

BAB II

STUDI PUSTAKA

2.1. Penelitian Terkait

Penelitian yang dilakukan oleh Suhailah dan Hartatik (2023) membahas pengembangan sistem rekomendasi pariwisata di Yogyakarta dengan menerapkan metode *Triangle Multiplaying Jaccard*. Metode tersebut digunakan untuk mengukur tingkat kemiripan (*similarity*) antara objek wisata sebagai dasar dalam menghasilkan rekomendasi kepada pengguna. Evaluasi kinerja sistem dilakukan menggunakan beberapa parameter, yaitu *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *Confusion Matrix*. Berdasarkan hasil pengujian terhadap 20 data, sistem menghasilkan nilai akurasi rata-rata sebesar 70%. Sementara itu, nilai *precision* yang diperoleh sebesar 45% dan *recall* sebesar 1%, sedangkan nilai *F1-Score* mencapai 65%. Hasil tersebut menunjukkan bahwa metode *Triangle Multiplaying Jaccard* dapat diterapkan dalam membangun sistem rekomendasi pariwisata, meskipun masih terdapat keterbatasan terutama pada kemampuan sistem dalam menghasilkan rekomendasi yang relevan berdasarkan nilai *recall* yang relatif rendah.

Penelitian yang dilakukan oleh Lubis dkk. (2020) membahas penerapan metode *Hybrid Filtering* yang menggabungkan pendekatan *Collaborative Filtering* dan *Content-Based Filtering* dalam membangun sistem rekomendasi pariwisata. Penggabungan kedua metode tersebut bertujuan untuk menghasilkan rekomendasi objek wisata yang lebih sesuai dengan preferensi pengguna, sekaligus mengatasi permasalahan *sparsity* pada data dan meningkatkan ketepatan hasil prediksi. Dalam

proses evaluasinya, beberapa eksperimen dilakukan untuk memperoleh konfigurasi algoritma yang memberikan nilai *Mean Absolute Error* (MAE) paling rendah. Konfigurasi dengan nilai MAE terendah kemudian digunakan sebagai dasar dalam menghasilkan rekomendasi objek wisata. Hasil penelitian menunjukkan bahwa pendekatan *Hybrid Filtering* dapat diterapkan untuk membangun sistem rekomendasi pariwisata. Berdasarkan hasil eksperimen, nilai MAE terendah yang diperoleh adalah sebesar 0,3741 dengan nilai k sebesar 25%, yang menunjukkan bahwa pendekatan gabungan tersebut mampu menghasilkan prediksi rekomendasi dengan tingkat kesalahan yang relatif rendah. Metode *hybrid filtering* terbukti lebih baik dari metode *content-based filtering* yang menghasilkan nilai MAE yaitu 1,174201 dan metode *collaborative filtering* yang menghasilkan nilai MAE 0,3768 dengan k 25%. Berdasarkan hasil tersebut, pengisian data *sparse* terbukti efektif berdasarkan nilai MAE yang diperoleh pada metode *hybrid filtering*.

Penelitian yang dilakukan oleh Minarni dan Sigit (2023) berjudul “Pengembangan Sistem Informasi Rekomendasi Wisata Kotawaringin Timur Berbasis Web Menggunakan Metode *Item-Based Collaborative Filtering*”. Penelitian tersebut menerapkan metode *Item-Based Collaborative Filtering* untuk menghasilkan rekomendasi destinasi wisata berdasarkan data dan preferensi pengguna. Metode ini dipilih karena memiliki mekanisme yang relatif sederhana serta mudah diterapkan dalam sistem rekomendasi, khususnya untuk mendukung proses personalisasi secara *online*. Hasil penerapan metode tersebut menunjukkan bahwa sistem mampu memberikan rekomendasi objek wisata di Kotawaringin Timur kepada pengguna. Rekomendasi yang ditampilkan berupa objek wisata

dengan nilai *rating* yang tinggi dan dilengkapi dengan informasi mengenai fasilitas pendukung yang tersedia pada masing-masing destinasi. Dengan demikian, sistem dapat membantu pengguna memperoleh informasi wisata sekaligus menentukan destinasi yang sesuai berdasarkan rekomendasi yang diberikan.

