

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi dalam satu dekade terakhir telah mendorong transformasi signifikan pada model bisnis digital, khususnya pada sektor perdagangan elektronik (*e-commerce*). Platform *e-commerce* modern tidak hanya berfungsi sebagai media transaksi, tetapi juga sebagai ekosistem data yang menghimpun interaksi pengguna dalam skala besar, meliputi riwayat penelusuran, preferensi produk, ulasan, serta aktivitas pembelian. Kondisi ini memicu pertumbuhan volume informasi yang masif dan berpotensi menimbulkan fenomena *information overload*, di mana pengguna mengalami kesulitan dalam menemukan produk yang relevan secara cepat dan akurat.

Dalam konteks tersebut, sistem rekomendasi (*recommender system*) berperan strategis sebagai komponen inti yang mampu menyaring informasi, mempersonalisasi konten, serta meningkatkan kualitas pengalaman pengguna. Penelitian terdahulu menunjukkan bahwa penerapan sistem rekomendasi pada platform *e-commerce* berkontribusi langsung terhadap peningkatan kepuasan pelanggan dan efektivitas pemasaran digital [1]. Metode konvensional yang umum digunakan, seperti *Collaborative Filtering* dan *Content-Based Filtering*, telah terbukti mampu memberikan rekomendasi dasar, namun masih memiliki keterbatasan signifikan dalam menangkap makna kontekstual dari data teks pengguna yang bersifat heterogen dan tidak terstruktur.

Sebagaimana diuraikan oleh Rachmaniar et al. [1], pendekatan berbasis kesamaan historis dan pencocokan kata kunci cenderung bergantung pada frekuensi kemunculan fitur eksplisit, sehingga kurang efektif dalam merepresentasikan preferensi laten pengguna. Selain itu, permasalahan *cold-start* baik pada pengguna baru maupun produk baru masih menjadi tantangan utama yang berdampak pada rendahnya relevansi rekomendasi pada fase awal interaksi sistem.

Seiring dengan pesatnya perkembangan kecerdasan buatan, khususnya *Large Language Models (LLM)*, pendekatan baru dalam sistem rekomendasi mulai diarahkan pada pemanfaatan teknik *Vector Embedding*. Model *LLM* modern, seperti *Google Gemini*, memiliki kemampuan untuk mentransformasi teks naratif menjadi representasi vektor berdimensi tinggi yang mampu menangkap hubungan semantik antar konsep secara lebih mendalam. Sejumlah penelitian terkini telah mengkaji pemanfaatan *API Gemini* dalam konteks *e-commerce*.

Kiareni et al. [2] menekankan penggunaan *Gemini* untuk otomatisasi deskripsi produk pada aplikasi *e-commerce* berbasis *mobile*. Penelitian lain oleh Nazarius et al. [3] menunjukkan bahwa *Gemini AI* mampu meningkatkan kualitas representasi bahasa alami dalam pembuatan konten produk. Selanjutnya, Handoko et al. [4] mengonfirmasi efektivitas *Gemini* dalam pengembangan aplikasi *e-commerce mobile* berbasis generasi teks. Meskipun demikian, keseluruhan penelitian tersebut masih berfokus pada pengayaan konten produk, dan belum secara eksplisit mengkaji representasi profil pelanggan sebagai dasar sistem rekomendasi.

Penggunaan model *embedding Gemini*, khususnya *models/text-embedding-004*, membuka peluang untuk merepresentasikan preferensi pengguna tidak lagi sebatas kecocokan kata, melainkan berdasarkan kedekatan makna semantik. Namun demikian, representasi vektor semantik yang dihasilkan memerlukan pengelolaan lanjutan agar dapat dimanfaatkan secara optimal dalam sistem rekomendasi. Salah satu pendekatan yang relevan adalah segmentasi pelanggan melalui teknik *clustering*, yang bertujuan mengelompokkan pengguna berdasarkan kemiripan preferensi.

Algoritma *K-Means Clustering* merupakan salah satu metode yang banyak digunakan karena kesederhanaan dan efisiensinya dalam mengelola data berdimensi tinggi. Studi Noviandy et al. [5] menunjukkan bahwa *K-Means* efektif digunakan untuk analisis pengelompokan berbasis data kuantitatif, namun penentuan jumlah klaster yang optimal tetap menjadi isu krusial. Oleh karena itu, diperlukan pendekatan kuantitatif yang terukur, seperti *Elbow Method* untuk

menentukan nilai K yang optimal, serta *Silhouette Score* sebagai metrik evaluasi kualitas klusterisasi.

