

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pariwisata merupakan bagian penting dalam sistem penggerak ekonomi daerah hingga nasional yang memiliki peranan penting dalam menggerakkan roda ekonomi dan penyumbang pendapatan nasional dan daerah (Prasetyo, Atina, and Purwanto 2021). Potensi pariwisata di Indonesia yang besar dengan adanya geografis yang luas serta keberagaman alam, budaya, dan tradisi yang terdapat pada masyarakat dengan keunikan setiap daerah. Indonesia yang merupakan negara kepulauan dengan 17000 pulau yang terdapat didalamnya memberikan keragaman dan daya tarik wisata. Budaya dan tradisi lokal melekat dalam sistem kehidupan masyarakat menghadirkan ciri khas setiap daerah seperti upacara adat, kesenian, kuliner dan peninggalan zaman sejarah seperti bangunan candi dan bangunan kolonial memberikan minat daya tarik. Potensi ini harus dikelola dan diperhatikan oleh pemerintah untuk digunakan sebagai pilar dalam melakukan pengembangan ekonomi di sektor ekonomi. Oleh karena itu pengelolaan potensi pariwisata harus mendapatkan perhatian secara berkelanjutan untuk menonjolkan pariwisata dan membantu pemanfaatan ekonomi di seluruh daerah (Kumala, Soelistyo, and Nuraini 2017). Provinsi Jawa Timur sendiri memiliki potensi besar dalam hal pariwisata dengan beragam bentang alam, budaya dan keunikan tiap daerah. Dorongan pengembangan potensi pariwisata dalam membangun akomodasi fasilitas pendukung seperti sarana transportasi dan sentra wisata. Keberadaan destinasi wisata dan budaya di Jawa timur dapat menjadi komoditas unggulan sektor pariwisata misalnya Kawah Ijen yang terletak di Bondowoso dan Kesenian Ludruk dapat menjadi contoh dalam penguatana citra pariwisata di Jawa Timur.

Sistem rekomendasi adalah sistem yang digunakan untuk membantu seseorang dalam membantu pengguna untuk menentukan pilihan terbaik dari informasi data berdasarkan preferensi pengguna (Afridi 2019). Dalam pariwisata rekomendasi berperan dalam membantu pengguna untuk mempermudah wisatawan menemukan destinasi, aktivitas, akomodasi, atau paket wisata yang paling sesuai dengan minat mereka. rekomendasi dapat memberikan saran yang lebih personal dan relevan sehingga pengalaman perjalanan menjadi lebih optimal. Dalam konteks ini,

penggunaan metode *Seasonal Context Aware* berbasis *Content Based Filtering* karena dalam sistem ini memberikan pertimbangan pada faktor musiman, kondisi cuaca, atau periode kunjungan tertentu untuk memberikan hasil akurat berdasarkan ulasan dari setiap pengguna yang relevan. *Content Based Filtering* berfokus pada analisis pada kemiripan tempat destinasi, kategori, karakteristik dari tempat yang disukai oleh tiap user. Pendekatan *Content Based Filtering* dapat memberikan hasil yang akurasi prediksi relevan dengan memberikan preferensi secara langsung berdasarkan pengalaman pengguna (Javed et al. n.d.)(Fields 2022). Metode ini dinilai paling tepat dalam sistem *seasonal context-aware* karena mampu memadukan analisis preferensi pengguna dan karakteristik destinasi dengan pertimbangan faktor temporal dan musiman (Huang et al. 2018).

Kebutuhan sistem diperlukan dalam mendukung rekomendasi pariwisata untuk menentukan destinasi yang sesuai dengan wisatawan terhadap destinasi yang akan dikunjungi. Informasi yang umum tersedia dalam platform wisata tergolong bersifat umum dan tidak mampu menyesuaikan dengan preferensi pengguna secara spesifik. Dalam beberapa aplikasi seperti, situs pariwisata daerah biasanya hanya menampilkan rekomendasi berdasarkan popularitas umum bukan karena preferensi personal dengan pendekatan *seasonal-context aware*. Oleh karena itu, Sistem rekomendasi diperlukan untuk melakukan pengurangan keterbasan dalam memuat informasi untuk setiap pengguna. Sistem yang relevan dapat menentukan karakteristik dan ketentuan dari tiap user seperti adanya ulasan destinasi, penilaian pengguna lain, minat, musim, akses, transportasi dan ciri khas dalam tiap pariwisata. Peranan sistem rekomendasi untuk menyaring dan menampilkan data yang sesuai dengan karakteristik pengunjung melalui relevansi data yang diambil dari preferensi ulasan, penilaian pengunjung dan pola perilaku pengguna sebelumnya. Pendekatan yang akurat perlu akurat perlu digunakan untuk memberikan hasil rekomendasi yang akurat.

