 ISSN:2302-8149
e-ISSN:2540-9719

Sistem Rekomendasi Hotel di Yogyakarta berbasis *Content-Based Filtering* dengan Pemodelan *FastText*

Hotel Recommendation System in Yogyakarta based on Content-Based Filtering with FastText Modeling

[1]Dzikri Albantani*, [2]Arif Nur Rohman
[1,2]Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta
[1,2]Jl. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Sleman, Daerah Istimewa Yogyakarta, Indonesia
*e-mail: dzikrialbantani@students.amikom.ac.id

(***received*:** 16 April 2026, ***revised*:** 21 October 2026, ***accepted*:** 22 October 2026)

Abstrak

Pesatnya pertumbuhan pariwisata di Yogyakarta menyebabkan banyaknya pilihan akomodasi yang dapat menyulitkan wisatawan dalam menemukan hotel sesuai dengan kebutuhan. Sistem rekomendasi berbasis *Term Frequency-Inverse Document Frequency* (TF-IDF) memiliki keterbatasan dalam menangani variasi bentuk kata dan bahasa informal karena pencocokannya bergantung pada representasi leksikal. Penelitian ini bertujuan mengembangkan sistem rekomendasi hotel berbasis Content-Based Filtering dengan mengintegrasikan *word embedding* FastText serta membandingkannya dengan TF-IDF dalam menghadapi variasi bentuk bahasa formal dan informal. Penelitian menggunakan dataset 100 hotel di Yogyakarta yang dikumpulkan melalui *web scraping* dan validasi data. Evaluasi dilakukan menggunakan lima pasangan kueri formal–informal dengan intensi pencarian yang ekuivalen. Untuk setiap kueri dan metode, lima rekomendasi teratas digunakan untuk menghitung Mean Similarity@5 dan perubahan skor dari kueri formal ke informal. Hasil pengujian menunjukkan bahwa perubahan Mean Similarity@5 pada FastText cenderung lebih kecil dibandingkan TF-IDF pada lima pasangan kueri yang diuji. Perubahan terbesar pada TF-IDF mencapai 26,09%, sedangkan pada FastText sebesar 5,41%. Hasil tersebut mengindikasikan bahwa FastText relatif lebih stabil terhadap variasi bentuk bahasa formal–informal pada skenario pengujian penelitian ini. Namun, penelitian belum memiliki ground truth relevansi yang diperoleh secara independen sehingga hasil evaluasi tidak digunakan untuk menyatakan perbedaan akurasi rekomendasi. Analisis juga menunjukkan keterbatasan dalam menangani atribut numerik seperti harga dan hubungan sintaksis tertentu seperti negasi.

Kata kunci: *content-based filtering, fasttext, natural language processing,* sistem rekomendasi, pariwisata

Abstract

The rapid growth of tourism in Yogyakarta has resulted in a wide range of accommodation options, making it challenging for tourists to find hotels that meet their needs. Recommendation systems based on Term Frequency–Inverse Document Frequency (TF-IDF) have limitations in handling variations in word forms and informal language because their matching process relies on lexical representations. This study aims to develop a hotel recommendation system using Content-Based Filtering integrated with FastText word embeddings and to compare its performance with TF-IDF in handling variations between formal and informal language. The study uses a dataset of 100 hotels in Yogyakarta collected through web scraping and data validation. The evaluation was conducted using five pairs of formal and informal queries with equivalent search intents. For each query and method, the top five recommendations were used to calculate Mean Similarity@5 and measure score changes from formal to informal queries. The experimental results show that changes in Mean Similarity@5 for FastText tended to be smaller than those for TF-IDF across the five query pairs tested. The largest change observed for TF-IDF was 26.09%, compared with 5.41% for FastText. These findings indicate that FastText is relatively more robust to variations between formal and informal language under the evaluation scenarios considered in this study. However, the study does not use independently

 ISSN:2302-8149
e-ISSN:2540-9719

established ground-truth relevance labels; therefore, the evaluation results are not used to claim differences in recommendation accuracy. The analysis also reveals limitations in handling numerical attributes, such as price, and certain syntactic relationships, such as negation.
Keywords: *content-based filtering, fasttext, natural language processing, recommender system, tourism.*

1 Pendahuluan

Sektor pariwisata di Daerah Istimewa Yogyakarta (DIY) terus menunjukkan tren pemulihan yang signifikan pasca-pandemi, mengukuhkan posisinya sebagai salah satu destinasi unggulan nasional sebagaimana tercatat dalam data statistik [1]. Peningkatan aktivitas pariwisata tersebut juga diikuti oleh perkembangan industri perhotelan. Namun, di tengah tingginya adopsi teknologi digital di Indonesia [2], banyaknya pilihan akomodasi yang tersedia di berbagai platform *Online Travel Agent* (OTA) sering kali memunculkan fenomena *information overload*. Wisatawan kerap mengalami kesulitan dalam menyaring informasi secara manual untuk menemukan hotel yang benar-benar sesuai dengan preferensi spesifik mereka. Oleh karena itu, pengembangan sistem rekomendasi dapat menjadi salah satu pendekatan untuk membantu wisatawan menyaring informasi secara efisien [3].

Dalam domain sistem rekomendasi berbasis teks, Content-Based Filtering (CBF) dapat digunakan dengan memanfaatkan karakteristik atau representasi konten dari item untuk menghasilkan rekomendasi. Kusuma dan Musdholifah [4], misalnya, menerapkan CBF pada sistem rekomendasi topik tugas akhir dengan memanfaatkan representasi teks dan pembobotan TF-IDF untuk mencocokkan masukan pengguna dengan karakteristik dokumen yang direkomendasikan. Beberapa penelitian terdahulu telah berupaya membangun sistem rekomendasi hotel berbasis konten, seperti yang dilakukan oleh Sari et al. [5] untuk hotel di Palangka Raya dan Melyani [6] untuk studi kasus di Yogyakarta. Kedua penelitian tersebut mengandalkan algoritma pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengekstraksi fitur teks dari deskripsi hotel. Meskipun efisien secara komputasi, Al Rasyid et al. [7] menyoroti bahwa TF-IDF memiliki kelemahan mendasar karena cara kerjanya yang sangat bergantung pada kecocokan kata kunci eksak (*exact keyword matching*).

