai relevan dibandingkan dengan seluruh destinasi yang direkomendasikan oleh sistem [16]. *Precision* ini dihitung menggunakan persamaan (2).

$$Precision@K = \frac{TP}{TP+FP} \quad (2)$$

Recall merupakan salah satu metrik evaluasi yang digunakan untuk mengukur kemampuan sistem dalam menemukan kembali item yang relevan dari seluruh item relevan yang tersedia [17]. *Recall* ini dihitung menggunakan persamaan (3).

$$Recall@N = \frac{TP}{TP+FN} \quad (3)$$

Keterangan: TP (*True Positive*) = destinasi relevan yang berhasil direkomendasikan oleh sistem; FP (*False Positive*) = destinasi tidak relevan yang direkomendasikan oleh sistem; FN (*False Negative*) = destinasi relevan yang tidak masuk dalam hasil rekomendasi.

Evaluasi pengguna (*user study*) digunakan untuk mengetahui persepsi pengguna terhadap sistem rekomendasi destinasi wisata yang dikembangkan. Evaluasi melibatkan 20 responden dengan menggunakan kuesioner yang terdiri atas lima pernyataan. Setiap pernyataan dinilai menggunakan skala Likert lima tingkat, yaitu Sangat Tidak Setuju (1), Tidak Setuju (2), Netral (3), Setuju (4), dan Sangat Setuju (5).

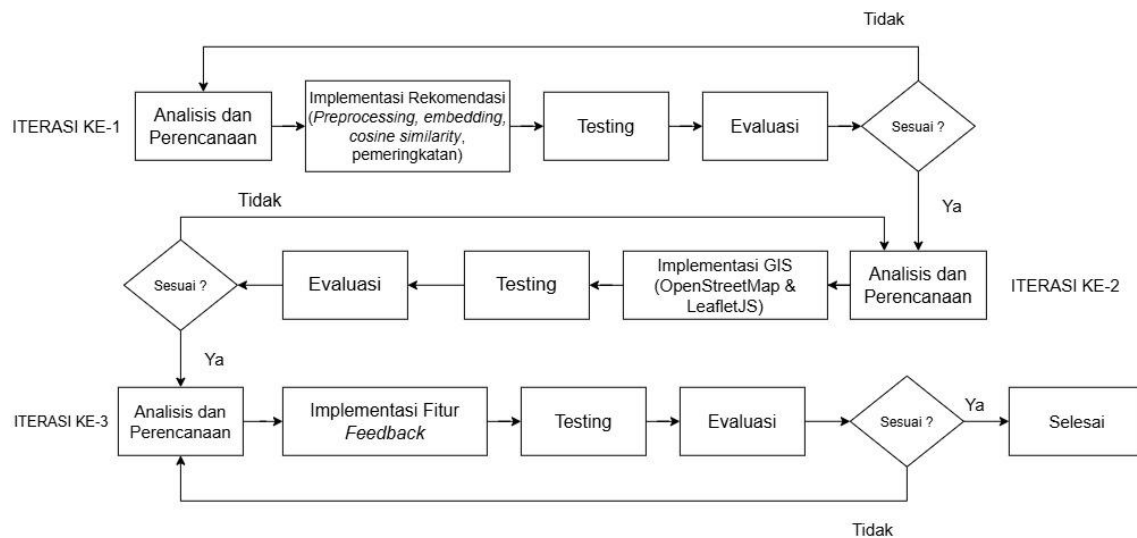
Persentase hasil penilaian setiap pernyataan dihitung dengan membandingkan total skor yang diperoleh dari seluruh responden terhadap skor maksimum. Dengan 20 responden dan skor tertinggi 5, skor maksimum untuk setiap pernyataan adalah 100. Persentase skor dihitung menggunakan persamaan (4).

$$Rumus : T * Pn \quad (4)$$

Keterangan: T = jumlah responded, Pn = bobot yang ada di setiap pilihan pernyataan
Hasil persentase kemudian diinterpretasikan menggunakan lima kategori, yaitu 20% sangat lemah, 21–40% lemah, 41–60% cukup, 61–80% kuat, dan 81–100% sangat kuat.

Iterative Development

Pada penelitian ini pendekatan untuk mengembangkan sistem rekomendasi menggunakan *iterative development*, dengan pendekatan *iterative development* pengembangan sistem ini dapat dibagi menjadi beberapa iterasi – iterasi alur iterasi setiap iterasi dimulai dari analisis dan perencanaan sampai dengan evaluasi, jika hasil dari evaluasi sudah sesuai maka akan berlanjut ke iterasi selanjutnya, untuk penelitian ini pengembangan sistem dibagi menjadi 3 iterasi yaitu, integrasi proses rekomendasi, integrasi fitur GIS dan integrasi fitur *feedback*. Untuk alur dari iterasi tersebut ditunjukkan pada Gambar 2.