Pada penelitian (Flórez et al. 2025), pengembangan sistem rekomendasi pariwisata berbasis *context-aware* dengan mengombinasikan *deep learning* dan *ontology-based knowledge*. Pada sistem ini data yang digunakan adalah melalui preferensi pengguna, lokasi geografis, biaya, dan popularitas destinasi. Penelitian ini menunjukkan hasil rekomendasi personal secara real-time dengan memberikan

rekomendasi pariwisata berkelanjutan melalui fitur geofencing. Pada penelitian (Alrasheed 2020), melalui pendekatan *multi-level collaborative filtering* dan *Content Based Filtering* dengan perangkingan melalui preferensi pengguna dengan data yang didapatkan melalui *web scraping* dari berbagai situs pariwisata. Sistem rekomendasi pada penelitian ini bekerja dalam dua tahap, yaitu penyaringan destinasi berdasarkan kesamaan pengguna dan pengurutan destinasi berdasarkan preferensi individual. Pemanfaatan data ulasan pengguna masih terbatas pada analisis sentimen, sehingga belum mampu menangkap kesamaan makna preferensi pengguna secara semantik mendalam.

Pada penelitian (Yulfihani and Zakariyah 2024), dikembangkan sistem rekomendasi destinasi wisata di Kabupaten Batang menggunakan metode *Content-Based Filtering* dengan mempertimbangkan kesamaan kategori wisata dan kedekatan lokasi. Kesamaan kategori direpresentasikan secara biner, sedangkan kedekatan lokasi dihitung menggunakan rumus Haversine, kemudian kedua nilai tersebut digabungkan dengan bobot tertentu untuk menghasilkan skor kemiripan akhir yang digunakan dalam pemeringkatan rekomendasi. Sistem menunjukkan kinerja yang tinggi dengan rata-rata F1-score sebesar 0,965, khususnya pada skenario preferensi pengguna yang sederhana. Namun, penulis juga melaporkan bahwa performa sistem menurun pada kasus dengan kombinasi preferensi yang lebih kompleks, serta bahwa cakupan kategori dan jumlah data masih terbatas. Penelitian yang dilakukan oleh.

Pada penelitian (Agrawal, Roy, and Mitra 2021), sistem rekomendasi *Point of Interest* dibangun dengan memanfaatkan tag *embedding* berbasis Word2Vec untuk merepresentasikan *Point Of Interest* (POI) dan profil pengguna secara semantik. Hasil penelitian menunjukkan bahwa representasi berbasis *embedding* lebih unggul dibandingkan metode *one-hot* dalam meningkatkan kualitas rekomendasi. Namun, sumber informasi yang digunakan masih terbatas pada tag singkat sehingga belum mampu merepresentasikan preferensi pengguna secara mendalam melalui ulasan teks yang kaya makna. Selain itu, *embedding* yang digunakan masih berbasis *Word2Vec* yang bersifat non-kontekstual dan belum dioptimalkan untuk kebutuhan *semantic retrieval modern*. Sementara itu, penelitian (Kulkarni and Rodd 2020), menunjukkan bahwa sebagian besar *Context Aware Recommendation Systems*

masih bertumpu pada pendekatan *Content-Based Filtering*, *Collaborative Filtering*, serta teknik statistik dan bio-inspired computing yang mengandalkan rating, atribut eksplisit, dan kata kunci sebagai representasi item. Meskipun berbagai teknik kontekstual telah dikembangkan, sistem rekomendasi masih menghadapi permasalahan utama seperti cold start, data sparsity, serta keterbatasan dalam menangkap perubahan preferensi pengguna yang bersifat laten dan semantik.