Selanjutnya, penelitian yang dilakukan oleh Utama dkk. (2024) berjudul “Sistem Rekomendasi Tempat Wisata di Malang Raya Menggunakan Metode *K-Means Clustering*”. Penelitian ini menerapkan algoritma *K-Means Clustering* untuk mengelompokkan objek wisata ke dalam beberapa *cluster* berdasarkan lokasi atau karakteristik wilayah tempat wisata tersebut berada. Pengelompokan tersebut digunakan sebagai dasar dalam menghasilkan rekomendasi destinasi wisata di wilayah Malang Raya. Pengujian dilakukan dengan menggunakan variasi jumlah *cluster*, yaitu 2, 3, 4, dan 5 *cluster* pada wilayah Kota Malang, Kota Batu, dan Kabupaten Malang. Hasil penelitian menunjukkan bahwa metode *K-Means Clustering* mampu digunakan untuk membentuk kelompok data destinasi wisata sebagai dasar rekomendasi. Dari beberapa konfigurasi yang diuji, penggunaan 5 *cluster* memberikan hasil yang lebih optimal pada wilayah Kota Malang.

Penelitian lainnya dilakukan oleh Mi’Roj dkk. (2023) dengan judul “Implementasi Metode *Cosine Similarity* untuk Rekomendasi Pariwisata Berbasis Website”. Penelitian tersebut menggunakan metode *Cosine Similarity* untuk mengukur tingkat kemiripan antara data wisata sebagai dasar dalam menentukan rekomendasi destinasi kepada pengguna. Selain penerapan metode, penelitian juga melakukan pengujian terhadap penerimaan pengguna melalui *User Acceptance Testing* (UAT). Berdasarkan hasil pengujian, metode *Cosine Similarity* dapat

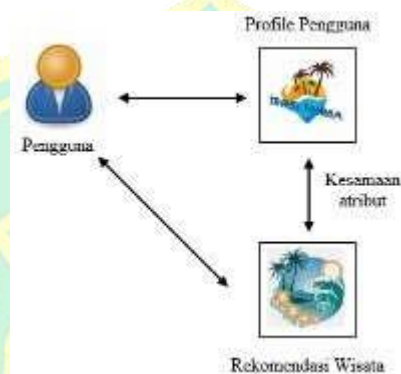
diterapkan pada sistem rekomendasi pariwisata berbasis website dan mampu menghasilkan rekomendasi yang sesuai dengan data yang digunakan. Hasil pengujian penerimaan sistem menunjukkan tingkat keberhasilan sebesar 90% dari 100%, sehingga sistem dinilai dapat diterima dan digunakan sebagai media untuk membantu pengguna memperoleh rekomendasi destinasi wisata.

2.2. Content-Based Filtering

Content-Based Filtering (CBF) merupakan salah satu pendekatan dalam sistem rekomendasi yang menghasilkan rekomendasi berdasarkan kesesuaian karakteristik suatu item dengan preferensi pengguna. Proses rekomendasi dilakukan dengan membandingkan atribut atau informasi yang terdapat pada item dengan profil pengguna, kemudian menentukan tingkat kemiripan di antara item tersebut. Item yang memiliki karakteristik paling mendekati preferensi pengguna akan memperoleh tingkat relevansi yang lebih tinggi dan menjadi prioritas untuk direkomendasikan. Pendekatan ini memiliki keunggulan karena pengguna dapat memperoleh informasi mengenai item yang memiliki karakteristik sesuai dengan minatnya. Dalam prosesnya, informasi atau deskripsi yang melekat pada suatu item menjadi bagian penting dalam menentukan tingkat kesesuaian rekomendasi (Nastiti, 2019).

Pada penelitian ini, metode *Content-Based Filtering* memanfaatkan algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk melakukan pembobotan terhadap atribut atau informasi tekstual yang terdapat pada data wisata. TF-IDF digunakan untuk menentukan tingkat kepentingan suatu kata dalam sebuah dokumen berdasarkan frekuensi kemunculannya pada dokumen tersebut dan

tingkat kemunculannya pada keseluruhan dokumen. Bobot yang dihasilkan selanjutnya dapat digunakan sebagai dasar untuk menghitung tingkat kemiripan antarobjek wisata sehingga sistem dapat memberikan rekomendasi berdasarkan karakteristik yang memiliki kesesuaian dengan preferensi pengguna (Mariani Widia Putri & Achmad Muchayan, 2018).