Permasalahan utama yang muncul dari penggunaan TF-IDF pada penelitian-penelitian sebelumnya [5], [6] adalah ketidakmampuannya menangkap konteks semantik atau makna kata, yang menyebabkan terjadinya kesenjangan leksikal (*lexical gap*). Kelemahan ini menjadi krusial karena penelitian-penelitian terdahulu tersebut belum secara spesifik mengevaluasi ketahanan model terhadap kueri bahasa informal, padahal variasi bahasa tidak baku merupakan karakteristik utama dari *User Generated Content* pada ulasan pariwisata. Metode berbasis statistik frekuensi ini sering gagal memproses kueri pengguna yang mengandung variasi kata, sinonim, istilah tidak baku (*slang*), atau kesalahan penulisan (*typo*) yang lumrah ditemukan pada ulasan pengguna di platform digital. Dalam representasi berbasis TF-IDF, kata yang memiliki bentuk permukaan berbeda seperti ‘murmer’ dan ‘murah’ tidak secara otomatis memperoleh representasi yang sama apabila keduanya diperlakukan sebagai token yang berbeda.

Untuk mengatasi keterbatasan representasi berbasis pencocokan kata, pendekatan *word embedding* dapat digunakan untuk membentuk representasi kata dalam ruang vektor. Salah satu metode yang relevan adalah FastText, yang merepresentasikan kata menggunakan karakter n-gram sehingga struktur sub-kata dapat dimanfaatkan dalam pembentukan representasi kata [8]. Pendekatan ini juga telah digunakan dalam pemrosesan teks berbahasa Indonesia, termasuk pada penelitian berbasis klasifikasi teks [8], [9]. Namun, penerapan *FastText* pada penelitian-penelitian tersebut masih sebatas pada tugas klasifikasi sentimen, dan belum dieksplorasi kemampuannya dalam mengukur kemiripan dokumen (*document similarity*) untuk konteks sistem rekomendasi.

Berdasarkan pembandingan kesenjangan pada literatur tersebut, penelitian ini bertujuan untuk mengembangkan sistem rekomendasi hotel di Yogyakarta dengan memperbarui metode konvensional TF-IDF menjadi pendekatan semantik menggunakan pemodelan *FastText*. Kebaruan (*novelty*) dari penelitian ini terletak pada integrasi pemodelan *FastText* untuk menjembatani *lexical gap* pada *Content-Based Filtering* berbasis ulasan pengguna, sekaligus mengevaluasi secara komparatif ketahanan model terhadap kueri bahasa informal dibandingkan dengan *baseline* TF-IDF. Manfaat

 ISSN:2302-8149
e-ISSN:2540-9719

praktis dari penelitian ini diharapkan dapat meningkatkan pengalaman wisatawan dalam menemukan akomodasi yang akurat dan relevan meskipun menggunakan preferensi bahasa yang tidak baku.

2 Tinjauan Literatur

Dinamika perkembangan teknologi informasi telah mendorong transformasi masif dalam industri pariwisata, khususnya dalam mekanisme penyebaran informasi akomodasi. Di tengah ledakan data digital saat ini, wisatawan sering kali dihadapkan pada masalah kelebihan informasi (*information overload*) yang menghambat proses pengambilan keputusan [2]. Secara garis besar, pendekatan sistem rekomendasi untuk mengatasi masalah ini terbagi menjadi dua paradigma utama: *Collaborative Filtering* (CF) dan *Content-Based Filtering* (CBF). Studi komparasi oleh Obajha, Sari, dan Pranatawijaya [10] menegaskan bahwa meskipun CF unggul dalam memanfaatkan data komunitas pengguna, metode ini dapat menghadapi masalah cold-start ketika informasi interaksi pengguna belum tersedia. Sebaliknya, Faurina dan Sitanggang [11] serta Putra, Rohman, dan Norhikmah [12] menunjukkan bahwa CBF menawarkan stabilitas yang lebih baik untuk domain pariwisata karena rekomendasi dibangun berdasarkan pencocokan karakteristik item secara mandiri, sehingga tetap dapat menghasilkan rekomendasi berdasarkan karakteristik konten item meskipun data interaksi pengguna terbatas.

Dalam implementasi CBF untuk rekomendasi hotel, sejumlah penelitian terdahulu masih terpaku pada ekstraksi fitur berbasis statistik. Penelitian oleh Sari, Licantik, dan Zahra [5] serta Melyani [6] berhasil menerapkan algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk merekomendasikan hotel di Palangka Raya dan Yogyakarta. Penerapan Content-Based Filtering pada domain pariwisata juga telah dilakukan dengan menggunakan representasi berbasis karakteristik item. Salsabilla dan Utomo [13] mengembangkan sistem rekomendasi wisata Kabupaten Pekalongan dengan menggunakan TF-IDF untuk memberikan bobot pada kategori wisata dan Cosine Similarity untuk mengukur tingkat kesamaan antaritem. Penelitian tersebut menggunakan delapan kategori wisata dan menunjukkan bahwa pendekatan tersebut dapat menghasilkan rekomendasi berdasarkan kesesuaian kategori dengan preferensi pengguna. Pendekatan serupa juga diterapkan oleh Maulana, Rohman, dan Pristyanto [14] pada sistem rekomendasi destinasi wisata Provinsi Aceh dengan memanfaatkan Cosine Similarity berdasarkan deskripsi destinasi. Penelitian tersebut mengombinasikan kemiripan konten dengan kedekatan geografis menggunakan Haversine Formula, setelah data teks melalui tahapan pra-pemrosesan seperti *case folding*, tokenisasi, *stopword removal*, dan *stemming*. Kedua penelitian tersebut menunjukkan bahwa *Content-Based Filtering* dapat diterapkan pada sistem rekomendasi pariwisata dengan memanfaatkan karakteristik konten sebagai dasar pembentukan kemiripan. Namun, penelitian tersebut belum membahas ketahanan representasi teks terhadap perubahan bentuk kueri formal dan informal, yang menjadi fokus penelitian ini. Wibowo, Indriati, dan Dewi [3] juga memperkuat tren ini dengan menyimpulkan efisiensi komputasi dari kombinasi TF-IDF dan *Cosine Similarity*. Meskipun demikian, pendekatan berbasis TF-IDF memiliki keterbatasan ketika variasi bentuk kata menyebabkan perubahan pada representasi leksikal. Al Rasyid, Handayani, dan Ningsih [7] mengidentifikasi bahwa pencocokan berbasis TF-IDF bergantung pada kesesuaian istilah dalam representasi teks. Situngkir dan Sianipar [15] juga membahas keterbatasan representasi berbasis TF-IDF dalam menangkap hubungan semantik antarkata. Keterbatasan tersebut menjadi relevan ketika sistem rekomendasi harus menangani kueri dengan variasi ejaan, singkatan, atau penggunaan bahasa informal.

Untuk memberikan gambaran komparatif yang komprehensif terkait transisi teknologi dalam mengatasi *lexical gap* tersebut, analisis kritis terhadap kelebihan dan keterbatasan masing-masing metode ekstraksi fitur disajikan pada Tabel 1.

