

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan bisnis kuliner, khususnya usaha kedai kopi, menunjukkan pertumbuhan yang pesat seiring dengan perubahan gaya hidup masyarakat modern (Judawinata 2023). Konsumsi kopi tidak lagi hanya dipandang sebagai aktivitas pemenuhan kebutuhan minuman, tetapi telah berkembang menjadi bagian dari aktivitas sosial dan budaya masyarakat perkotaan, di mana *coffee shop* berfungsi sebagai ruang interaksi, tempat berkumpul, berdiskusi, hingga menjalankan berbagai kegiatan sehari-hari terutama di kalangan generasi muda (Herlyana 2012). Kehadiran *coffeeshop* dengan berbagai konsep, suasana, dan fasilitas menjadikannya tidak hanya sebagai tempat menikmati kopi, tetapi juga sebagai ruang yang mendukung beragam aktivitas seperti bekerja, belajar, maupun bersantai, sehingga *coffee shop* menjadi bagian dari pola aktivitas masyarakat modern (G. J. E. Putri 2013).

Di sisi lain, pertumbuhan usaha kedai kopi di berbagai daerah, termasuk Kota Malang, menunjukkan adanya inovasi berkelanjutan dalam konsep, pelayanan, dan strategi pemasaran untuk menarik minat konsumen, yang berdampak pada semakin banyaknya pilihan *coffeeshop* dengan karakteristik berbeda (JALAALI 2018). Banyaknya pilihan *coffeeshop* dengan karakteristik yang beragam sering membuat pengguna kesulitan menentukan tempat yang sesuai dengan preferensi dan kebutuhannya. Informasi mengenai karakteristik tersebut sebenarnya telah tersedia dalam bentuk ulasan pengguna, namun jumlah data yang besar serta sifatnya yang tidak terstruktur menyebabkan informasi tersebut sulit dimanfaatkan secara langsung. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengekstraksi dan merepresentasikan karakteristik *coffeeshop* secara otomatis berdasarkan pola pengalaman pengguna, sehingga sistem dapat menghasilkan rekomendasi yang lebih relevan.

Sejalan dengan permasalahan tersebut, diperlukan pendekatan yang mampu mengidentifikasi pola dan tema yang tersembunyi di dalam teks secara otomatis. Salah satu pendekatan dalam text mining yang mampu mengidentifikasi pola tersembunyi dari kumpulan dokumen adalah topic modeling. *Latent Dirichlet Allocation* (LDA) merupakan metode probabilistik yang digunakan untuk menemukan topik laten berdasarkan distribusi kemunculan kata dalam dokumen (Maulidiya 2023). Berbeda dengan pendekatan klasifikasi yang memerlukan pelabelan data, LDA mampu mengekstraksi struktur topik secara otomatis dari teks yang tidak terstruktur (Sahria and Fudholi 2020).

Selain menggunakan representasi kata tunggal (unigram), penelitian ini juga memanfaatkan pendekatan N-Gram, khususnya bigram, untuk mempertahankan hubungan konteks antar kata yang muncul secara berurutan pada ulasan pengguna sehingga frasa yang memiliki makna tertentu dapat direpresentasikan dengan lebih baik (Kelvin et al. 2025). Penggunaan bigram dinilai mampu memberikan representasi teks yang lebih informatif karena beberapa frasa tertentu memiliki makna yang lebih spesifik dibandingkan kata tunggal, seperti “tempat nyaman”, “kopi enak”, atau “harga murah”. Representasi berbasis bigram membantu sistem mempertahankan konteks lokal antar kata sehingga hubungan makna pada ulasan pengguna dapat direpresentasikan secara lebih baik melalui pembentukan frasa yang lebih bermakna (Aqshal et al. 2026). Dalam proses topic modeling, penggunaan bigram dapat membantu model mengenali keterkaitan istilah yang merepresentasikan pengalaman pengguna secara lebih kontekstual sehingga topik yang dihasilkan menjadi lebih interpretatif dan relevan terhadap karakteristik coffeeshop. Dengan demikian, penggunaan bigram pada penelitian ini berperan dalam meningkatkan kualitas representasi teks sebelum dilakukan proses pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA).

Penelitian sistem rekomendasi di Indonesia pada umumnya masih menggunakan pendekatan berbasis kemiripan kata atau atribut eksplisit, misalnya melalui content-based filtering yang mengukur kesamaan deskripsi teks maupun metode pengambilan keputusan multikriteria berbasis atribut rating atau rasio kuantitatif (Muhammad Ridho Fasya, Muhamad Alda, and Adnan Buyung Nasution 2025) (Budi Suryono, Sumarlinda, and Oktaviani 2025). Pendekatan tersebut umumnya merepresentasikan data berdasarkan kemunculan kata atau atribut eksplisit, sehingga hubungan yang ditangkap masih bersifat permukaan dan belum sepenuhnya merepresentasikan konteks makna yang terkandung dalam ulasan pengguna. Akibatnya, sistem rekomendasi cenderung bergantung pada kesamaan istilah tanpa mempertimbangkan pola pengalaman yang lebih kompleks. Oleh karena itu, diperlukan pendekatan yang mampu membangun representasi semantik coffeeshop berdasarkan pola pengalaman pengguna yang muncul secara alami dalam data.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan sistem rekomendasi coffeeshop berbasis analisis topik ulasan pengguna menggunakan *Latent Dirichlet Allocation* (LDA). Hasil pemodelan topik digunakan untuk membentuk profil karakter setiap *coffeeshop*, yang selanjutnya dimanfaatkan dalam proses pemberian rekomendasi berdasarkan kesamaan karakteristik.