Gambar 2. Alur Iterasi Pengembangan Sistem

3. HASIL DAN PEMBAHASAN

Hasil Pengumpulan Data

Penelitian ini menggunakan data destinasi wisata di wilayah DKI Jakarta yang digunakan sebagai dasar dalam proses rekomendasi destinasi. Data yang diperoleh berjumlah 178 destinasi wisata dan terdiri atas beberapa atribut yang menggambarkan karakteristik destinasi, meliputi nama destinasi, deskripsi, alamat, kota, kategori, aktivitas, serta koordinat geografis berupa latitude dan longitude. Rincian atribut yang digunakan dalam penelitian ditunjukkan pada Tabel 1. Atribut nama, kategori, deskripsi, dan aktivitas digunakan sebagai informasi tekstual dalam proses *content-based filtering*, sedangkan atribut latitude dan longitude digunakan untuk mendukung penyajian informasi geografis dan rute destinasi. Data destinasi yang telah dikumpulkan selanjutnya digunakan sebagai masukan pada tahap *preprocessing* sebelum dilakukan proses vektorisasi menggunakan model *multilingual-e5-small*.

Data destinasi yang digunakan dalam penelitian memiliki informasi yang terdiri atas beberapa atribut sesuai dengan karakteristik masing-masing destinasi. Contoh data faktual destinasi yang digunakan dalam penelitian ditunjukkan pada Tabel 2.

Hasil Preprocessing Data dan Pembentukan Dokumen Destinasi

Data destinasi yang telah diperoleh selanjutnya melalui tahap *preprocessing* untuk mempersiapkan informasi tekstual sebelum digunakan dalam proses vektorisasi. Tahap *preprocessing* diterapkan pada data teks yang digunakan sebagai karakteristik destinasi, meliputi

proses mengubah seluruh teks menjadi huruf kecil (*case folding*), menghapus karakter yang tidak diperlukan (*Remove Special*), dan menormalisasi spasi (*Normalize Space*). Lalu setelah itu setiap attribute data destinasi dijadikan satu narasi yang konsisten untuk membentuk dokumen destinasi wisata, Contoh hasil *preprocessing* pada query pengguna dan data destinasi ditunjukkan pada Tabel 3.

Tabel 1. Attribut data destinasi wisata

Attribute	Embedding Vektor
Nama	Nama destinasi wisata
Deskripsi	Informasi yang menjelaskan karakteristik destinasi
Alamat	Alamat lokasi destinasi
Kota	Wilayah/kota tempat destinasi berada
Kategori	Jenis atau kategori destinasi
Aktivitas	Aktivitas yang dapat dilakukan wisatawan
Latitude	Koordinat lintang destinasi
Longitude	Koordinat bujur destinasi

Tabel 2. Contoh data destinasi wisata

No	Nama Destinasi	Kota	Kategori	Aktivitas	Deskripsi
1	Taman Margasatwa Ragunan	Jakarta Selatan	Alam	melihat satwa, edukasi, berjalan kaki, piknik, bersepeda	Taman Margasatwa Ragunan adalah kawasan konservasi, edukasi, dan rekreasi alam.....
2	Museum Bank Indonesia	Jakarta Barat	Sejarah	edukasi, wisata sejarah, wisata budaya	Museum Bank Indonesia adalah museum sejarah modern...

Tabel 3. Contoh hasil preprocessing data

No	Data	Sebelum Preprocessing	Setelah Preprocessing
1	Query	Tempat yang enak untuk jogging di Jakarta Selatan!	query:tempat yang enak untuk jogging di jakarta selatan
2	Destinasi A	Museum Bank Indonesia, Jakarta Barat, Sejarah, edukasi, wisata sejarah, wisata budaya, Museum Bank Indonesia adalah museum sejarah modern....	passage:museum bank indonesia jakarta barat sejarah edukasi wisata sejarah wisata budaya museum bank indonesia adalah museum sejarah modern ...
...	...	...	...

Hasil Embedding

Hasil *preprocessing* pada *query* pengguna dan data destinasi selanjutnya diubah ke dalam bentuk representasi numerik menggunakan model *intfloat/multilingual-e5-small*. Proses vektorisasi dilakukan dengan model yang sama pada kedua jenis data sehingga *query* pengguna dan karakteristik destinasi direpresentasikan dalam ruang vektor yang sama. Setiap teks menghasilkan embedding berdimensi 384 yang selanjutnya digunakan sebagai dasar dalam proses perhitungan *cosine similarity*. Contoh hasil vektorisasi pada *query* pengguna dan data destinasi ditunjukkan pada Tabel 4.