Gambar 2.1 Skema *Content-Based Filtering*

Pada Gambar 2.1 ditunjukkan mekanisme kerja *Content-Based Filtering* (C-BF) dalam menghasilkan rekomendasi kepada pengguna. Sistem bekerja dengan membandingkan karakteristik atau atribut yang terdapat pada suatu item dengan preferensi pengguna untuk mengetahui tingkat kesesuaiannya. Setiap atribut konten diberikan bobot tertentu berdasarkan tingkat kepentingannya, kemudian sistem melakukan proses perhitungan untuk menentukan tingkat kemiripan antara profil pengguna dan item yang tersedia. Item yang memiliki tingkat kemiripan paling tinggi akan memperoleh prioritas lebih besar untuk ditampilkan sebagai rekomendasi. Dalam proses tersebut, metode TF-IDF dapat digunakan untuk memberikan bobot pada setiap kata atau atribut tekstual sehingga informasi yang terdapat pada konten dapat diolah dan digunakan dalam menentukan tingkat relevansi rekomendasi.

Tahapan TF-IDF melibatkan dua komponen utama. Pertama, *Trem Frequency* (TF) menghitung seberapa sering sebuah kata tertentu muncul dalam sebuah dokumen, dan kedua, *Content-Based Filtering (CBF)* menghitung seberapa jarang sebuah kata tertentu muncul dalam sebuah dokumen. Untuk mengukur nilai TF atau seberapa sering kata (i) muncul dalam dokumen (j), digunakan rumus pada persamaan 2.1.

$$TF_{(i,j)} = \frac{freq(i,j)}{maxOther(i,j)}$$

Keterangan :

$freq(i,j)$ adalah frekuensi kata (i) dalam suatu dokumen (j)

$maxOthers(i,j)$ adalah frekuensi kata selain kata (i) dalam dokumen (i),

Perhitungan selanjutnya adalah perhitungan IDF, untuk mengukur seberapa penting kata (i) dalam seluruh dokumen (N), bertujuan untuk memberikan bobot yang lebih tinggi pada kata lebih jarang muncul dalam dokumen, semakin sedikit kata muncul, maka semakin tinggi nilai IDF nya. Perhitungan IDF dapat dilakukan dengan menggunakan rumus pada persamaan 2.2.

$$IDF_j = \log \frac{N}{n(i)}$$

Keterangan

N = total jumlah dokumen dalam koleksi,

$n(i)$ = total jumlah dokumen yang mengandung kata (i).

Langkah selanjutnya adalah perhitungan bobot pada kata, dengan mempertimbangkan seberapa sering kata tersebut muncul dalam dokumen (nilai TF) dan seberapa pentingnya kata tersebut dalam seluruh dokumen (nilai IDF).

Nilai TF dan IDF akan dilakukan untuk mendapatkan nilai bobotnya. Perhitungan bobot bertujuan untuk mengidentifikasi kata yang lebih relevan dan penting dalam suatu dokumen. Untuk melakukan perhitungan bobot maka dengan menggunakan rumus pada persamaan 2.3.

$$TF - IDF_{(i,j)} = TF(i, j) \times IDF(j)$$

2.3. *Cosine Similarity*

Cosine Similarity merupakan salah satu metode yang banyak digunakan untuk mengukur tingkat kemiripan antara dua objek yang direpresentasikan dalam bentuk vektor. Dalam sistem rekomendasi, metode ini dapat digunakan untuk mengetahui tingkat kesesuaian antara preferensi pengguna sebagai *query* dengan karakteristik atau atribut suatu objek wisata sebagai *document*. Pengukuran dilakukan berdasarkan nilai cosinus dari sudut yang terbentuk antara kedua vektor. Semakin kecil sudut antara vektor tersebut, semakin tinggi nilai *Cosine Similarity* yang diperoleh, sehingga menunjukkan bahwa kedua objek memiliki karakteristik yang semakin mirip.