Pada penelitian ini penggunaan *BAIL General Embedding* (BGE) dalam sistem digunakan untuk meningkatkan akurasi dalam pencarian kesamaan makna (*semantic similarity*) pada data teks. Penggunaan vektorisasi dapat memberikan hasil identifikasi dari hasil ulasan yang terdapat pada ulasan tiap pengguna dalam membantu merekomendasikan tempat pariwisata. BGE dikembangkan dari arsitektur M3-Embedding yang menggunakan *self-knowledge distillation* dalam menghasilkan representasi semantik lintas bahasa yang akurat dan efisien, mendukung berbagai fungsi retrieval dalam satu model tunggal (Chen et al. 2024). Setelah itu, teknik *Cosine Similarity* digunakan dalam menghitung tingkat kemiripan antar vektor yang mempresentasikan ulasan pengguna dalam menentukan destinasi wisata yang relevan dengan preferensi pengguna, sehingga rekomendasi yang diberikan dapat sesuai dengan setiap personal, karakteristik dan minat wisatawan. Dengan demikian, sistem ini diharapkan mampu memberikan rekomendasi destinasi pariwisata yang paling relevan dengan topik penelitian peneliti, serta mempertimbangkan faktor-faktor lain yang penting dalam proses menentukan destinasi wisata. Penelitian ini, menggabungkan kekuatan pemodelan semantik modern, personalisasi berbasis konten, serta faktor kontekstual yang dinamis. *Content-Based Filtering* (CBF) berbasis *BAIL General Embedding* (BGE) dengan penerapan *seasonal context-aware* dalam penelitian ini, digunakan karena mampu memberikan rekomendasi berdasarkan karakteristik konten dan preferensi individual pengguna dengan hasil vektor yang diberikan melalui proses *embedding* dalam hasil vektor BGE. Selain itu integrasi *seasonal context-aware* dalam penelitian ini dilakukan untuk mengakomodasi perubahan karakteristik destinasi wisata berdasarkan musim. Aspek cuaca dipehitungkan karena terhadap kenyamanan, daya tarik, dan kesesuaian destinasi bagi wisatawan.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Provinsi Jawa Timur memiliki sejumlah potensi besar pada sektor pariwisata dimana potensi yang dimiliki melalui keberagaman alam dan budaya. Keanekaragaman destinasi dan tempat wisata yang dimiliki membuat banyak wisatawan mengalami kesulitan dalam menentukan destinasi yang dituju. Selain itu, penentuan kesuaian preferensi, kondisi, dan waktu kunjungan membuat wisatawan juga harus memperhatikan destinasi yang tepat dalam menentukan lokasi wisata. Keterbatasan informasi juga menjadi masalah yang dihadapi oleh wisatawan dalam memilih destinasi.

Platform pariwisata sebagian besar hanya menyediakan informasi secara umum tanpa mempertimbangkan preferensi pribadi dengan pendekatan *seasonal context-aware*. Kondisi ini membuat wisatawan sulit menemukan destinasi yang sesuai dengan minat dan waktu kunjungan mereka. Permasalahan tersebut menimbulkan dengan informasi yang tersedia, di mana rekomendasi yang diberikan sering kali tidak relevan dengan preferensi individu. Selain itu, variasi kondisi lingkungan, musim, serta perbedaan karakteristik tiap daerah wisata juga memperumit proses pengambilan keputusan wisatawan dalam memilih destinasi yang tepat.

1.2.2 Metode Usulan

Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan metode *Content-Based Filtering* berbasis *seasonal context-aware* dengan integrasi BAII *General Embedding* (BGE) dalam sistem rekomendasi pariwisata. Metode ini menggabungkan bekerja dengan pemanfaatan karakteristik destinasi wisata untuk menghasilkan rekomendasi yang sesuai dengan preferensi pengguna. *Content-Based Filtering*, memiliki kinerja dalam memanfaatkan karakteristik destinasi wisata seperti kategori wisata, lokasi, dan ulasan pengguna yang menyesuaikan hasil rekomendasi berdasarkan kesamaan dengan destinasi sebelumnya yang diminati oleh pengguna.

BAII *General Embedding* (BGE) berfungsi sebagai representasi tesk yang mengubah ulasan dalam bentuk vektor numerik. Pendekatan ini berfungsi dalam

memberikan makna semantik . Integrasi *seasonal context-aware* digunakan untuk menambahkan aspek waktu dan kondisi musiman ke dalam proses rekomendasi. Dengan mempertimbangkan faktor-faktor seperti musim hujan atau kemarau, periode liburan, dan kecenderungan kunjungan wisatawan, sistem dapat menyesuaikan hasil rekomendasi agar lebih relevan dengan kondisi aktual pariwisata di Jawa Timur. Pendekatan ini diharapkan mampu memberikan hasil rekomendasi yang lebih personal, relevan, dan adaptif terhadap perubahan konteks musiman yang memengaruhi minat wisatawan.