Berdasarkan Tabel 4, *query* pengguna dan setiap data destinasi berhasil direpresentasikan dalam bentuk vektor berdimensi 384. Meskipun representasi *embedding* terdiri atas 384 nilai, hanya beberapa nilai awal yang ditampilkan pada tabel untuk menjaga keterbacaan. Seluruh nilai

embedding digunakan dalam tahap perhitungan cosine similarity untuk menentukan tingkat kemiripan antara preferensi pengguna dan karakteristik destinasi.

Tabel 4. Hasil vektorisasi

Data	Dimensi	Embedding Vektor
Query	384	[0.0691, 0.0166, -0.0329, -0.0266, 0.0554, ...]
Destinasi A	384	[0.0464, -0.0208, -0.0344, -0.0285, 0.0488, ...]
Destinasi B	384	[0.0536, -0.0038, -0.0156, -0.0383, 0.0579, ...]
Destinasi C	384	[0.0793, -0.0301, -0.0213, -0.0365, 0.0685, ...]
Destinasi n	384	...

Hasil Cosine Similarity dan Perankingan

Embedding yang dihasilkan dari *query* pengguna dan data destinasi selanjutnya dibandingkan menggunakan cosine similarity untuk memperoleh nilai kemiripan antara preferensi pengguna dan karakteristik setiap destinasi. Perhitungan dilakukan dengan membandingkan *embedding query* terhadap seluruh *embedding* destinasi yang tersedia. Nilai *cosine similarity* yang diperoleh kemudian digunakan untuk mengurutkan destinasi berdasarkan tingkat kemiripannya, dengan nilai yang lebih tinggi menunjukkan tingkat kesesuaian yang lebih besar terhadap preferensi pengguna. Hasil perhitungan kemiripan vektor dan peringkat destinasi untuk query “Tempat yang enak untuk jogging di Jakarta Selatan!” ditunjukkan pada Tabel 5.

Tabel 5. Hasil perhitungan kemiripan vektor

No	Nama Destinasi	Cosine Similarity	Peringkat	Interpretasi Semantik
1	Destinasi A	0.8751	1	Memiliki kesesuaian semantik yang sangat tinggi karena langsung menyebut adanya jogging track dan aktivitas utamanya jogging di pagi hari. Lokasinya juga berada di Jakarta Selatan, Selain itu, karakteristik taman yang luas, asri, dan memiliki jalan setapak mendukung konteks tempat untuk aktivitas jogging.
2	Destinasi B	0.8702	2	Memiliki kesesuaian semantik yang tinggi karena merupakan ruang publik di Jakarta Selatan dan menyediakan jogging track serta aktivitas olahraga. tetapi aktivitas lebih terfokus pada kegiatan skate board
3	Destinasi C	0.8674	3	Memiliki kesesuaian semantik yang kuat karena berada di Jakarta Selatan, menyediakan jalur jogging dan fasilitas kebugaran outdoor, tetapi ditaman ini fokus aktivitasnya adalah berjalan kaki dan rekreasi
4	Destinasi D	0.8672	4	Memiliki kesesuaian semantik yang cukup kuat karena berlokasi di jakarta selatan serta fasilitas olahraga dan ruang publik, namun aktivitas yang banyak dilakukan berkaitan dengan skate board, BMX, dan berjalan kaki
5	Destinasi E	0.8589	5	Memiliki kesesuaian semantic karena berada di jakarta selatan, dan secara langsung menyediakan sarana jogging track, tetapi aktivitas umum yang dilakukan adalah rekreasi seperti piknik dan healing

Berdasarkan Tabel 5. hasil rekomendasi dengan *query* “Tempat yang enak untuk jogging di Jakarta Selatan!”, destinasi pada peringkat pertama memperoleh nilai *cosine similarity* sebesar 0.8751. Tingginya nilai tersebut dapat diinterpretasikan melalui kesesuaian semantik antara *query* dengan karakteristik destinasi. Destinasi A memiliki informasi yang secara langsung berkaitan

dengan kebutuhan pengguna, yaitu keberadaan lintasan lari (*jogging track*), penggunaan taman untuk aktivitas jogging, serta lokasi yang berada di kawasan Jakarta Selatan. Selain itu, karakteristik taman yang luas, asri, dan memiliki jalan setapak juga mendukung konteks destinasi sebagai tempat untuk melakukan aktivitas luar ruang.