Penerapan *Cosine Similarity* pada penelitian ini digunakan untuk menghitung tingkat kemiripan antara preferensi pengguna dengan atribut yang dimiliki oleh objek wisata. Setelah data atribut diubah menjadi bentuk vektor melalui proses pembobotan TF-IDF, nilai kemiripan setiap objek wisata dapat dihitung dan dibandingkan. Objek wisata yang memperoleh nilai *Cosine Similarity* lebih tinggi menunjukkan tingkat kesesuaian yang lebih besar terhadap preferensi pengguna dan akan memperoleh prioritas yang lebih tinggi dalam hasil rekomendasi. Dengan demikian, penggunaan metode ini diharapkan dapat membantu sistem

menghasilkan rekomendasi destinasi wisata yang lebih relevan dan sesuai dengan kebutuhan atau minat masing-masing pengguna.

Keterangan:

$$sim(q, d) = \frac{q \cdot d}{|q| |d|}$$

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

q = vektor *query*

d_j = vektor dokumen ke-j

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

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

2.4. Sistem Rekomendasi

Sistem rekomendasi merupakan sebuah sistem yang dirancang untuk membantu pengguna dalam menentukan pilihan terhadap suatu item berdasarkan informasi dan preferensi yang tersedia. Penerapannya dapat ditemukan pada berbagai bidang, seperti rekomendasi film, musik, buku, produk, maupun destinasi wisata. Sistem ini bekerja dengan mengolah data tertentu untuk menghasilkan alternatif pilihan yang dianggap sesuai dengan kebutuhan atau minat pengguna. Dalam penerapannya, terdapat beberapa pendekatan yang umum digunakan, di antaranya *Content-Based Filtering*, *Collaborative Filtering*, dan *Knowledge-Based Filtering*.

Content-Based Filtering menghasilkan rekomendasi dengan mempertimbangkan karakteristik item yang sebelumnya disukai atau dipilih oleh pengguna. Pendekatan ini berfokus pada kesesuaian antara atribut suatu item

dengan preferensi pengguna. Sementara itu, *Collaborative Filtering* menghasilkan rekomendasi dengan memanfaatkan pola interaksi atau preferensi dari sejumlah pengguna. Data tersebut dapat direpresentasikan dalam bentuk matriks pengguna-item, kemudian sistem mencari pengguna atau item yang memiliki tingkat kemiripan untuk menghasilkan rekomendasi. Adapun *Knowledge-Based Filtering* menggunakan pengetahuan khusus mengenai suatu domain untuk menentukan item yang paling sesuai dengan kebutuhan pengguna. Pendekatan ini mempertimbangkan hubungan antara karakteristik item dan kebutuhan atau kriteria yang ditentukan oleh pengguna (Arfisko & Wibowo, 2022).

2.5. Wisata

Wisata merupakan aktivitas perjalanan yang dilakukan oleh individu maupun kelompok dengan mendatangi suatu wilayah atau destinasi tertentu dalam kurun waktu tertentu. Kegiatan tersebut dapat dilakukan untuk berbagai tujuan, seperti memperoleh hiburan, melakukan rekreasi, mengembangkan diri, maupun mengenal karakteristik dan daya tarik yang terdapat pada suatu destinasi. Sementara itu, objek wisata merupakan suatu tempat atau kawasan yang memiliki daya tarik tertentu sehingga mampu memberikan pengalaman dan kepuasan bagi pengunjung yang datang (Harahap, 2018).

Objek wisata dapat berupa sumber daya alam maupun hasil karya manusia yang memiliki nilai daya tarik bagi wisatawan. Bentuknya antara lain kawasan pegunungan, pantai, flora dan fauna, taman atau kebun binatang, peninggalan sejarah, bangunan kuno, monumen, candi, kesenian tradisional, pertunjukan, serta berbagai bentuk kebudayaan lokal lainnya (Ananto, 2018). Suatu daerah dapat

dikembangkan menjadi destinasi wisata apabila mempunyai karakteristik atau keunikan yang mampu menarik minat pengunjung. Keunikan tersebut dapat berasal dari kondisi alam dan lingkungan, keberagaman flora dan fauna, kebudayaan masyarakat, perkembangan teknologi, maupun nilai-nilai sosial dan spiritual yang menjadi ciri khas daerah tersebut.