Tabel 1 Analisis kritis metode ekstraksi fitur teks pada sistem rekomendasi

Metode	Pendekatan	Kelebihan	Keterbatasan Utama
TF-IDF	Statistik Frekuensi Kata	Merepresentasikan bobot berdasarkan frekuensi kata	Tidak merepresentasikan hubungan semantik secara eksplisit

 ISSN:2302-8149
e-ISSN:2540-9719

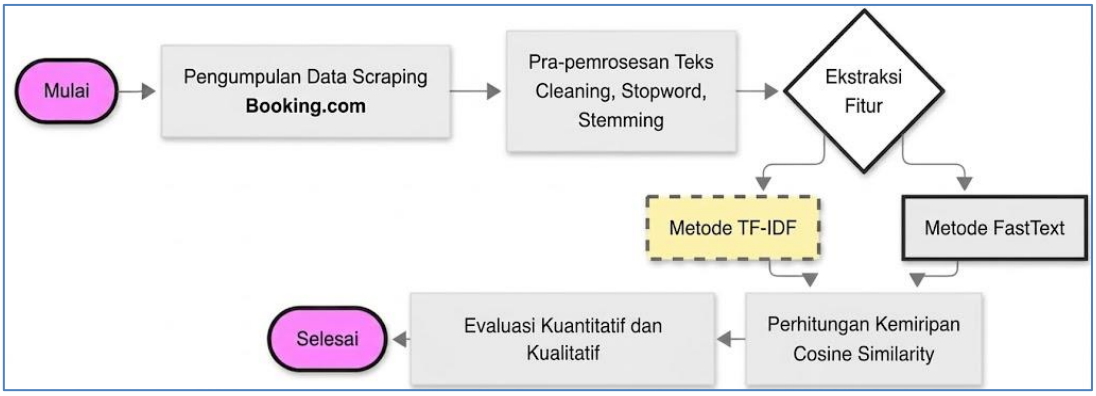
Word2Vec / GloVe	*Word embedding* (Tingkat Kata)	Merepresentasikan kata dalam ruang vektor	Kata yang tidak terdapat dalam kosakata pelatihan sulit direpresentasikan
FastText	*Sub-word embedding* (Karakter *n-grams*)	Memanfaatkan karakter n-gram	Representasi berbasis pooling tidak mempertahankan seluruh struktur sintaksis
Transformer (IndoBERT)	Contextualized Embedding	Mempertimbangkan konteks token dalam kalimat	Membutuhkan sumber daya komputasi lebih besar

Berdasarkan sintesis analisis kritis pada Tabel 1, dapat dilihat bahwa metode berbasis statistik frekuensi memiliki batasan inheren pada aspek semantik karena mengabaikan hubungan linguistik antar-kata. Gelombang penelitian terbaru mulai beralih pada pemodelan *Word embedding* seperti *FastText*. Berbeda dengan *Word2Vec* yang memperlakukan kata sebagai unit atomik, Gunawan, Young, dan Rusli [16] menjelaskan bahwa *FastText* memecah kata menjadi komponen karakter *n-grams*. Representasi *karakter n-gram* ini memberikan karakteristik yang berbeda, terutama dalam representasi *sub-word* dan penanganan variasi bentuk kata. Kemampuan model untuk "memahami" struktur internal kata ini dibuktikan secara empiris oleh Putri et al. [17], yang menyimpulkan bahwa *FastText* mampu menangkap informasi kontekstual yang jauh lebih kaya dibandingkan TF-IDF dalam mengelola data berisikan OOV. Validitas ini juga diperkuat oleh studi analisis sentimen dari Naldi dan Agustian [18] serta Sepriadi et al. [9] yang menemukan bahwa representasi *n-grams FastText* memiliki ketahanan (*robustness*) tinggi untuk mengasosiasikan kata tidak baku (seperti *slang* atau singkatan) dengan bentuk bakunya berdasarkan kemiripan karakter.

Meskipun keunggulan teoretis FastText dalam memproses teks informal (*User Generated Content*) telah banyak divalidasi pada domain klasifikasi teks, pemanfaatannya dalam sistem rekomendasi *Content-Based Filtering* masih sangat minim dieksplorasi. Penelitian sistem rekomendasi hotel di Yogyakarta sebelumnya [6] hanya berfokus pada pendekatan TF-IDF deskriptif. Oleh karena itu, penelitian ini mengisi *research gap* tersebut dengan merancang sistem rekomendasi berbasis FastText untuk menguji secara komparatif apakah pendekatan semantik tingkat sub-kata ini mampu menjadi salah satu pendekatan untuk mengurangi dampak *lexical gap* yang ditinggalkan oleh penelitian-penelitian sebelumnya.

3 Metode Penelitian

Penelitian ini dirancang menggunakan pendekatan eksperimental untuk menguji performa sistem rekomendasi akomodasi pariwisata. Fokus utama penelitian adalah melakukan analisis komparasi antara metode konvensional TF-IDF yang bekerja berdasarkan statistik frekuensi kata (*lexical matching*), dengan metode FastText yang berbasis *word embedding* dengan memanfaatkan informasi sub-kata. Alur kerja sistem disusun secara sistematis mulai dari pengumpulan data, pra-pemrosesan teks, ekstraksi fitur, hingga evaluasi relevansi rekomendasi sebagaimana diilustrasikan pada Gambar 1.



Gambar 1 Alur penelitian

 ISSN:2302-8149
e-ISSN:2540-9719

Berdasarkan Gambar 1, kerangka kerja penelitian disusun secara sekuensial. Tahapan diawali dengan Pengumpulan Data untuk membangun dataset profil hotel di Yogyakarta. Data mentah selanjutnya memasuki tahap Pra-pemrosesan Teks guna membersihkan *noise* dan menstandarisasi format teks. Setelah data bersih, proses dilanjutkan ke tahap Ekstraksi Fitur menggunakan dua skenario terpisah: TF-IDF (sebagai *baseline*) dan FastText (sebagai metode usulan). Tahap akhir meliputi perhitungan *Cosine Similarity* untuk menghasilkan daftar rekomendasi dan mengevaluasi ketahanan representasi terhadap perubahan bentuk kueri.