Interpretasi tersebut menunjukkan bahwa nilai cosine similarity yang dihasilkan tidak semata-mata didasarkan pada keberadaan satu kata yang sama dengan *query*. Model *multilingual-E5* merepresentasikan keseluruhan informasi teks destinasi ke dalam bentuk *embedding*, sehingga kemiripan dapat mencerminkan kesesuaian konteks antara preferensi pengguna dan karakteristik destinasi. Oleh karena itu, nilai similarity digunakan untuk menentukan tingkat kemiripan antarrepresentasi teks, sedangkan karakteristik destinasi digunakan untuk memberikan interpretasi terhadap hasil pemeringkatan yang diperoleh.

Hasil Evaluasi Sistem

Evaluasi dilakukan terhadap hasil rekomendasi untuk mengetahui tingkat relevansi destinasi yang dihasilkan oleh sistem terhadap preferensi pengguna. Evaluasi menggunakan Precision@K dan Recall@K yang dimana K merepresentasikan *top-K* yaitu jumlah dari destinasi yang direkomendasikan. Untuk memperoleh evaluasi yang lebih menyeluruh, pengujian dilakukan menggunakan beberapa skenario *query* dengan karakteristik kebutuhan pengguna yang berbeda serta nilai *top-K* sebesar 1, 3, 5, dan 10.

Skenario *query* yang digunakan mencakup kebutuhan berdasarkan aktivitas, kategori, dan lokasi. Daftar skenario *query* ditunjukkan pada Tabel 6.

Tabel 6. Skenario *Query* Pengujian

Kode	<i>Query</i>
Q1	Tempat yang enak untuk jogging di Jakarta Selatan
Q2	Destinasi wisata sejarah di jakarta pusat
Q3	Pusat perbelanjaan yang ada di sekitaran Jakarta
Q4	tempat wisata edukasi yang asyik di Jakarta Utara
Q5	Spot kuliner di Jakarta Selatan

Pembentukan *ground truth* dilakukan melalui *manual relevance analysis* oleh peneliti sendiri terhadap data destinasi berdasarkan kesesuaian antara *query* dan karakteristik destinasi. Penentuan relevansi mempertimbangkan informasi pada atribut destinasi, meliputi kategori, kota, aktivitas, dan deskripsi. Hasil dari manual relevance analysis yang dilakukan untuk membentuk *ground truth* setiap skenario *query* di tunjukan pada Tabel 7.

Tabel 7. *Ground truth* berdasarkan *query*

Kode	<i>Ground truth</i>
Q1	7
Q2	18
Q3	20
Q4	7
Q5	23

Destinasi wisata apabila karakteristiknya sesuai dengan kebutuhan yang dinyatakan pada *query* akan dianggap relevan dan tidak relevan apabila tidak sesuai. Evaluasi hasil penilaian relevansi rekomendasi destinasi wisata yang diberikan oleh sistem terhadap skenario *top-K* di tunjukan pada Tabel 8.

Berdasarkan hasil penilaian evaluasi relevansi, destinasi yang relevan dan berhasil ditemukan sistem dikategorikan sebagai *True Positive* (TP), sedangkan destinasi yang tidak

relevan tetapi masuk ke dalam hasil rekomendasi dikategorikan sebagai *False Positive* (FP). Adapun destinasi yang relevan tetapi tidak termasuk ke dalam hasil rekomendasi atau tidak ditemukan oleh sistem dikategorikan sebagai *False Negative* (FN). Hasil TP, FP dan FN pada setiap skenario *Top-K* tersebut ditunjukkan pada Tabel 9.

Tabel 8. Hasil Evaluasi Relevansi pada setiap Skenario *Top-K*

Kode	K	Hasil Relevan	Hasil Tidak Relevan	Hasil Relevan Tapi Tidak Ditemukan
Q1	1	1	0	6
	3	3	0	4
	5	5	0	2
	10	7	3	0
Q2	1	1	0	17
	3	3	0	15
	5	5	0	13
	10	9	1	9
Q3	1	1	0	19
	3	3	0	17
	5	5	0	15
	10	10	0	10
Q4	1	1	0	6
	3	3	0	4
	5	5	0	2
	10	7	3	0
Q5	1	1	0	22
	3	3	0	20
	5	5	0	18
	10	10	0	13

