

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan platform ulasan daring seperti *Yelp* telah berperan signifikan dalam membentuk keputusan konsumen terhadap restoran dan café[1]. Ulasan berbasis teks memuat informasi emosional dan pengalaman pengguna yang tidak tercermin melalui *rating* numerik semata[2]. Model berbasis *transformer* seperti *Bidirectional Encoder Representations from Transformers (BERT)* memiliki kemampuan lebih unggul dalam memahami konteks linguistik dan polaritas sentimen dibandingkan model *Long Short Term Memory (LSTM)* tradisional[3]. Sadikin menambahkan bahwa kualitas pra-pemrosesan dan pemilihan sampel data sangat memengaruhi performa model analisis sentimen pada dataset *Yelp*[4]. Hal ini menegaskan bahwa ulasan teks merupakan sumber informasi yang kompleks dan membutuhkan pendekatan analisis yang lebih canggih untuk memahami preferensi emosional pengguna.

Untuk menjawab kompleksitas tersebut, berbagai pendekatan analisis telah dikembangkan. Pendekatan tradisional seperti *Support Vector Machine (SVM)* masih banyak digunakan dalam analisis sentimen berbasis aspek, namun terbatas dalam menangkap konteks emosional yang kompleks[5]. Sebagai pengembangan, Kaur memperkenalkan metode *hybrid feature extraction* yang terbukti meningkatkan stabilitas dan akurasi hasil analisis[6]. Selain itu, *ensemble learning* terbukti efektif dalam menjaga konsistensi analisis opini pelanggan pada ulasan restoran[7]. Lebih lanjut, *pipeline* memiliki peran penting dalam menjaga struktur dan keandalan proses analisis berbasis pembelajaran mesin[8]. Dengan demikian, model pembelajaran dan teknik ekstraksi fitur modern memberikan landasan kuat bagi peningkatan kualitas analisis ulasan pengguna.

Hasil analisis sentimen yang akurat dapat dimanfaatkan lebih lanjut dalam sistem rekomendasi untuk menghasilkan rekomendasi yang lebih personal dan kontekstual[9]. Representasi semantic berbasis *word embeddings* terbukti mampu

memperdalam pemahaman hubungan antar item dalam sistem rekomendasi[10]. Selain itu, pendekatan *content-based filtering* efektif dalam mencocokkan profile pengguna dengan karakteristik item[11]. Sistem rekomendasi yang *context-aware* juga memiliki peran penting dalam menyesuaikan hasil rekomendasi dengan preferensi pengguna[12]. Oleh karena itu, kombinasi antara analisis sentimen dan model rekomendasi berbasis konten menjadi arah penelitian yang relevan untuk meningkatkan relevansi dan kepuasan pengguna.

Dalam sistem rekomendasi, pengukuran kesamaan antar item merupakan elemen krusial dalam menentukan tingkat relevansi rekomendasi. Salah satu metode yang umum digunakan untuk tujuan tersebut adalah *cosine similarity*, karena kemampuannya menghitung kedekatan antar fitur dengan efisiensi dan akurasi tinggi[13]. Li dan Zitong memperkuat temuan ini dengan menunjukkan efektivitas *cosine similarity* dalam pipeline *embedding-similarity-ranking* pada berbagai domain seperti pendidikan dan sistem rekomendasi ilmiah[14], [15]. Meskipun penggunaan *BERT* dan *cosine similarity* telah diterapkan di berbagai domain[16], [17], integrasi keduanya dalam sistem rekomendasi café berbasis ulasan teks masih jarang dilakukan. Berdasarkan hal tersebut, penelitian ini bertujuan mengembangkan sistem rekomendasi yang mampu menghasilkan saran café yang kontekstual dan adaptif terhadap variasi pengguna melalui pemahaman makna serta sentimen dalam ulasan.

1.2 Rumusan Masalah

Seberapa baik kinerja model *BERT* dalam mengekstraksi fitur sentimen dan suasana (*mood*) dari ulasan teks, serta penerapan *cosine similarity* pada sistem rekomendasi café berbasis *content-based filtering* dalam menghasilkan rekomendasi yang relevan dan kontekstual berdasarkan metrik evaluasi?

1.3 Batasan Masalah

1. Penelitian ini menggunakan 38.000 data ulasan dari *Yelp Academic Dataset* karena keterbatasan komputasi yang tidak memungkinkan pemrosesan seluruh dataset.