Berdasarkan uraian tersebut, dapat diidentifikasi adanya celah penelitian pada integrasi antara model *embedding LLM*, khususnya *Google Gemini*, dengan pendekatan *clustering* yang tervalidasi secara kuantitatif dalam sistem rekomendasi *e-commerce*. Penelitian ini berupaya mengisi celah tersebut dengan menyusun analisis kuantitatif sistem rekomendasi produk berbasis model *embedding Gemini* yang teruji secara statistik. Oleh karena itu, penelitian ini diberi judul “Analisis Kuantitatif Sistem Rekomendasi Produk pada Platform E-commerce Berbasis Model Embedding Gemini”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

- 1) Bagaimana efektivitas model *embedding Gemini (models/text-embedding-004)* dalam merepresentasikan profil pelanggan yang heterogen ke dalam vektor semantik dibandingkan dengan pendekatan ekstraksi fitur tradisional?
- 2) Berapa jumlah segmen pelanggan (*cluster*) yang paling optimal berdasarkan analisis *Elbow Method* dalam memetakan variasi minat pengguna pada platform *e-commerce*?
- 3) Bagaimana kinerja hasil pengelompokan pelanggan dan validitas sistem rekomendasi yang dibangun berdasarkan pengukuran kuantitatif menggunakan metrik *Silhouette Score*?

1.3 Batasan Masalah

Untuk menjaga fokus penelitian dan mencegah perluasan pembahasan yang tidak relevan, batasan masalah dalam penelitian ini ditetapkan sebagai berikut:

- 1) Data yang digunakan berupa data simulasi pelanggan *e-commerce* yang bersumber dari berkas *customer_data_collection.csv* yang diperoleh dari platform *Kaggle*, mencakup riwayat penelusuran, riwayat pembelian, dan data demografis dalam bentuk teks tidak terstruktur.

- 2) Proses representasi vektor teks dibatasi pada penggunaan *Google Gemini API* dengan spesifikasi model *models/text-embedding-004*.
- 3) Pengukuran tingkat kemiripan antar profil pengguna dilakukan menggunakan algoritma *Cosine Similarity*.
- 4) Segmentasi pelanggan dilakukan menggunakan algoritma *K-Means Clustering*, dengan penentuan jumlah kluster optimal divalidasi melalui *Elbow Method*.
- 5) Penelitian difokuskan pada analisis kuantitatif kinerja sistem rekomendasi pada sisi *backend* dan tidak mencakup perancangan antarmuka pengguna.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah:

- 1) Mengimplementasikan model *embedding Gemini (models/text-embedding-004)* untuk merepresentasikan profil pelanggan *e-commerce* ke dalam vektor semantik.
- 2) Menentukan jumlah kluster pelanggan yang optimal menggunakan pendekatan *Elbow Method*.
- 3) Mengevaluasi kualitas pengelompokan pelanggan dan validitas sistem rekomendasi secara kuantitatif menggunakan *Silhouette Score*.

1.5 Manfaat Penelitian

- 1) Manfaat Teoritis.

Penelitian ini diharapkan dapat memberikan kontribusi akademik dalam pengembangan kajian sistem rekomendasi berbasis *Large Language Models*, khususnya terkait pemanfaatan *vector embedding* dan validasi klusterisasi pada data tekstual.

- 2) Manfaat Praktis.

Hasil penelitian ini dapat menjadi referensi teknis bagi pengembang *platform e-commerce* dalam memanfaatkan data profil pelanggan yang tidak terstruktur guna meningkatkan relevansi rekomendasi produk dan strategi segmentasi pasar.

1.6 Sistematika Penulisan

Penulisan tesis ini disusun dalam lima bab, yaitu:

BAB I PENDAHULUAN, yang memuat Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, dan Sistematika Penulisan.

BAB II TINJAUAN PUSTAKA, yang membahas Penelitian Terdahulu dan Landasan Teori terkait Sistem Rekomendasi, *Vector Embedding*, *Google Gemini*, *K-Means Clustering*, *Elbow Method*, dan *Silhouette Score*.

BAB III METODE PENELITIAN, yang menjelaskan Desain Penelitian, Sumber Data, Tahapan *Preprocessing*, Proses *Embedding*, Pembentukan Matriks Kemiripan, serta Prosedur *Clustering*.

BAB IV HASIL DAN PEMBAHASAN, yang menyajikan Hasil Eksperimen, Analisis Kluster Optimal, Visualisasi Data, serta Evaluasi Kinerja Sistem.

BAB V PENUTUP, yang berisi Kesimpulan dan Rekomendasi untuk penelitian selanjutnya.