Berdasarkan hasil TP, FP dan FN pada Tabel 9, nilai $Precision@K$ dihitung untuk mengetahui tingkat ketepatan rekomendasi yang diberikan oleh sistem sesuai dengan skenario jumlah skenario *Top-K*. Perhitungan $Precision@K$ dilakukan dengan membandingkan jumlah destinasi relevan yang berhasil direkomendasikan terhadap seluruh hasil rekomendasi. Sebagai contoh, untuk perhitungan $Precision@K$ pada skenario Q1 dengan nilai $K = 10$ ditunjukkan pada persamaan (4).

$$Precision@10 = \frac{7}{7+3} = 0.7 \quad (4)$$

Berdasarkan perhitungan tersebut, diperoleh nilai $Precision@10$ sebesar 70%. Nilai tersebut menunjukkan bahwa sebagian besar destinasi yang termasuk dalam hasil rekomendasi merupakan destinasi yang relevan dengan preferensi pengguna.

Selanjutnya, $Recall@K$ digunakan untuk mengetahui kemampuan sistem dalam menemukan destinasi yang relevan dari seluruh destinasi yang telah ditetapkan relevan berdasarkan ground truth. Perhitungan $Recall@K$ pada skenario Q1 dengan nilai $K = 10$ dilakukan menggunakan persamaan (5).

$$Recall@10 = \frac{7}{7+0} = 1.0 \quad (5)$$

Hasil perhitungan menunjukkan bahwa nilai Recall@10 pada skenario Q1 yang diperoleh sebesar 100%. Nilai tersebut menunjukkan bahwa sistem mampu menemukan seluruh destinasi yang relevan untuk hasil rekomendasi. Seluruh hasil perhitungan precision dan recall dari berbagai variasi skenario query dan skenario Top-K kemudian dirangkum dan ditunjukkan pada Tabel 10.

Tabel 9. TP, FP dan FN pada setiap skenario *Top-K*

Kode	K	TP	FP	FN
Q1	1	1	0	6
	3	3	0	4
	5	5	0	2
	10	7	3	0
Q2	1	1	0	17
	3	3	0	15
	5	5	0	13
	10	9	1	9
Q3	1	1	0	19
	3	3	0	17
	5	5	0	15
	10	10	0	10
Q4	1	1	0	6
	3	3	0	4
	5	5	0	2
	10	7	3	0
Q5	1	1	0	22
	3	3	0	20
	5	5	0	18
	10	10	0	13

Tabel 10. Hasil perhitungan precision dan recall pada setiap skenario

Kode	P@1	R@1	P@3	R@3	P@5	R@5	P@10	R@10
Q1	1.0	0.14	1.0	0.43	1.0	0.71	0.70	1.0
Q2	1.0	0.06	1.0	0.16	1.0	0.27	0.90	0.50
Q3	1.0	0.05	1.0	0.15	1.0	0.25	1.0	0.50
Q4	1.0	0.14	1.0	0.43	1.0	0.71	0.70	1.0
Q5	1.0	0.04	1.0	0.13	1.0	0.22	1.0	0.43
Rata - Rata	1.0	0.09	1.0	0.26	1.0	0.43	0.86	0.69

Berdasarkan hasil rata-rata yang ditunjukkan pada Tabel 10, terlihat gambaran performa sistem rekomendasi pada berbagai skenario *query* dan nilai *top-K*. Nilai rata-rata *precision* berada pada angka 1.0 pada K=1, K=3, dan K=5, kemudian sedikit menurun menjadi 0,86 pada K=10. Sementara itu, nilai rata-rata *recall* meningkat dari 0,09 pada K=1 menjadi 0,69 pada K=10. Peningkatan *recall* dari K=1 hingga K=10 menunjukkan bahwa penambahan jumlah rekomendasi memungkinkan sistem menemukan lebih banyak destinasi yang relevan. Berdasarkan hasil tersebut, K=10 menghasilkan nilai rata-rata *recall* tertinggi pada skenario pengujian yang dilakukan.

Hasil Evaluasi User Study

Evaluasi pengguna ini menggunakan data dari 20 responden dan dilakukan menggunakan *skala Likert* dengan lima pilihan jawaban, yaitu Sangat Tidak Setuju (1), Tidak Setuju (2), Netral (3), Setuju (4), dan Sangat Setuju (5). Persentase skor setiap pernyataan dihitung berdasarkan jumlah responden pada masing-masing pilihan, kemudian dibandingkan dengan skor maksimum yang diperoleh dari perkalian skor tertinggi dengan jumlah responden. Dengan jumlah 20 responden, skor maksimum adalah 100 dan skor minimum adalah 20. Interpretasi hasil evaluasi menggunakan lima kategori, yaitu 20% sangat lemah, 21–40% lemah, 41–60% cukup, 61–80% kuat, dan 81–100% sangat kuat. Hasil evaluasi pengguna ditunjukkan pada Tabel 11.

