

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan transformasi digital telah mendorong peningkatan signifikan terhadap ketersediaan data teks tidak terstruktur, khususnya dalam bentuk *user-generated content* seperti ulasan pengguna (*user reviews*). Ulasan tersebut merepresentasikan opini, pengalaman, dan persepsi konsumen terhadap suatu produk atau layanan, sehingga menjadi sumber informasi strategis dalam analisis perilaku konsumen dan pengembangan sistem cerdas berbasis data. Namun, volume data yang besar, keragaman bahasa, serta kompleksitas struktur kalimat menimbulkan tantangan dalam proses ekstraksi informasi apabila hanya menggunakan pendekatan konvensional. Oleh karena itu, dibutuhkan metode analisis teks yang mampu memahami konteks semantik secara mendalam dan menghasilkan representasi topik yang koheren serta interpretatif.

Dalam bidang *Natural Language Processing* (NLP), *topic modeling* merupakan teknik dalam *text mining* yang digunakan untuk mengidentifikasi pola dan tema tersembunyi dalam sekumpulan dokumen teks. Salah satu metode klasik yang paling banyak digunakan adalah *Latent Dirichlet Allocation* (LDA). Metode ini bekerja dengan pendekatan probabilistik untuk menentukan distribusi topik berdasarkan kemunculan kata dalam dokumen. Namun, LDA memiliki keterbatasan karena hanya mengandalkan frekuensi dan *co-occurrence* kata tanpa memahami makna kontekstualnya, sehingga kurang mampu menangkap hubungan semantik, sinonim, maupun variasi bahasa yang berbeda tetapi memiliki arti serupa.

Seiring berkembangnya model *transformer* seperti BERT, muncul pendekatan *topic modeling* berbasis *contextual embedding* yang mampu memahami konteks kalimat secara lebih mendalam. Salah satu metode yang banyak digunakan adalah BERTopic, yaitu metode pemodelan topik yang memanfaatkan representasi *embedding* dari model *transformer*, kemudian melakukan reduksi dimensi dan pengelompokan dokumen menggunakan algoritma *clustering* berbasis densitas. Representasi topik selanjutnya dibentuk

menggunakan *class-based TF-IDF* (c-TF-IDF) sehingga menghasilkan topik yang lebih koheren dan informatif (Hill et al., 2025).

Dalam penelitian yang dilakukan oleh Alessandro Hill, Kalen Goo, dan Puneet Agarwal, BERTopic digunakan untuk mengekstraksi minat mahasiswa melalui pendekatan *interest mining* dalam sistem rekomendasi program akademik. Hasil penelitian tersebut menunjukkan bahwa BERTopic mampu menangkap hubungan semantik antar dokumen secara lebih efektif dibandingkan metode pemodelan topik tradisional, sehingga menghasilkan topik yang lebih stabil dan relevan (Hill et al., 2025).

Meskipun pendekatan seperti *Explicit Factor Model* (EFM) telah memberikan peningkatan dalam aspek *Explainability* pada sistem rekomendasi, masih terdapat beberapa keterbatasan. Penelitian oleh Shunsuke Kido menunjukkan bahwa sebagian besar sistem *explainable recommendation* hanya memanfaatkan ulasan yang ditulis langsung oleh pengguna. Dalam praktiknya, banyak pengguna hanya memberikan sedikit ulasan sehingga informasi preferensi yang tersedia menjadi terbatas dan menimbulkan *sparsity problem*. Selain itu, kualitas ulasan yang tidak merata juga dapat memengaruhi ketepatan proses ekstraksi fitur (Kido et al., 2021).

Dalam konteks aspect-based recommendation, integrasi metode *topic modeling* berbasis *embedding* dengan model eksplisit seperti EFM berpotensi menghasilkan representasi aspek yang lebih koheren dan stabil. Ekstraksi aspek tidak lagi sepenuhnya bergantung pada aturan phrase-level manual, melainkan dapat diperoleh melalui pemodelan topik yang kontekstual dan adaptif terhadap domain data. Hal ini membuka peluang untuk meningkatkan kualitas representasi preferensi pengguna sekaligus mempertahankan interpretabilitas sistem rekomendasi.