3.1 Pengumpulan Data

Penelitian ini menggunakan pendekatan hybrid dalam pengumpulan data, menggabungkan teknik otomasi *web scraping* dan validasi semi-manual untuk menjamin kualitas data (*data fidelity*). Sumber data utama adalah platform *Online Travel Agent* (OTA) Booking.com dengan fokus wilayah studi Yogyakarta. Proses akuisisi hingga finalisasi data dilakukan melalui tiga tahapan utama:

1. **Akuisisi Data Otomatis**
   Tahap awal dilakukan menggunakan alat bantu ekstensi peramban *Instant Data Scraper*. Alat ini digunakan untuk mengekstraksi atribut data mentah (*raw data*) secara massal, meliputi nama hotel, harga, rating numerik, dan lokasi properti. Hasil ekstraksi disimpan dalam format *Comma-Separated Values* (*.csv*).
2. **Pembersihan Data Komputasional**
   Data mentah diproses menggunakan skrip bahasa pemrograman Python (dengan pustaka Pandas) untuk standardisasi struktur. Proses ini meliputi penyeragaman label kolom, dan konversi tipe data string ke numerik
3. **Pengayaan & Validasi Data Semi-Manual**
   Untuk memastikan data deskriptif yang sesuai dan bermakna, dilakukan proses kurasi manual terhadap kolom Deskripsi dan *Review User*

Setelah melalui proses pembersihan, dataset final terdiri dari atribut: Nama Hotel, URL, Rating, Harga, Lokasi, *Review User*, dan Deskripsi. Dimana Atribut Harga, Rating, dan Lokasi digunakan sebagai informasi pendukung pada analisis hasil, sedangkan representasi teks untuk proses rekomendasi dibentuk dari atribut Deskripsi dan Ulasan Pengguna.

3.2 Pra-pemrosesan Data

Mengingat data deskripsi dan ulasan hotel berasal dari platform publik yang sering kali mengandung format tidak baku, dilakukan tahapan pra-pemrosesan komprehensif untuk menjaga kualitas fitur. Langkah-langkah yang dilakukan meliputi:

1. **Case Folding & Data Cleaning:** Mengonversi seluruh karakter menjadi huruf kecil (*lowercase*) serta membersihkan dokumen dari karakter *noise* (tanda baca, angka, simbol mata uang, *emoticon*, dan URL).
2. **Stopword Removal**: Menghilangkan kata-kata umum yang tidak memiliki bobot makna spesifik menggunakan pustaka NLTK. Pada tahap ini juga mendefinisikan *custom stopwords* tambahan untuk mengeliminasi kata hubung tidak baku yang tidak memiliki nilai semantik pada korpus pariwisata.
3. **Stemming:** Mengubah kata berimbuhan menjadi kata dasar menggunakan pustaka Sastrawi agar algoritma dapat berfokus pada inti makna kata (misalnya: "menginap" direduksi menjadi "inap").

3.3 Algoritma Pembobotan & Ekstraksi Fitur

Profil tekstual hotel dibentuk dari gabungan atribut Deskripsi dan Ulasan Pengguna. Kedua atribut tersebut terlebih dahulu melalui tahap pra-pemrosesan secara terpisah. Hasil pra-pemrosesan dari kedua atribut kemudian digabungkan ke dalam atribut text_feature. Atribut text_feature selanjutnya digunakan sebagai input dalam pembentukan representasi teks menggunakan TF-IDF dan FastText.

1. **Term Frequency-Inverse Document Frequency (TF-IDF)**

 ISSN:2302-8149
e-ISSN:2540-9719

Metode TF-IDF digunakan sebagai *baseline* (tolok ukur). Metode ini menghitung bobot kata berdasarkan frekuensi kemunculannya dalam satu dokumen (TF) berbanding terbalik dengan kemunculannya di seluruh koleksi dokumen (IDF) [7]. Bobot kata (W_{ij}) dihitung menggunakan Persamaan (1).

$$W_{ij} = TF_{ij} \times log(\frac{N}{df_1}) \quad (1)$$

Keterangan: TF_{ij} adalah frekuensi kemunculan kata i dalam deskripsi hotel. j ,N adalah total jumlah hotel, dan df_i adalah jumlah dokumen hotel yang mengandung kata i.

2. FastText Embedding

Skenario kedua menerapkan model FastText yang merepresentasikan kata sebagai kumpulan sub-kata atau karakter *n-grams*. Pendekatan ini memungkinkan model untuk mempelajari morfologi kata secara internal dan menangani kata *Out-of-Vocabulary* (OOV). Fungsi penilaian (*scoring function*) antara kata dan konteks didefinisikan pada Persamaan (2).

$$s(w, c) = \sum_{g \epsilon G_w} \quad z_g v_c \quad (2)$$

Dimana G_w adalah himpunan *n-grams* yang menyusun kata, z_g adalah vektor representasi untuk setiap *n-gram*, dan v_c adalah vektor representasi untuk konteks kata. Representasi vektor untuk satu entitas hotel diperoleh melalui proses *Mean Pooling* (merata-ratakan vektor kata) untuk menghasilkan satu vektor padat (*dense vector*).

3.4 Perhitungan Kemiripan (Similarity Measure)

Setelah atribut Deskripsi dan Ulasan Pengguna melalui pra-pemrosesan secara terpisah, keduanya digabungkan menjadi text_feature. Tingkat kemiripan antara hotel yang disukai pengguna kueri dengan hotel lain di basis data dihitung menggunakan metode *Cosine Similarity* [15]. Nilai kemiripan dihitung menggunakan Persamaan (3).

$$Similarity(A, B) = \frac{\sum_{n=1}^{n} \quad A_i B_i}{\sqrt{\sum_{n=1}^{n} \quad (A_i)^2} \times \sqrt{\sum_{n=1}^{n} \quad (B_i)^2}} \quad (3)$$

Nilai hasil perhitungan berkisar antara −1 sampai 1 secara umum. Nilai yang semakin mendekati 1 menunjukkan kemiripan arah vektor yang semakin tinggi.

3.5 Evaluasi Ketahanan Representasi

Evaluasi dilakukan untuk membandingkan karakteristik representasi TF-IDF dan FastText dalam menghasilkan rekomendasi hotel ketika bentuk bahasa pada kueri mengalami perubahan. Karena penelitian ini tidak memiliki data interaksi pengguna, riwayat klik, rating relevansi, maupun ground truth relevansi yang diperoleh dari anotator independen, evaluasi kuantitatif tidak digunakan untuk menentukan benar atau salahnya rekomendasi. Sebagai gantinya, evaluasi difokuskan pada ketahanan representasi terhadap perubahan kueri dari bentuk formal menjadi bentuk informal dengan intensi pencarian yang tetap dipertahankan.

1. Evaluasi Kuantitatif

Evaluasi kuantitatif dilakukan untuk membandingkan ketahanan representasi TF-IDF dan FastText terhadap perubahan bentuk kueri dari formal menjadi informal. Penelitian ini tidak memiliki data interaksi pengguna maupun *ground truth* relevansi yang diperoleh dari anotator independen. Oleh karena itu, Cosine Similarity tidak digunakan untuk menentukan benar atau salahnya rekomendasi, tetapi sebagai indikator internal untuk mengamati perubahan konsistensi representasi ketika bentuk bahasa kueri berubah. Pendekatan ini sesuai dengan tujuan evaluasi ketahanan (*robustness*) terhadap variasi bahasa formal–informal.

