

ABSTRAK

Pariwisata merupakan salah satu penggerak sektor perekonomian daerah hingga nasional di Indonesia dengan potensi yang sangat besar dari segi alam, budaya, dan tradisi. Provinsi Jawa Timur sebagai salah satu provinsi dengan keberagaman destinasi wisata dan ragam budaya didalamnya memberikan potensi besar dalam sektor pariwisata. Namun, Informasi pariwisata yang tersedia saat ini masih bersifat umum dan belum mampu menyesuaikan dengan preferensi spesifik pengguna. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem rekomendasi destinasi wisata berbasis *Content-Based Filtering* dengan pendekatan *Seasonal Context-Aware* menggunakan *BAll General Embedding* (BGE) sebagai semantik vektor. BGE digunakan untuk merepresentasikan ulasan pengguna dan karakteristik destinasi dalam bentuk vektor semantik sehingga mampu menangkap kesamaan makna secara mendalam. Sistem ini dirancang untuk melakukan rekomendasi pariwisata berdasarkan Top 10 rekomendasi paling relevan berdasarkan *query* yang dimasukkan oleh pengguna berdasarkan preferensi pengguna dan musim. Hasil evaluasi menunjukkan bahwa model BGE *Base* mencapai nilai *Precision@10* sebesar 0,928, *MAP@10* sebesar 0,900, dan *NDCG@10* sebesar 0,977. Hasil tersebut menunjukkan, sistem rekomendasi berbasis BGE terbukti mampu menghasilkan rekomendasi destinasi wisata di Provinsi Jawa Timur yang lebih personal, relevan, dan adaptif terhadap preferensi pengguna serta kondisi musiman.

Kata Kunci : Sistem Rekomendasi, *Content-Based Filtering*, *Seasonal Context-Aware*, *BAll General Embedding* (BGE), Pariwisata

ABSTRACT

Tourism is a primary driver of both the regional and national economy in Indonesia, possessing immense potential in terms of natural, cultural, and traditional aspects. East Java Province, characterized by its diverse tourist destinations and rich cultural heritage, offers significant potential within the tourism sector. However, currently available tourism information remains generalized and is not yet capable of adapting to specific user preferences. Therefore, this research aims to develop a tourist destination recommendation system based on Content-Based Filtering with a Seasonal Context-Aware approach, utilizing BAAI General Embedding (BGE) as semantic vectors. BGE is employed to represent user reviews and destination characteristics in the form of semantic vectors, enabling the profound capture of semantic similarities. The system is designed to provide the Top 10 most relevant tourism recommendations based on queries inputted by users, accounting for both user preferences and seasons. The evaluation results indicate that the BGE Base model achieved a Precision@10 of 0.928, a MAP@10 of 0.900, and an NDCG@10 of 0.977. These findings demonstrate that the BGE-based recommendation system is proven capable of generating tourist destination recommendations in East Java Province that are more personalized, relevant, and adaptive to user preferences and seasonal conditions.

Keyword : Recommendation System, Content-Based Filtering, Seasonal Context-Aware, BAAI General Embedding (BGE), Tourism

DAFTAR ISI

HALAMAN JUDUL	iii
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN ORISINALITAS	iii
MOTTO DAN PERSEMBAHAN.....	iv
KATA PENGANTAR.....	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
DAFTAR KODE PROGRAM	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.2.1 Permasalahan	5
1.2.2 Metode Usulan	5
1.2.3 Pertanyaan Penelitian.....	6
1.3 Tujuan Penelitian	6
1.4 Manfaat Penelitian	7
1.5 Batasan Masalah	7
1.6 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 Sistem Rekomendasi.....	9
2.2 <i>Natural Language Processing (NLP)</i>	10
2.3 <i>Content Based Filtering</i>	11
2.4 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	11
2.5 <i>BAAI General Embedding (BGE)</i>	13
2.6 <i>Seasonal Context Aware</i>	15
2.7 <i>Cosine Similarity</i>	16
2.8 Tinjauan Penelitian Terdahulu	17
BAB III METODE USULAN.....	26