1.2 Rumusan Masalah

Untuk mengidentifikasi masalah tersebut, sebuah rumusan masalah dibuat berdasarkan latar belakang di atas:

1. Bagaimana membangun representasi karakter coffeeshop berbasis distribusi topik laten yang diekstraksi dari ulasan pengguna menggunakan metode Latent Dirichlet Allocation (LDA)?
2. Sejauh mana distribusi topik yang dihasilkan mampu merepresentasikan karakteristik coffeeshop secara semantik?

3. Bagaimana memanfaatkan representasi distribusi topik tersebut untuk menghasilkan rekomendasi coffeeshop berbasis kesamaan makna terhadap kebutuhan pengguna dalam bentuk teks bebas?
4. Bagaimana mengevaluasi kualitas topik dan relevansi rekomendasi yang dihasilkan oleh sistem?

1.3 Batasan Masalah

Batasan masalah yang diperlukan untuk menyelesaikan penelitian ini adalah sebagai berikut:

1. Data yang digunakan berupa ulasan teks *coffeeshop* yang diperoleh dari sumber daring (Google Maps Reviews), dan tidak mencakup data non-teks seperti gambar atau audio.
2. Analisis dilakukan menggunakan metode *topic modeling* LDA tanpa proses pelabelan manual.
3. Parameter pencarian yang dimasukkan pengguna ke dalam sistem dibatasi hanya menggunakan query teks bebas.
4. Pemeringkatan rekomendasi murni ditentukan berdasarkan perhitungan kemiripan semantik (Cosine Similarity) antara query pengguna dengan distribusi topik tiap coffeeshop hasil LDA.
5. Ruang lingkup dataset dan objek penelitian dibatasi secara spesifik pada 30 kedai kopi (*coffeeshop*) yang berlokasi di wilayah Kota Malang.
6. Penelitian ini hanya mengambil ulasan dengan data minimal rating = 3
7. Sistem yang dibangun berupa prototipe berbasis web sebagai media pengujian model, dan tidak membahas aspek implementasi skala industri.

1.4 Tujuan Penelitian

Berikut adalah tujuan dari penelitian ini:

1. Mengidentifikasi topik laten pada ulasan pengguna coffeeshop menggunakan metode *Latent Dirichlet Allocation* (LDA).
2. Membentuk profil karakter setiap *coffeeshop* berdasarkan distribusi topik yang dihasilkan.

3. Mengembangkan sistem rekomendasi coffeeshop berbasis kesamaan karakteristik tempat menggunakan pendekatan LDA yang dibantu dengan boost menggunakan TF-IDF.
4. Mengevaluasi kemampuan sistem dalam menghasilkan rekomendasi yang relevan berdasarkan representasi semantik ulasan pengguna.

1.5 Manfaat

1. Manfaat bagi Pengguna (User)

Penelitian ini diharapkan dapat membantu pengguna dalam memperoleh rekomendasi *coffeeshop* yang sesuai dengan preferensi berdasarkan analisis ulasan pengguna.

2. Manfaat Bagi Peneliti

Penelitian ini dapat menambah pemahaman dan pengalaman dalam menerapkan metode text mining, khususnya pemodelan topik menggunakan Latent Dirichlet Allocation (LDA).

3. Manfaat Bagi Penelitian Selanjutnya

Penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan analisis teks, topic modeling, maupun pengembangan sistem rekomendasi.

1.6 Sistematika Penulisan

Agar mempermudah pemahaman pada pembahasan penulisan skripsi ini, maka sistematika penulisan diperoleh sebagai berikut:

BAB I Pendahuluan : Bab ini berisi latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan yang digunakan sebagai gambaran umum penelitian.

BAB II Tinjauan Pustaka : Bab ini membahas teori-teori yang mendukung penelitian, meliputi konsep sistem rekomendasi, *text mining*, *Natural Language Processing* (NLP), tahapan *preprocessing* teks, representasi dokumen menggunakan *Bag-of-Words*, konsep *topic modeling*, metode *Latent Dirichlet Allocation* (LDA), pengukuran kemiripan dokumen menggunakan *cosine similarity* serta penelitian terdahulu yang relevan dengan analisis topik dan sistem rekomendasi berbasis teks

BAB III Analisis dan Perancangan : Bab ini menjelaskan analisis kebutuhan sistem, perancangan arsitektur sistem, alur proses, serta desain sistem rekomendasi yang akan dibangun berdasarkan metode yang digunakan dalam penelitian.

BAB IV Implementasi dan Pengujian : Bab ini menjelaskan tahapan implementasi dari rancangan sistem rekomendasi ke dalam bentuk perangkat lunak (antarmuka pengguna dan *backend*). Selain itu, bab ini juga memaparkan skenario pengujian fungsionalitas, evaluasi algoritma LDA menggunakan

perhitungan *Coherence Score*, serta pengujian tingkat akurasi hasil rekomendasi berbasis perhitungan *Cosine Similarity*.

BAB V Penutup : Bab ini menyajikan kesimpulan akhir berdasarkan hasil analisis dan pengujian yang telah dilakukan selama penelitian.