Pengujian menggunakan lima pasangan kueri yang dipilih secara purposif untuk merepresentasikan beberapa bentuk variasi informal, yaitu slang, singkatan, perubahan ejaan, dan penggunaan bentuk preposisional informal, dengan mempertahankan intensi pencarian.

 ISSN:2302-8149
e-ISSN:2540-9719

Setiap pasangan diuji menggunakan TF-IDF dan FastText, kemudian masing-masing metode menghasilkan lima rekomendasi teratas beserta nilai Cosine Similarity. Rata-rata skor lima rekomendasi dihitung menggunakan Mean Similarity@5:

$$\boldsymbol{Mean\ Similarity@5} = \frac{\boldsymbol{S_1+S_2+S_3+S_4+S_5}}{\boldsymbol{5}} \quad (4)$$

$\boldsymbol{S_1, S_2, S_3, S_4,}$ dan $\boldsymbol{S_5}$ masing-masing merupakan nilai *Cosine Similarity* pada rekomendasi peringkat pertama hingga kelima. Selanjutnya, perubahan nilai Mean Similarity@5 antara kueri formal dan kueri informal dihitung menggunakan Persamaan (5).

$$\boldsymbol{Perubahan}(\%) = \frac{\boldsymbol{S_{formal-S_{informal}}}}{\boldsymbol{S_{formal}}} \times \boldsymbol{100}\% \quad (5)$$

S_{formal} merupakan Mean Similarity@5 pada kueri formal dan $S_{informal}$ merupakan Mean Similarity@5 pada kueri informal. Nilai positif menunjukkan penurunan skor pada kueri informal, sedangkan nilai negatif menunjukkan peningkatan skor. Nilai yang mendekati 0% menunjukkan bahwa perubahan skor antara kedua bentuk kueri relatif kecil. Perbandingan dilakukan dalam metode yang sama, sedangkan perbedaan nilai absolut *Cosine Similarity* antara TF-IDF dan FastText tidak ditafsirkan sebagai perbedaan akurasi karena keduanya menggunakan ruang representasi yang berbeda.

2. Evaluasi Kualitatif

Evaluasi kualitatif dilakukan melalui analisis *Top-N Recommendation* secara deskriptif dengan membandingkan lima rekomendasi teratas yang dihasilkan TF-IDF dan FastText pada kueri formal dan informal. Analisis difokuskan pada perubahan susunan rekomendasi dan karakteristik hotel yang muncul ketika bentuk bahasa kueri berubah.

Analisis ini digunakan untuk melengkapi hasil kuantitatif dengan mengamati kemampuan masing-masing metode dalam mempertahankan pola rekomendasi pada variasi bahasa formal–informal, sekaligus mengidentifikasi keterbatasan representasi yang digunakan. Hasil analisis kualitatif juga digunakan untuk melihat kasus ketika kemiripan semantik yang tinggi belum tentu mencerminkan seluruh intensi pengguna, terutama pada atribut terstruktur seperti harga dan hubungan sintaksis seperti negasi.

4 Hasil dan Pembahasan

4.1 Karakteristik Dataset

Sesuai dengan tahap pengumpulan data, penelitian ini menggunakan dataset 100 hotel di Yogyakarta. Setiap entitas data memiliki dua komponen teks utama: deskripsi formal dari manajemen hotel dan ulasan informal dari pengguna. Kombinasi tersebut menyediakan variasi teks formal dan informal yang digunakan sebagai sumber representasi teks dalam penelitian. Sampel data mentah disajikan pada Tabel 2.

Tabel 2 Contoh dataset hotel yogyakarta

Nomor	Nama Hotel	Cuplikan Ulasan Pengguna	Cuplikan Deskripsi
1	Hotel Limaran 1	Sebagai tempat menginap murah bersama pasangan, hotel ini cukup baik...	Berjarak 5 menit berkendara dari kawasan Malioboro yang ramai...
2	Puri Sagan	Kamar keluarga di Puri Sagan, Yogyakarta, menawarkan tempat tidur yang nyaman...	Puri Sagan mempunyai balkon dan terletak di Yogyakarta...
…	…	…	…
100	KHAS Tugu Hotel	Dapat dicapai dengan berjalan kaki ke jalan Malioboro...	Pesonna Tugu Yogyakarta menawarkan kamar-kamar dengan perabotan modern...

Berdasarkan Tabel 2, terlihat bahwa dataset memiliki dua karakteristik linguistik yang kontras. Kolom Deskripsi menggunakan bahasa baku yang terstruktur, sedangkan kolom *Review User*

 ISSN:2302-8149
e-ISSN:2540-9719

mengandung banyak opini subjektif dengan gaya bahasa bebas. Keberadaan ulasan pengguna ini penting karena menyediakan variasi kata yang tidak ditemukan dalam deskripsi resmi, seperti kata "murah", "nyaman", atau "strategis", yang sering menjadi kata kunci pencarian utama wisatawan.

4.2 Hasil Pra-pemrosesan Teks

Tahapan pra-pemrosesan teks bertujuan untuk mengurangi *noise* dan menyeragamkan bentuk teks sebelum dilakukan ekstraksi fitur. Hasil pra-pemrosesan pada beberapa contoh data ditunjukkan pada Tabel 3.

Tabel 3 Perbandingan data mentah dan hasil pra-pemrosesan

Atribut	Data Mentah (Raw Data)	Data Hasil Pra-pemrosesan
Deskripsi	Berjarak 5 menit berkendara dari kawasan Malioboro yang ramai, Hotel Limaran 1 terletak di Yogyakarta…	jarak menit kendara kawasan malioboro ramai limar letak yogyakarta...
Review User	Sebagai tempat menginap murah bersama pasangan, hotel ini cukup baik. Kamarnya kecil tetapi memiliki AC...	murah pasang kamar milik ac tv pancur bagus bangun tua tambah karakter...

Sebagaimana ditunjukkan pada Tabel 3, proses pra-pemrosesan mengubah teks menjadi bentuk token yang lebih seragam melalui tahapan pembersihan, penghapusan stopword, dan stemming. Sebagai contoh, kata “berkendara” diubah menjadi “kendara”, sedangkan kata yang termasuk stopword seperti “dari” dan “yang” dihilangkan. Hasil tersebut menghasilkan representasi teks yang lebih ringkas dengan mempertahankan kata-kata yang digunakan dalam tahap ekstraksi fitur.

Pada data ulasan pengguna, kata-kata seperti “murah”, “ac”, dan “tv” tetap dipertahankan karena dapat menjadi bagian dari informasi yang digunakan dalam proses pencocokan antara kueri dan profil hotel. Hasil pra-pemrosesan selanjutnya digunakan sebagai masukan untuk pembentukan representasi TF-IDF dan FastText pada tahap ekstraksi fitur.