3.1	Tahapan Alur Penelitian.....	26
3.1.1	<i>Crawling dan Scrapping</i>	27
3.1.2	<i>Preprocessing</i>	27
3.1.3	TF-IDF	29
3.1.4	<i>BAAI General Embedding</i>	31
3.1.5	<i>Content Based Filtering</i>	32
3.1.6	Rekomendasi Destinasi	39
3.1.7	Evaluasi.....	39
3.2	Dataset.....	42
3.3	Alur Kerja Sistem	44
3.4	Jenis Penelitian.....	45
BAB IV HASIL DAN PEMBAHASAN		46
4.1.	Proses Pengumpulan Data.....	46
4.2.	Proses Pengelolaan Data	48
4.2.1	<i>CaseFolding</i>	48
4.2.2	<i>Tokenization</i>	50
4.2.3	<i>Cleaning Text</i>	51
4.2.4	Normalisasi Data.....	53
4.2.5	<i>Stopword Removal</i>	57
4.2.6	<i>Whitespace Handling</i>	61
4.2.7	<i>Stemming</i>	61
4.3.	Proses Implementasi terhadap Rekomendasi	62
4.3.1.	Penerapan Seasonal Context Aware.....	62
4.3.2.	Implementasi Rekomendasi Menggunakan <i>Cosine Similarity</i>	68
4.4.	Implementasi Website Rekomendasi	71
4.4.1	Landing Page	71
4.4.2	Laman <i>Login</i>	72
4.4.3	Laman Register	73
4.4.4	Laman Dashboard Admin	73
4.4.5	Laman Manage Destinasi.....	74
4.4.6	Laman Manage Users.....	74
4.4.7	Laman Dashboard Pengguna	75
4.4.8	Laman Profil	76
4.4.9	Laman Edit Profil.....	76
4.4.10	Laman Detail Destinasi.....	77
4.4.11	Laman Pencarian.....	77

4.5.	Evaluasi.....	78
4.5.1.	TF-IDF	78
4.5.2.	BGE <i>Baseline Multilanguage</i>	79
4.5.3.	BGE Indonesia	80
4.5.4.	BGE <i>Hyperparameter</i>	81
4.5.5.	Perbandingan Model	82
BAB V SARAN DAN KESIMPULAN		84
DAFTAR PUSTAKA		86
LAMPIRAN.....		91
Lampiran 1	Kode Program.....	91
Lampiran 2	Hasil Cek Turnitin	113
Lampiran 3	Perhitungan Manual.....	114

DAFTAR GAMBAR

Gambar 2. 1. Karakteristik M3 Embedding (Chen et al. 2024)	14
Gambar 2. 2 Arsitektur Model BGE M3 Embedding (Chen et al. 2024).....	15
Gambar 3. 1 Arsitektur Tahapan Penelitian.....	26
Gambar 3. 2 Perhitungan TF-IDF	30
Gambar 3. 3. Proses embedding BGE.....	31
Gambar 3. 4 Alur Kerja Sistem	44
Gambar 4. 1. Flowchart Urutan Proses Scrapping	46
Gambar 4. 2. Hasil BGE Base Model	69
Gambar 4. 3. Hasil BGE Base <i>HyperParameter</i>	69
Gambar 4. 4. Gambar 4.3. Hasil BGE Indonesia	70
Gambar 4. 5. Hasil TF-IDF	71
Gambar 4. 6. Landing Page	71
Gambar 4. 7. Laman Login	72
Gambar 4. 8. Laman Register.....	73
Gambar 4. 9. Laman Dashboard.....	73
Gambar 4. 10. Gambar Manage Destinasi	74
Gambar 4. 11. Gambar Manage Users	74
Gambar 4. 12. Gambar Dashboard Pengguna	75
Gambar 4. 13. Laman Profil.....	76
Gambar 4. 14. Laman Edit Profile	76
Gambar 4. 15. Laman Detail Destinasi	77
Gambar 4. 16. Laman Pencarian	77
Gambar 4. 17. Evaluasi Hasil BGE Base Model.....	79
Gambar 4. 18. Evaluasi Hasil BGE Indonesia	80
Gambar 4. 19. Evaluasi Hasil BGE HyperParameter.....	81
Gambar 4. 20. Perbandingan Model.....	82

DAFTAR TABEL

Tabel 3. 1 Deskripsi wisata dan input query pengguna	33
Tabel 3. 2 <i>Preprocessing</i>	33
Tabel 3. 3. Kemunculan term frequency	33
Tabel 3. 4. Normalisasi Panjang Dokumen	34
Tabel 3. 5. Tabel Hasil Vektorisasi Menggunakan TF-IDF	35
Tabel 3. 6 Vektorisasi menggunakan BGE	37
Tabel 3. 7 Skenario Pengujian.....	42
Tabel 3. 8. Dataset	43
Tabel 4. 1. Data Scrapping	47
Tabel 4. 2. Casefolding.....	49
Tabel 4. 3. Tokenisasi	50
Tabel 4. 4. Cleaning Text.....	52
Tabel 4. 5. Normalisasi Data Text	56
Tabel 4. 6. Proses Stopword Removal.....	60
Tabel 4. 7. Proses Stemming	61

DAFTAR KODE PROGRAM

Kode Program 4. 1. Pseudocode Proses Normalisasi	53
Kode Program 4. 2. Pseudocode Stopword Removal.....	58
Kode Program 4. 3. Pseudocode Inisiasi Seasonal	63
Kode Program 4. 4 Pseudocode Similarity Musim	64
Kode Program 4. 5. Tabel Hasil Seasonal	66

DAFTAR LAMPIRAN

Lampiran 1 Kode Program.....	91
Lampiran 2 Hasil Cek Turnitin.....	113
Lampiran 3 Perhitungan Manual.....	114