1.2.3 Pertanyaan Penelitian

Berdasarkan uraian permasalahan dan metode yang dijelaskan, muncul beberapa pertanyaan pada penelitian sebagai berikut :

1. Bagaimana perbandingan kinerja metode *Content-Based Filtering* berbasis TF-IDF dengan *Content-Based Filtering* berbasis *embedding* BGE dalam sistem rekomendasi pariwisata?
2. Bagaimana penerapan transformasi *max-pooling* pada *embedding* BGE berpengaruh signifikan terhadap peningkatan kualitas fitur dan hasil rekomendasi dalam metode *Content-Based Filtering*?
3. Bagaimana penggunaan *embedding* BGE baseline multibahasa menghasilkan performa rekomendasi yang lebih baik dibandingkan pendekatan *monolanguage* pada metode *Content-Based Filtering*?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengidentifikasi permasalahan yang dihadapi oleh wisatawan dalam menentukan destinasi wisata di Provinsi Jawa Timur untuk memberikan sistem rekomendasi yang membantu proses pengambilan keputusan. Penerapan metode *Content-Based Filtering* untuk meningkatkan relevansi rekomendasi pariwisata berdasarkan relevansi pengguna. Integrasi *BAAI General Embedding* (BGE) sebagai metode representasi teks dalam analisis ulasan wisatawan. Semantik teks dari hasil vektorisasi untuk melakukan penilaian makna secara mendalam untuk memberikan rekomendasi lebih relevan, personal, dan adaptif terhadap perubahan konteks musiman maupun preferensi wisatawan.

1.4 Manfaat Penelitian

Penelitian ini diharapkan memberikan pengembangan kajian ilmiah dalam bidang sistem rekomendasi khususnya penerapan metode *Content-Based Filtering* yang dipadukan dengan pendekatan *seasonal context-aware*. Dengan adanya sistem rekomendasi ini, analisis perilaku pengguna dan faktor musiman dapat meningkatkan relevansi rekomendasi, terutama pada konteks pariwisata. Melalui karakteristik dinamis dan kontekstual serta memberikan hasil rekomendasi relevan untuk wisatawan. Metode *embedding* yang terdapat pada *BAAI General Embedding (BGE)* dalam pemrosesan data teks pariwisata, yang selama ini masih jarang diterapkan pada sistem rekomendasi berbasis konteks di tingkat regional. Dengan demikian, sistem rekomendasi yang dihasilkan mampu memberikan saran yang lebih personal, relevan, dan efisien, sehingga dapat meningkatkan pengalaman berwisata di Provinsi Jawa Timur. Selain itu, melalui penelitian ini diharapkan dapat mendukung pihak terkait dalam memahami pola kunjungan wisatawan untuk pengembangan strategi promosi pariwisata yang lebih tepat sasaran.

1.5 Batasan Masalah

Batasan masalah dari penelitian ini adalah :

1. Dataset yang digunakan dalam ulasan dan penilaian pengguna didapatkan melalui *scrapping* yang terdapat pada platform *Google Maps*
2. Dataset yang diambil berupa tempat pariwisata yang terdapat pada Provinsi Jawa Timur
3. Pendekatan *seasonal context-aware* yang mempertimbangkan faktor waktu, musim, dan kondisi cuaca umum, tanpa melibatkan data *real-time* atau aspek sosial ekonomi wisatawan.

1.6 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun untuk memberikan gambaran yang terstruktur dan memudahkan pembaca dalam memahami alur penelitian secara keseluruhan. Adapun sistematika penulisan dalam penelitian ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini menguraikan mengenai gambaran umum penelitian yang meliputi latar belakang masalah, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, serta pertanyaan penelitian. Selain itu, pada bab ini juga disajikan sistematika penulisan sebagai panduan alur pembahasan dalam penelitian.

BAB 2 TINJAUAN PUSTAKA

Pada bagian ini disajikan dasar teori dan metode yang digunakan sebagai dasar dan alat untuk mengatasi masalah serta uraian beberapa hal yang terkait dalam penelitian.

BAB 3 METODE PENELITIAN

Bab ini menjelaskan secara rinci metodologi penelitian yang digunakan. Pembahasan mencakup tahapan penelitian, perancangan sistem, sumber dan karakteristik data, teknik pengumpulan data, metode pengolahan dan prapemrosesan data, skema pengujian, metode evaluasi sistem, serta flowchart atau alur kerja penelitian