4.3 Hasil Ekstraksi Fitur dan Analisis Cosine Similarity

Teks hasil pra-pemrosesan selanjutnya diubah menjadi representasi numerik menggunakan TF-IDF dan FastText. Pada metode FastText, model dilatih menggunakan korpus lokal dengan dimensi vektor 100, window size 3, min_count 1, dan 200 epoch. Representasi dokumen hotel diperoleh melalui proses *Mean Pooling*, sedangkan tingkat kemiripan antara vektor kueri dan vektor hotel dihitung menggunakan *Cosine Similarity*.

Sebagai contoh, pengujian dilakukan menggunakan kueri informal “penginapan murmer dkt tugu” untuk melihat karakteristik hasil rekomendasi yang dihasilkan oleh kedua metode. Lima rekomendasi teratas beserta nilai *Cosine Similarity* ditampilkan pada Tabel 4.

Tabel 4 Komparasi skor cosine similarity TF-IDF dan fasttext pada kueri informal

Peringkat	Metode TF-IDF (Baseline)	Skor	Metode FastText (Usulan)	Skor
1	The Cabin Tugu Hostel	0.2189	Atrium Premiere Hotel	0.8718
2	KHAS Tugu Hotel	0.1762	The Cabin Tugu Hostel	0.8688
3	Sans Hotel Nagari	0.1445	All Nite & Day Hotel	0.8634
4	RedDoorz near Galeria	0.1400	Hotel Andono near YIA	0.8614
5	Laxston Hotel	0.1398	Sans Hotel Nagari	0.8562

Berdasarkan Tabel 4, kedua metode menghasilkan urutan rekomendasi dan nilai *Cosine Similarity* yang berbeda. Pada contoh kueri informal tersebut, nilai *Cosine Similarity* FastText berada pada rentang 0.8562–0.8718, sedangkan TF-IDF berada pada rentang 0.1398–0.2189. Perbedaan tersebut menunjukkan bahwa kedua metode menghasilkan representasi dan ruang vektor yang berbeda dalam mencocokkan kueri dengan profil hotel.

 ISSN:2302-8149
e-ISSN:2540-9719

Pada TF-IDF, pencocokan sangat dipengaruhi oleh keberadaan dan bobot kata dalam representasi dokumen. Sementara itu, FastText menggunakan informasi subkata sehingga variasi bentuk kata dapat direpresentasikan melalui karakter penyusunnya. Hal tersebut menjadi relevan pada kueri informal yang mengandung bentuk seperti "murmer" dan "dkt". Namun, nilai *Cosine Similarity* yang lebih tinggi tidak secara langsung menunjukkan bahwa suatu rekomendasi lebih relevan secara objektif, karena penelitian ini tidak memiliki *ground truth* relevansi yang diperoleh secara independen.

4.4 Hasil Evaluasi Ketahanan

Evaluasi ketahanan dilakukan untuk mengamati perubahan nilai kemiripan ketika kueri formal diubah menjadi kueri informal dengan intensi pencarian yang tetap sama. Pengujian menggunakan lima pasangan kueri formal dan informal yang masing-masing diuji menggunakan TF-IDF dan FastText. Dari setiap kueri, lima rekomendasi teratas digunakan untuk menghitung nilai Mean Similarity@5. Perubahan nilai antara kueri formal dan informal kemudian digunakan untuk mengamati konsistensi representasi masing-masing metode terhadap variasi bentuk bahasa.

Tabel 5 Perbandingan mean similarity@5 pada kueri formal dan informal

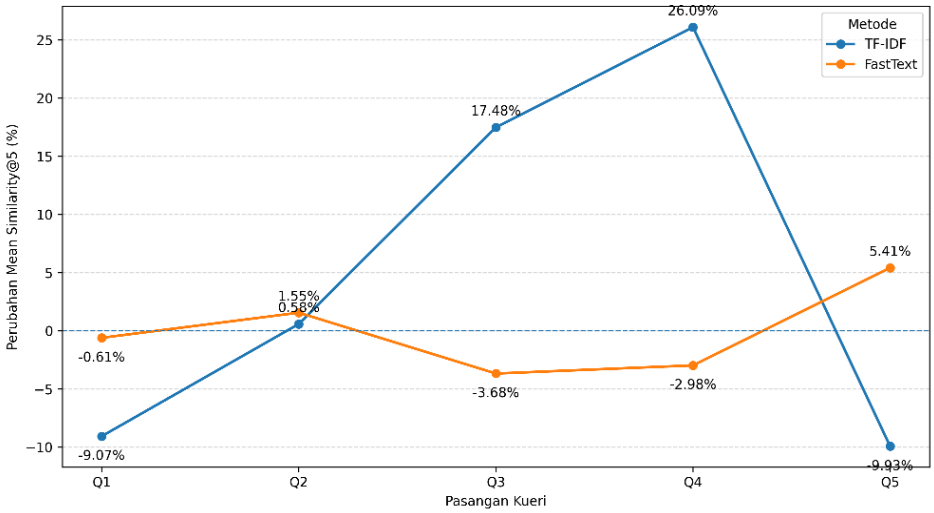
Pasangan	Metode	Kueri formal	Kueri Informal	Mean Sim.@5 Formal	Mean Sim.@5 Informal	Perubahan Relatif (%)
Q1	TF-IDF	hotel murah dekat malioboro	hotel murmer dkt malioboro	0.116673	0.127253	-9.07
Q1	FastText	hotel murah dekat malioboro	hotel murmer dkt malioboro	0.7672	0.7718	-0.61
Q2	TF-IDF	penginapan bersih dengan fasilitas kolam renang	penginapan bersih ada kolam renang	0.2531	0.2516	0.58
Q2	FastText	penginapan bersih dengan fasilitas kolam renang	penginapan bersih ada kolam renang	0.8465	0.8334	1.55
Q3	TF-IDF	hotel dekat malioboro dengan sarapan	hotel deket malioboro ada sarapan	0.1227	0.1012	17.48
Q3	FastText	hotel dekat malioboro dengan sarapan	hotel deket malioboro ada sarapan	0.8298	0.8603	-3.68
Q4	TF-IDF	penginapan nyaman untuk keluarga	penginapan nyaman buat keluarga	0.1683	0.1244	26.09
Q4	FastText	penginapan nyaman untuk keluarga	penginapan nyaman buat keluarga	0.8007	0.8245	-2.98
Q5	TF-IDF	hotel strategis dekat stasiun tugu	hotel strategis dkt stasiun tugu	0.1452	0.1597	-9.93
Q5	FastText	hotel strategis dekat stasiun tugu	hotel strategis dkt stasiun tugu	0.8077	0.7642	5.41

 ISSN:2302-8149
e-ISSN:2540-9719

Berdasarkan Tabel 5, TF-IDF menunjukkan perubahan Mean Similarity@5 yang lebih bervariasi ketika bentuk kueri diubah dari formal menjadi informal. Pada Q3, nilai Mean Similarity@5 menurun dari 0.1227 menjadi 0.1012 atau mengalami perubahan sebesar 17.48%. Penurunan yang lebih besar terjadi pada Q4, yaitu dari 0.1683 menjadi 0.1244 dengan perubahan sebesar 26.09%. Sebaliknya, pada Q1 dan Q5, nilai pada kueri informal meningkat dibandingkan kueri formal, masing-masing sebesar 9.07% dan 9.93%, yang ditunjukkan oleh nilai perubahan negatif.