Tabel 11. Evaluasi Pengguna

Pernyataan	Persentase
Kesesuaian rekomendasi dengan preferensi	82%
Kemudahan informasi lokasi dan GIS	84%
Kemudahan antarmuka dan navigasi	83%
Kesesuaian fitur dengan kebutuhan	85%
Manfaat sistem secara keseluruhan	83%
Rata -Rata	83.4%

Berdasarkan hasil evaluasi pengguna terhadap 20 responden, diperoleh persentase rata-rata sebesar 83,4%. Penilaian tertinggi terdapat pada aspek kesesuaian fitur dengan kebutuhan pengguna sebesar 85%, sedangkan penilaian terendah terdapat pada aspek kesesuaian rekomendasi dengan preferensi sebesar 82%. Hasil tersebut menunjukkan bahwa responden memberikan penilaian positif terhadap aspek rekomendasi, informasi lokasi dan GIS, antarmuka, fitur, serta manfaat sistem secara keseluruhan.

Pembahasan

Hasil pengujian menunjukkan bahwa penerapan *Content-Based Filtering* menggunakan representasi *embedding* dan *cosine similarity* menghasilkan pemeringkatan rekomendasi berdasarkan kesesuaian antara preferensi pengguna dan karakteristik destinasi. Representasi karakteristik destinasi melalui *intfloat/multilingual-e5-small* memungkinkan informasi tekstual seperti nama, kategori, kota, aktivitas, dan deskripsi direpresentasikan dalam bentuk vektor sehingga kemiripan antara preferensi pengguna dan destinasi dapat dihitung secara semantik. Proses tersebut menghasilkan pemeringkatan destinasi berdasarkan nilai *cosine similarity*, sehingga destinasi dengan representasi yang memiliki kedekatan semantik terhadap preferensi pengguna dapat ditempatkan pada posisi teratas dalam hasil rekomendasi.

Berdasarkan hasil evaluasi pada beberapa skenario *query* dan nilai K, diperoleh gambaran performa sistem dari sisi ketepatan dan cakupan rekomendasi. Nilai rata-rata *precision* sebesar 1,0 diperoleh pada K=1, K=3, dan K=5, kemudian menurun menjadi 0,86 pada K=10. Sementara itu, nilai rata-rata *recall* meningkat dari 0,08 pada K=1 menjadi 0,69 pada K=10. Peningkatan *recall* dari K=1 hingga K=10 menunjukkan bahwa penambahan jumlah rekomendasi memungkinkan sistem mencakup lebih banyak destinasi yang relevan.

Hasil evaluasi juga menunjukkan bahwa tingkat relevansi rekomendasi masih bervariasi pada setiap skenario *query* dan nilai *top-K*. Kondisi tersebut dapat terjadi karena sistem menentukan pemeringkatan berdasarkan kemiripan representasi tekstual antara *query* pengguna dan informasi destinasi. Destinasi yang memiliki kedekatan semantik dengan *query* dapat memperoleh nilai *similarity* yang relatif tinggi. Hasil rekomendasi juga bergantung pada kelengkapan dan karakteristik informasi yang tersedia pada setiap destinasi. Hal ini menunjukkan bahwa pendekatan *Content-Based Filtering* memiliki keterbatasan dalam menangkap kebutuhan

pengguna apabila informasi yang merepresentasikan karakteristik destinasi belum cukup menggambarkan konteks preferensi pengguna.

Selain menghasilkan rekomendasi berdasarkan preferensi, integrasi *Geographic Information System* dalam sistem memberikan informasi lokasi dan rute menuju destinasi yang dipilih. Fitur tersebut melengkapi sistem dengan memberikan informasi geografis sehingga pengguna tidak hanya memperoleh alternatif destinasi yang sesuai dengan preferensi, tetapi juga dapat mengetahui lokasi dan rute menuju destinasi tersebut. Di sisi lain, fitur ulasan menyediakan mekanisme bagi wisatawan untuk memberikan pengalaman dan penilaian terhadap destinasi yang dikunjungi. Informasi tersebut dapat menjadi data pendukung bagi pengelola dalam memahami pengalaman wisatawan dan melakukan evaluasi terhadap destinasi.

Untuk memperjelas perbedaan pendekatan yang digunakan, penelitian ini dibandingkan dengan beberapa penelitian terdahulu berdasarkan metode rekomendasi, representasi konten, pemanfaatan informasi geografis, fitur umpan balik, dan metode evaluasi yang ditunjukkan pada tabel 12.