Di sisi lain, perkembangan *Sistem Pendukung Keputusan* (SPK) menuntut integrasi data tidak terstruktur sebagai dasar pengambilan keputusan yang lebih komprehensif. SPK modern tidak lagi terbatas pada pengolahan data numerik terstruktur, tetapi juga memanfaatkan opini pengguna sebagai sumber pengetahuan eksplisit. Ulasan produk batik cenderung menggunakan bahasa yang beragam dan tidak seragam. Satu konsumen bisa menulis "kain adem", sementara

konsumen lain menulis "nyaman dipakai" atau "tidak panas" padahal ketiganya merujuk pada hal yang sama, yaitu kenyamanan. Keragaman ini membuat ulasan batik sulit dianalisis menggunakan metode konvensional yang hanya mengandalkan frekuensi kata. Berdasarkan analisis terhadap 827 ulasan berteks dalam dataset penelitian ini, ditemukan bahwa topik yang paling sering dibahas konsumen adalah kualitas bahan dan kain (muncul di 70,4% ulasan), desain dan motif batik (69,6%), serta kenyamanan pemakaian (40,7%). Temuan ini menunjukkan bahwa konsumen batik lebih banyak membicarakan aspek estetika dan material produk.

Berdasarkan penelitian oleh Dhiraj Murthy dkk. (2024), pendekatan BERTopic berbasis *contextual embedding* terbukti mampu menghasilkan kluster topik yang stabil dan tidak saling tumpang tindih pada data teks berskala besar. Kemampuan ini relevan untuk domain batik karena memungkinkan ekstraksi aspek produk secara semantik dan kontekstual, sehingga hasil topik dapat dimanfaatkan secara langsung sebagai fitur eksplisit dalam *Explicit Factor Model* untuk membangun sistem *aspect-based recommendation*, yang lebih akurat dan *explainable* (Murthy et al., 2024).

Selain itu, penerapan BERTopic dalam *explainable recommender system* terbukti mampu meningkatkan transparansi sistem dengan memanfaatkan representasi topik sebagai alasan rekomendasi kepada pengguna (Ante Panjkota et al., 2025).

Dalam ranah sistem rekomendasi, kebutuhan terhadap transparansi dan interpretabilitas semakin meningkat karena pengguna tidak hanya membutuhkan hasil rekomendasi, tetapi juga alasan yang mendasarinya. Untuk mengatasi keterbatasan tersebut, dikembangkan pendekatan berbasis aspek seperti *Explicit Factor Model* (EFM), yang secara eksplisit memodelkan keterkaitan antara aspek produk yang diekstraksi dari ulasan dan preferensi pengguna terhadap aspek tersebut.

Berdasarkan penelitian oleh Peng Bai, Yang Xia dalam artikel *Fusing Knowledge and Aspect Sentiment for Explainable Recommendation*, sistem rekomendasi yang dapat dijelaskan (*explainable recommender system*) perlu memanfaatkan informasi aspek dan sentimen dari ulasan pengguna secara lebih

terstruktur. Penelitian tersebut menekankan bahwa pemodelan berbasis aspek memungkinkan sistem tidak hanya memprediksi rating, tetapi juga memberikan justifikasi rekomendasi berdasarkan aspek tertentu yang relevan bagi pengguna.

Dengan menggabungkan informasi pengetahuan dan analisis sentimen aspek, pendekatan ini meningkatkan kualitas penjelasan sekaligus mempertahankan performa prediksi rekomendasi (Bai et al., 2020).

Lebih lanjut, Liao et al. (2025) mengembangkan model *Aspect-Enhanced Explainable Recommendation* yang mengintegrasikan informasi aspek dalam pembelajaran representasi pengguna dan item, sehingga meningkatkan akurasi sekaligus transparansi rekomendasi (Liao et al., 2025).

Dalam konteks nasional, penelitian terkait analisis sentimen dan pemodelan topik berbasis BERTopic juga menunjukkan performa yang signifikan. Studi pada data ulasan serta media sosial berbahasa Indonesia mengindikasikan bahwa pendekatan berbasis Transformer mampu menghasilkan representasi topik yang lebih koheren dibandingkan metode probabilistik tradisional seperti LDA, khususnya dalam memahami persepsi dan opini pengguna terhadap suatu produk atau layanan (Ningrum et al., 2025). Penelitian lain yang membandingkan LDA dan BERTopic dalam analisis ulasan pelanggan menunjukkan bahwa pendekatan berbasis *contextual embedding* lebih efektif dalam mengidentifikasi aspek dominan yang memengaruhi kepuasan konsumen (Pratama, A. R., & Sarno, R. 2025).

Dalam konteks industri kreatif Indonesia, batik merupakan komoditas budaya dengan nilai ekonomi dan simbolik yang tinggi. Digitalisasi pemasaran melalui *e-commerce* telah meningkatkan jumlah ulasan pelanggan terhadap produk batik. Maharani dan Yel (2025) menunjukkan bahwa ulasan pelanggan batik mengandung informasi sentimen yang kuat terkait kualitas bahan, motif, harga, dan layanan penjual (Maharani & Yel, 2025). Namun, penelitian tersebut masih terbatas pada klasifikasi sentimen dan belum mengintegrasikan ekstraksi aspek ke dalam sistem rekomendasi yang bersifat *explainable*.