Pada FastText, perubahan Mean Similarity@5 cenderung lebih kecil pada lima pasangan kueri yang diuji. Perubahan terbesar terjadi pada Q5 sebesar 5.41%. Pada Q1, Q3, dan Q4, nilai Mean Similarity@5 pada kueri informal justru lebih tinggi dibandingkan kueri formal, dengan perubahan masing-masing sebesar -0.61%, -3.68%, dan -2.98%. Sementara itu, Q2 mengalami penurunan sebesar 1.55%.

Perbandingan nilai Mean Similarity@5 antara kueri formal dan informal untuk kedua metode divisualisasikan pada Gambar 2. Visualisasi menunjukkan bahwa perubahan skor TF-IDF lebih beragam, sedangkan perubahan skor FastText cenderung berada pada rentang yang lebih sempit.



Gambar 2 Perubahan mean similarity@5 dari kueri formal ke informal

Gambar 2 menunjukkan perubahan Mean Similarity@5 ketika kueri formal diubah menjadi kueri informal pada lima pasangan kueri. Perubahan pada TF-IDF terlihat lebih berfluktuasi dibandingkan FastText. Pada TF-IDF, perubahan terbesar terjadi pada Q4 sebesar 26,09%, sedangkan pada Q3 sebesar 17,48%. Sebaliknya, perubahan pada FastText berada pada rentang yang lebih sempit, dengan perubahan terbesar sebesar 5,41% pada Q5. Nilai negatif menunjukkan bahwa Mean Similarity@5 pada kueri informal justru lebih tinggi dibandingkan kueri formal, seperti yang terlihat pada Q1, Q3, dan Q4 untuk FastText.

Pola tersebut mengindikasikan bahwa representasi FastText relatif lebih stabil ketika bentuk kueri berubah dari formal menjadi informal pada lima skenario pengujian. Sementara itu, perubahan yang lebih besar pada TF-IDF menunjukkan bahwa representasi berbasis pencocokan leksikal mengalami variasi skor yang lebih tinggi pada beberapa perubahan bentuk kueri. Temuan ini digunakan sebagai indikator ketahanan representasi terhadap variasi formal–informal, bukan sebagai ukuran akurasi atau relevansi objektif, karena penelitian ini tidak memiliki ground truth relevansi yang diperoleh secara independen.

4.5 Analisis Keterbatasan Model

Meskipun hasil evaluasi menunjukkan bahwa FastText relatif lebih stabil terhadap perubahan bentuk kueri formal dan informal pada skenario pengujian, hasil rekomendasi masih memiliki beberapa keterbatasan. Keterbatasan tersebut terutama berkaitan dengan pemenuhan atribut numerik

 ISSN:2302-8149
e-ISSN:2540-9719

dan pemahaman hubungan sintaksis dalam kueri. Keterbatasan pertama terlihat pada kueri "hotel murah dekat malioboro". Hasil rekomendasi FastText pada kueri tersebut ditunjukkan pada Tabel 6.

Tabel 6 Hasil rekomendasi fasttext (kueri: "hotel murah dekat malioboro")

Peringkat	Nama Hotel	Harga (IDR)	Lokasi	Skor Kemiripan
1	INNSIDE by Meliá	872.275	Yogyakarta	0.7790
2	Hotel Andono near YIA	325.954	Yogyakarta	0.7785
3	Cempaka Hotel	446.097	Mergangsan	0.7750
4	House Of Cokro at Malioboro	513.022	Tegalrejo	0.7721
5	Carani Hotel Yogyakarta	914.727	Gondokusuman	0.7714

Pada Tabel 6, beberapa hotel dengan harga relatif tinggi tetap memperoleh nilai *Cosine Similarity* yang tinggi untuk kueri yang mengandung kata "murah". Kondisi ini menunjukkan bahwa kemiripan semantik yang tinggi belum tentu merepresentasikan seluruh batasan yang dimaksudkan pengguna, khususnya ketika kueri mengandung preferensi harga. Dalam pendekatan yang digunakan, batas harga belum dimodelkan sebagai *hard constraint* numerik dalam proses pencocokan, sehingga sistem lebih berfokus pada kemiripan representasi teks.

Keterbatasan kedua berkaitan dengan pemahaman hubungan sintaksis, khususnya pada kalimat negasi. Penggunaan FastText dengan Mean Pooling menghasilkan representasi berdasarkan agregasi vektor kata sehingga hubungan antar-kata dalam struktur kalimat tidak sepenuhnya dipertahankan. Sebagai contoh, kueri "hotel yang tidak berisik" berpotensi memiliki kemiripan dengan dokumen yang memuat kata "berisik", meskipun konteksnya menunjukkan makna yang berbeda. Kondisi tersebut menunjukkan bahwa kemiripan berbasis representasi kata belum selalu mampu menangkap hubungan negasi dalam kalimat secara utuh.

Kedua keterbatasan tersebut menunjukkan bahwa representasi semantik FastText masih dapat dikembangkan melalui kombinasi dengan mekanisme pemrosesan atribut terstruktur dan pemahaman konteks kalimat. Salah satu pendekatan yang dapat dipertimbangkan adalah *hybrid filtering* dengan menerapkan aturan khusus pada atribut harga, sedangkan pemodelan berbasis Transformer seperti IndoBERT dapat menjadi alternatif untuk penelitian lanjutan dalam menangani hubungan antar-kata dan konteks kalimat secara lebih mendalam.