Tabel 12. Perbandingan pendekatan dan evaluasi penelitian terdahulu

Penelitian	Metode	Data yang digunakan	Evaluasi Hasil
Arkan et al,	CBF + TF-IDF	Nama, Kategori, Fasilitas, Deskripsi	Menghasilkan $P@5 = 0.72$, $R@5 = 0.45$, $P@10 = 0.66$ dan $R@10 = 0.825$
Widika et al,	CBF + TF-IDF + Haversine	Nama, Deskripsi, Kota, Kategori, Partner, Temperature	Precision = 0.63, Recall = 0.90 dan F1-Score = 0.73
Penelitian ini	CBF + E5	Nama, Kota, Kategori, Aktivitas, Deskripsi	Menghasilkan rata – rata $P@1 = 1.0$ dan $R@1 = 0.09$, $P@3 = 1.0$ dan $R@3 = 0.26$, $P@5 = 1.0$ dan $R@5 = 0.43$, $P@10 = 0.86$ dan $R@10 = 0.69$

Berdasarkan Tabel 10, penelitian ini menggunakan representasi *semantic embedding* dengan *multilingual-e5-small*, sedangkan beberapa penelitian terdahulu menggunakan representasi berbasis TF-IDF. Penelitian ini juga menggunakan GIS sebagai komponen pendukung untuk penyajian lokasi dan rute, sedangkan proses pemeringkatan rekomendasi dilakukan menggunakan *cosine similarity*. Perbandingan hasil evaluasi menunjukkan adanya perbedaan metrik antarpenelitian, namun nilai tersebut tidak dibandingkan secara langsung karena setiap penelitian menggunakan dataset, jumlah query, nilai *top-K*, dan *ground truth* yang berbeda.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem rekomendasi destinasi wisata berbasis web yang menerapkan metode *Content-Based Filtering* dengan model *intfloat/multilingual-e5-small* dan *cosine similarity* untuk menghasilkan rekomendasi berdasarkan kesesuaian antara preferensi pengguna dan karakteristik destinasi. Sistem juga mengintegrasikan *Geographic Information System* (GIS) untuk menyediakan informasi lokasi dan rute serta fitur ulasan sebagai media pengumpulan umpan balik wisatawan. Hasil evaluasi pada lima skenario query dengan nilai $K=1$, $K=3$, $K=5$, dan $K=10$ menunjukkan bahwa nilai rata-rata *precision* sebesar 1.0 pada $K=1$, $K=3$, dan $K=5$, kemudian sedikit menurun menjadi 0,86 pada $K=10$. Sementara itu, nilai rata-rata *recall* meningkat dari 0,09 pada $K=1$ menjadi 0,26 pada $K=3$ dan 0,43 pada $K=5$, kemudian menjadi 0,69 pada $K=10$. Hasil tersebut menunjukkan bahwa peningkatan jumlah rekomendasi hingga $K=10$ memungkinkan sistem menemukan lebih banyak destinasi yang relevan. Secara keseluruhan, sistem yang dikembangkan dapat membantu wisatawan menemukan destinasi yang sesuai dengan preferensinya serta menyediakan informasi lokasi dan umpan balik yang dapat dimanfaatkan sebagai pendukung evaluasi dan pengembangan destinasi wisata.