2. Data yang digunakan hanya berasal dari dataset *review* dan *business*, sehingga analisis terbatas pada teks ulasan, *rating* bintang (*stars*), nama *café*, dan kategori bisnis tanpa mencakup informasi seperti lokasi, harga, menu, fasilitas, maupun demografi pengguna.
3. Analisis sentimen dilakukan dengan menggunakan model *BERT* (*bert-base uncased*) yang dikembangkan untuk mengklasifikasikan sentimen ulasan menjadi kategori positif, netral dan negatif sebagai indikator kualitas *café*.
4. Sistem rekomendasi dibangun menggunakan pendekatan *content-based filtering* yang memanfaatkan *embedding* teks, *sentimen score* dan perhitungan *cosine similarity* untuk menentukan tingkat kemiripan antar *café*.
5. Lingkup penelitian difokuskan pada rekomendasi *café* berdasarkan analisis teks ulasan pengguna, yang mencakup makna semantic dan kualitas sentimen, tanpa menggunakan fitur non teks.

1.4 Tujuan Penelitian

Menganalisis efektivitas kombinasi model *BERT* dalam mengekstraksi fitur sentimen dan suasana (*mood*) dari ulasan *café*, serta mengevaluasi kinerja *content-based filtering* berbasis *cosine similarity* dalam menghasilkan rekomendasi *café* yang relevan dan kontekstual.

1.5 Manfaat Penelitian

1. Manfaat Teoritis
 Penelitian ini memberikan kontribusi pada pengembangan kajian di bidang *Natural Language Preprocessing (NLP)* dan sistem rekomendasi dengan menunjukkan penerapan metode *BERT* dalam analisis ulasan teks berbasis suasana (*mood*). Hasil penelitian ini juga dapat menjadi rujukan bagi studi lanjutan terkait integrasi model *transformer* dengan metode rekomendasi berbasis konten.
2. Manfaat Praktis
 Secara praktis, penelitian ini bermanfaat bagi konsumen dalam memperoleh rekomendasi *café* yang lebih relevan sesuai preferensi emosional (*mood*),

bagi pemilik cafe sebagai masukan untuk memahami persepsi pelanggan secara lebih mendalam guna meningkatkan kualitas layanan, serta bagi pengembang sistem sebagai acuan dalam membangun aplikasi rekomendasi yang memanfaatkan *machine learning* modern untuk menghasilkan layanan yang lebih personal dan kontekstual.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini berisi penjelasan mengenai latar belakang perkembangan sistem rekomendasi dan keterbatasan metode tradisional yang hanya mengandalkan *rating* numerik. Penjelasan mengenai perumusan masalah yang menjadi fokus penelitian, batasan ruang lingkup kerja, tujuan penelitian, serta manfaat penelitian secara teoritis maupun praktis. Selain itu, berisi penjelasan sistematika penulisan skripsi secara keseluruhan sebagai panduan bagi pembaca.

BAB II TINJAUAN PUSTAKA

Bab ini membahas teori yang mendasari penelitian, meliputi konsep sistem rekomendasi (*content-based filtering*, *collaborative filtering*, dan *hybrid methods*), *Natural Language Processing (NLP)*, serta model *BERT* sebagai metode ekstraksi fitur teks. Selain itu, dijelaskan pula metode pengukuran kesesuaian menggunakan *cosine similarity* dan evaluasi performa dengan metrik *Average Similarity* serta *nDCG*. Bab ini juga memuat kajian penelitian terdahulu yang relevan dengan sistem rekomendasi berbasis ulasan teks.

BAB III METODE PENELITIAN

Bab ini menjelaskan tahapan penelitian yang meliputi pengumpulan dataset ulasan cafe dari *Yelp*, *preprocessing* (pembersihan teks, penanganan data hilang, *outlier*, dan *stratified sampling*), serta *EDA*. Tahapan dilanjutkan dengan ekstraksi fitur menggunakan *BERT*, pembentukan profile sentimen cafe, dan penerapan *content-based filtering* berbasis *cosine similarity*.

Evaluasi sistem dilakukan menggunakan metrik *Average Similarity* dan *nDCG*. Diagram alur penelitian disertakan untuk menggambarkan proses penelitian secara menyeluruh.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi metode yang digunakan, mencakup hasil *preprocessing*, ekstraksi fitur dengan *BERT*, dan rekomendasi cafe berdasarkan preferensi pengguna. Disertakan pula hasil evaluasi sistem menggunakan metrik *Average Similarity* dan *nDCG*. Pembahasan berfokus pada analisis efektivitas model dalam menangkap konteks emosional ulasan serta relevansinya terhadap peningkatan kualitas rekomendasi di domain cafe.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian yang dilakukan berdasarkan hasil pengujian dan capaian tujuan penelitian. Selain itu, disampaikan pula saran untuk penelitian selanjutnya, seperti penggunaan dataset yang lebih luas, penggabungan pendekatan *hybrid filtering*, serta pengembangan sistem rekomendasi berbentuk aplikasi interaktif agar lebih mudah dimanfaatkan oleh pengguna dan pelaku bisnis.