5 Kesimpulan

Penelitian ini mengevaluasi penerapan pemodelan *word embedding* FastText pada sistem rekomendasi hotel berbasis *Content-Based Filtering* dan membandingkannya dengan metode TF-IDF menggunakan dataset 100 hotel di Yogyakarta. Evaluasi dilakukan menggunakan lima pasangan kueri yang dipilih secara purposif untuk merepresentasikan variasi bahasa informal, meliputi slang, singkatan, perubahan ejaan, dan penggunaan bentuk preposisional informal, dengan tetap mempertahankan intensi pencarian. Hasil pengujian menunjukkan bahwa perubahan *Mean Similarity*@5 pada FastText cenderung lebih kecil dibandingkan TF-IDF pada lima pasangan kueri yang diuji. Pada TF-IDF, perubahan terbesar terjadi pada Q4 sebesar 26,09% dan Q3 sebesar 17,48%, sedangkan pada FastText perubahan terbesar sebesar 5,41% terjadi pada Q5. Temuan ini mengindikasikan bahwa FastText relatif lebih stabil terhadap perubahan bentuk bahasa formal–informal dalam skenario pengujian penelitian ini, tetapi tidak dapat digunakan untuk menyimpulkan bahwa FastText memiliki akurasi rekomendasi yang lebih tinggi karena penelitian tidak menggunakan *ground truth* relevansi yang diperoleh secara independen. Selain itu, FastText masih memiliki keterbatasan dalam menangkap atribut numerik dan hubungan sintaksis tertentu. Pada kueri yang mengandung preferensi harga, kemiripan semantik yang tinggi belum tentu memenuhi batas harga yang dimaksudkan pengguna, sedangkan penggunaan *Mean Pooling* dapat menyebabkan hubungan antar-kata, khususnya pada kalimat negasi, tidak sepenuhnya dipertahankan dalam representasi dokumen. Oleh karena itu, penelitian selanjutnya dapat mengombinasikan representasi semantik FastText dengan mekanisme *hybrid filtering* yang menerapkan aturan khusus pada atribut terstruktur,

seperti harga, menggunakan dataset yang lebih besar dan bervariasi, serta membandingkan FastText dengan model berbasis Transformer, seperti IndoBERT, untuk mengevaluasi kemampuan pemahaman konteks kalimat secara lebih mendalam.

Referensi

[1] Badan Pusat Statistik Provinsi D.I. Yogyakarta, *Provinsi Daerah Istimewa Yogyakarta Dalam Angka 2024*, Yogyakarta: BPS DI Yogyakarta, 2024. [Online]. Available: https://yogyakarta.bps.go.id/

[2] S. Kemp, *Digital 2023: Indonesia*, We Are Social & Kepios, 2023. [Online]. Available: https://datareportal.com/reports/digital-2023-indonesia

[3] A. T. Wibowo, I. Indriati, and C. Dewi, "Sistem Rekomendasi Destinasi Wisata menggunakan Metode *Content Based Filtering*," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, Vol. 8, No. 2, pp. 277-284, 2021.

[4] H. S. Kusuma and A. Musdholifah, "*Recommendation System for Thesis Topics using Content-based Filtering*," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, Vol. 15, No. 1, pp. 65-74, 2021.

[5] N. N. K. Sari, Licantik, and M. Zahra, "Pemanfaatan Sistem Rekomendasi menggunakan Content-Based Filtering pada Hotel di Palangka Raya," *Technologia: Jurnal Ilmiah*, Vol. 15, No. 4, pp. 754–756, 2024.

[6] C. A. Melyani, "Hotel Recommendation System with Content-Based Filtering Approach (Case Study: Hotel in Yogyakarta on Nusatrip Website)," *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, Vol. 15, No. 1, pp. 325-336, 2022.

[7] R. Al Rasyid, D. Handayani, and U. Ningsih, "Penerapan Algoritma *TF-IDF* dan *Cosine Similarity* untuk *Query* Pencarian pada Dataset Destinasi Wisata," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, Vol. 8, No. 1, pp. 170-178, 2024.

[8] N. A. Hasanah, N. Suciati, and D. Purwitasari, "Pemantauan Perhatian Publik terhadap Pandemi COVID-19 melalui Klasifikasi Teks dengan *Deep Learning*," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, Vol. 5, No. 1, pp. 193-202, 2021.

[9] N. Sepriadi, E. Budianita, M. Fikry, and Pizaini, "Analisis Sentimen Review Aplikasi Mypertamina menggunakan *Word embedding Fasttext* dan *Algoritma K-Nearest Neighbor*," *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, Vol. 15, No. 1, pp. 91–109, 2023.

[10] K. Obajha, N. N. K. Sari, and V. H. Pranatawijaya, "Implementasi Metode *Collaborative Filtering* pada Aplikasi Rekomendasi Hotel dan Wisma di Kota Palangka Raya berbasis *Website*," *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, Vol. 3, No. 2, pp. 398–410, 2023.

[11] R. Faurina and E. Sitanggang, "Implementasi Metode *Content-based Filtering* dan *Collaborative Filtering* pada Sistem Rekomendasi Wisata di Bali," *Tecno.com*, Vol. 22, No. 4, pp. 870–881, 2023.

[12] K. H. Putra, A. N. Rohman, and Norhikmah, "*Implementation of a Hybrid Filtering Approach in a Website-based Football News Recommendation System*," *Sistemasi: Jurnal Sistem Informasi*, Vol. 13, No. 1, pp. 321-333, 2025.

[13] C. Salsabilla and D. W. Utomo, "Sistem Rekomendasi Wisata Kabupaten Pekalongan dengan *Content Based Filtering*," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, No. 1, pp. 262–270, 2025.

[14] A. Maulana, A. N. Rohman, and Y. Pristyanto, "Sistem Rekomendasi Destinasi Tempat Wisata Provinsi Aceh menggunakan Metode *Content based Filtering*," *Sistemasi: Jurnal Sistem Informasi*, Vol. 15, No. 2, pp. 418–432, 2026.

[15] R. R. Situngkir and R. Sianipar, "Penerapan Metode *Cosine Similarity* dan Pembobotan TF-IDF Untuk Mendeteksi Kemiripan Dokumen," *Jurnal Teknik Informatika (JUTIF)*, Vol. 3, No. 4, pp. 1023-1030, 2022.

[16] Y. Gunawan, J. C. Young, and A. Rusli, "FastText *Word Embedding* and *Random Forest Classifier* for User Feedback Sentiment Classification in Bahasa Indonesia," *Ultimatics: Jurnal Teknik Informatika*, Vol. 13, No. 2, pp. 101–107, 2022.

ISSN:2302-8149
e-ISSN:2540-9719

[17] A. Putri, S. Agustian, Jasril, and I. Afrianty, "Eksplorasi Fitur *FastText*, TF-IDF dan IndoBERT pada Metode K-Nearest Neighbor untuk Klasifikasi Sentimen," *ZONAsi: Jurnal Sistem Informasi*, Vol. 7, No. 1, pp. 49–60, 2025.

[18] A. Naldi and S. Agustian, "Klasifikasi Sentimen Vaksin Covid-19 menggunakan K-Nearest Neighbor berdasarkan *Word embeddings* Fasttext pada Twitter," *ZONAsi: Jurnal Sistem Informasi*, vol. 5, no. 2, pp. 323–333, 2023.