5. SARAN

Penelitian ini memiliki beberapa keterbatasan. Metode *Content-Based Filtering* (CBF) bergantung pada informasi dan atribut yang tersedia pada destinasi sehingga belum mempertimbangkan pola interaksi antar pengguna serta dapat membatasi keberagaman rekomendasi. Selain itu, *ground truth* ditentukan melalui *manual relevance analysis* oleh satu penilai, sehingga masih terdapat potensi subjektivitas dalam menentukan relevansi destinasi. Evaluasi pengguna juga melibatkan 20 responden, sehingga hasilnya belum sepenuhnya merepresentasikan persepsi pengguna dengan karakteristik yang lebih beragam. Penelitian selanjutnya dapat mempertimbangkan metode *hybrid* untuk memanfaatkan informasi interaksi pengguna, melibatkan beberapa penilai atau pengguna dalam penentuan *ground truth*, serta menggunakan jumlah dan karakteristik responden yang lebih beragam untuk memperoleh evaluasi yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] D. Shrestha, T. Wenan, D. Shrestha, N. Rajkarnikar, dan S.-R. Jeong, "Personalized Tourist Recommender System: A Data-Driven and Machine-Learning Approach," *Computation*, vol. 12, no. 3, hlm. 59, Mar 2024, doi: 10.3390/computation12030059.
- [2] M. S. Viñán-ludeña dan L. De Campos, "Evaluating Tourist Dissatisfaction with Aspect-Based Sentiment Analysis Using Social Media Data," *Advances in Hospitality and Tourism Research (AHTR)*, vol. 12, no. 3, hlm. 254–286, Sep 2024, doi: 10.30519/ahtr.1436175.
- [3] A. Solano-Barliza dkk., "Recommender systems applied to the tourism industry: a literature review," *Cogent Business & Management*, vol. 11, no. 1, hlm. 2367088, Des 2024, doi: 10.1080/23311975.2024.2367088.
- [4] S. Oyadila, D. Abdullah, dan 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, hlm. 1329–1339, Jul 2025, doi: 10.36341/rabit.v10i2.6532.
- [5] M. N. Arkan, A. P. Widodo, dan G. Aryotejo, "A Web-Based Tourism Recommendation System for Boyolali Using Content-Based Filtering and Cosine Similarity," *JMASIF*, vol. 17, no. 1, hlm. 39–53, Jan 2026, doi: 10.14710/jmasif.17.1.72520.
- [6] M. Hendre dan P. Mukherjee, "Efficacy of Deep Neural Embeddings- Based Semantic Similarity in Automatic Essay Evaluation," *International Journal of Cognitive Informatics and Natural Intelligence*, 2023.
- [7] A. Amelia, T. N. Ad, S. Suakanto, dan S. C. R. Rijadi, "Contextual Tourism Access Through Geographic Information System with Recommendation System: The Case of Greater Bandung," dalam *2025 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, Bali, Indonesia: IEEE, Jul 2025, hlm. 283–289. doi: 10.1109/IAICT65714.2025.11101505.
- [8] C. Salsabilla dan D. W. Utomo, "Pekalongan Regency Tourism Recommendation System with Content based Filtering," *SISTEMASI*, vol. 14, no. 1, hlm. 262, Jan 2025, doi: 10.32520/stmsi.v14i1.4839.
- [9] M. N. Arkan, A. P. Widodo, dan G. Aryotejo, "A Web-Based Tourism Recommendation System for Boyolali Using Content-Based Filtering and Cosine Similarity," *JMASIF*, vol. 17, no. 1, hlm. 39–53, Jan 2026, doi: 10.14710/jmasif.17.1.72520.
- [10] M. A. Widika, P. S. Susilo, A. I. Ramadhan, S. Nur, F. A. Sariasih, dan I. Sutoyo, "SISTEM REKOMENDASI DESTINASI WISATA BERBASIS CONTENT-BASED FILTERING DAN ANALISIS FITUR GEOSPASIAL," vol. 11, no. 1, 2026.

- [11] S. Huang, "Research on software reliability enhancement methods based on iterative development," *ACE*, vol. 76, no. 1, hlm. 21–26, Jul 2024, doi: 10.54254/2755-2721/76/20240558.
- [12] K. S. Nugroho, F. A. Bachtiar, dan W. F. Mahmudy, "Detecting Emotion in Indonesian Tweets: A Term-Weighting Scheme Study," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 8, no. 1, hlm. 61–70, Apr 2022, doi: 10.20473/jisebi.8.1.61-70.
- [13] L. Wang dkk., "Text Embeddings by Weakly-Supervised Contrastive Pre-training," 22 Februari 2024, arXiv: arXiv:2212.03533. doi: 10.48550/arXiv.2212.03533.
- [14] L. Wang, N. Yang, X. Huang, L. Yang, R. Majumder, dan F. Wei, "Multilingual E5 Text Embeddings: A Technical Report," 8 Februari 2024, arXiv: arXiv:2402.05672. doi: 10.48550/arXiv.2402.05672.
- [15] H. Hartatik dan A. Syafrianto, "PENERAPAN MODEL SENTENCE-BERT UNTUK SISTEM REKOMENDASI BUKU BERBASIS KONTEN DI PERPUSTAKAAN DIGITAL," *J. Dialektika Inform.*, vol. 6, no. 1, hlm. 12–19, Nov 2025, doi: 10.24176/detika.v6i1.15916.
- [16] T.-D. Nguyen, "An approach to improve the accuracy of rating prediction for recommender systems," *Automatika*, vol. 65, no. 1, hlm. 58–72, Jan 2024, doi: 10.1080/00051144.2023.2284026.
- [17] M. Acharya dan K. K. Mohbey, "Recency-based spatio-temporal similarity exploration for POI recommendation in location-based social networks," *Front. Sustain. Cities*, vol. 6, hlm. 1331642, Apr 2024, doi: 10.3389/frsc.2024.1331642.