Berdasarkan telaah literatur tersebut, dapat diidentifikasi adanya celah penelitian pada integrasi metode *topic modeling* berbasis *transformer* dengan model rekomendasi berbasis aspek yang mendukung transparansi dan

pengambilan keputusan. Hingga saat ini, penelitian yang secara eksplisit mengombinasikan BERTopic sebagai metode ekstraksi aspek dengan Explicit Factor Model (EFM) dalam kerangka *aspect-based recommendation, system* pada domain produk batik Indonesia masih sangat terbatas.

Oleh karena itu, penelitian ini berfokus pada implementasi model *aspect-based recommendation*, menggunakan BERTopic dan *Explicit Factor Model* (EFM) pada ulasan produk batik Indonesia. Penelitian ini bertujuan menghasilkan sistem rekomendasi yang tidak hanya akurat dalam memprediksi preferensi pengguna, tetapi juga mampu memberikan penjelasan yang transparan berdasarkan aspek produk yang dianalisis. Evaluasi kinerja dilakukan menggunakan metrik terstandar seperti *Root Mean Square Error* (RMSE) dan Precision-K untuk mengukur tingkat akurasi dan ketepatan rekomendasi yang dihasilkan. Dengan demikian, penelitian ini diharapkan memberikan kontribusi teoritis dalam integrasi NLP berbasis *transformer* dengan sistem rekomendasi berbasis aspek, serta kontribusi praktis dalam pengembangan Sistem Pendukung Keputusan berbasis ulasan teks pada industri batik Indonesia.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Meningkatnya jumlah ulasan produk batik Indonesia pada platform e-commerce menghasilkan sumber data yang kaya akan informasi mengenai pengalaman dan preferensi konsumen. Namun, ulasan tersebut umumnya berbentuk teks tidak terstruktur dan mencakup berbagai aspek produk. Berdasarkan analisis frekuensi kata kunci pada dataset penelitian ini, aspek yang paling sering dibahas adalah desain dan motif batik (69,6% ulasan), kualitas (bahan, jahitan, dan material) (70,4% ulasan), serta kenyamanan pemakaian (40,7% ulasan). Kondisi ini menunjukkan bahwa preferensi konsumen batik bersifat multidimensional dan didominasi oleh pertimbangan estetika dan kualitas material, bukan semata-mata faktor harga. Hal ini menyulitkan sistem rekomendasi konvensional yang hanya berfokus pada rating numerik untuk menghasilkan rekomendasi yang relevan dan informatif bagi konsumen batik.

Kondisi ini menyulitkan sistem rekomendasi konvensional yang masih berfokus pada rating numerik atau preferensi umum pengguna untuk menghasilkan rekomendasi yang relevan dan informatif.

Selain itu, sebagian besar sistem rekomendasi belum mampu mengidentifikasi aspek-aspek spesifik yang memengaruhi kepuasan pengguna serta belum memberikan transparansi terkait alasan di balik rekomendasi yang dihasilkan. Padahal, pemahaman terhadap aspek produk yang dibahas dalam ulasan konsumen sangat penting untuk meningkatkan akurasi rekomendasi, khususnya pada produk batik yang memiliki karakteristik dan preferensi subjektif yang beragam.

Berdasarkan permasalahan tersebut, diperlukan suatu pendekatan sistem rekomendasi yang mampu mengekstraksi aspek-aspek penting dari ulasan produk secara semantik serta memodelkan preferensi pengguna terhadap aspek-aspek tersebut secara eksplisit.

1.2.2 Metode Usulan

Untuk mengatasi permasalahan yang telah diuraikan, penelitian ini mengusulkan penerapan model *aspect-based recommendation*, dengan mengombinasikan BERTopic dan *Explicit Factor Model* (EFM). BERTopic digunakan untuk melakukan *topic modeling* berbasis representasi semantik guna mengekstraksi aspek-aspek utama yang terkandung dalam ulasan produk batik Indonesia. Metode ini dipilih karena kemampuannya dalam menghasilkan topik yang lebih koheren dan relevan pada data teks pendek.

Selanjutnya, aspek-aspek yang dihasilkan dari BERTopic dimanfaatkan dalam penerapan *Explicit Factor Model* (EFM) untuk memodelkan hubungan antara pengguna, produk, dan aspek secara eksplisit. Dengan pendekatan ini, sistem rekomendasi diharapkan mampu memahami preferensi pengguna terhadap aspek tertentu dari produk batik, sehingga dapat menghasilkan rekomendasi yang lebih akurat, transparan, dan sesuai dengan kebutuhan pengguna.

1.2.3 Pertanyaan Penelitian

Mengacu pada latar belakang permasalahan dan perumusan masalah yang telah disusun, penelitian ini merumuskan beberapa pertanyaan penelitian sebagai berikut:

1. Bagaimana evaluasi BERTopic dalam mengekstraksi aspek dari ulasan produk batik Indonesia?
2. Bagaimana implementasi *Explicit Factor Model* (EFM) dalam sistem rekomendasi berbasis aspek pada produk batik Indonesia?
3. Bagaimana integrasi BERTopic dan EFM dapat meningkatkan kualitas rekomendasi produk batik Indonesia dibandingkan pendekatan konvensional?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Penelitian ini bertujuan untuk:

1. Menganalisis evaluasi BERTopic dalam mengekstraksi dan mengidentifikasi aspek-aspek utama dari ulasan produk batik Indonesia.
2. Mengimplementasikan *Explicit Factor Model* (EFM) sebagai model sistem rekomendasi berbasis aspek pada produk batik Indonesia.
3. Mengintegrasikan hasil ekstraksi aspek dari BERTopic ke dalam EFM untuk memodelkan preferensi pengguna dan kualitas produk secara eksplisit.

1.3.2 Manfaat

Manfaat yang diharapkan dari penelitian ini meliputi:

1. Manfaat Akademis

Memberikan kontribusi ilmiah dalam pengembangan dan penerapan sistem rekomendasi berbasis aspek, khususnya dengan mengintegrasikan *topic modeling* modern dan model rekomendasi yang bersifat *explainable* pada produk budaya lokal Indonesia.

2. Manfaat Penelitian Selanjutnya

Menjadi referensi dan dasar pengembangan bagi penelitian selanjutnya yang mengkaji integrasi BERTopic atau metode *topic modeling* serupa dengan sistem rekomendasi berbasis aspek dan faktor eksplisit.

3. Manfaat Praktis

Memberikan solusi konseptual dan praktis bagi pelaku *e-commerce* batik dalam meningkatkan kualitas dan relevansi rekomendasi produk kepada konsumen berdasarkan aspek-aspek yang dianggap penting oleh pengguna.

1.4 Batasan Masalah

Adapun batasan masalah yang dihadapi ketika mengerjakan penulisan proposal skripsi ini adalah:

1. Data yang digunakan dalam penelitian ini berupa ulasan teks produk batik Indonesia yang diperoleh dari platform *e-commerce* Shopee.
2. Bahasa ulasan yang dianalisis dibatasi pada bahasa Indonesia.
3. Metode yang digunakan untuk ekstraksi aspek dan topik dalam ulasan adalah BERTopic, tanpa melakukan perbandingan dengan metode *topic modeling* lainnya.
4. Model sistem rekomendasi yang digunakan dalam penelitian ini adalah Explicit Factor Model (EFM).
5. Evaluasi penelitian difokuskan pada kualitas rekomendasi dan analisis aspek, tanpa membahas implementasi sistem secara penuh dalam lingkungan produksi (*production environment*).

1.5 Sistematika Penulisan

Sistematika penulisan dalam skripsi ini adalah sebagai berikut:

Bab I Pendahuluan

Bab ini memuat gambaran umum mengenai penelitian yang dilakukan. Pembahasan pada bab ini meliputi latar belakang penelitian, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan skripsi.

Bab II Tinjauan Pustaka

Bab ini berisi landasan teori dan kajian pustaka yang relevan dengan penelitian. Pembahasan mencakup konsep dasar sistem rekomendasi, *text mining* dan *natural language processing*, *aspect-based recommendation*, *topic modeling* dengan BERTopic, *Explicit Factor Model* (EFM), serta penelitian-penelitian terdahulu yang berkaitan dengan sistem rekomendasi berbasis ulasan teks. Pada bab ini juga

disajikan kerangka pemikiran penelitian sebagai dasar konseptual pelaksanaan penelitian.

Bab III Metodologi Penelitian

Bab ini menjelaskan tahapan dan metode yang digunakan dalam penelitian. Pembahasan meliputi desain penelitian, sumber dan teknik pengumpulan data ulasan produk batik Indonesia dari platform *e-commerce* Shopee, proses *preprocessing* data teks, metode ekstraksi aspek menggunakan BERTopic, perancangan dan implementasi model sistem rekomendasi menggunakan *Explicit Factor Model* (EFM), serta metode evaluasi yang digunakan untuk mengukur kualitas rekomendasi.

BAB IV Hasil dan Pembahasan

Bab ini menyajikan hasil dan analisis dari setiap tahap penelitian. Pembahasan diawali dengan eksplorasi data (EDA) untuk memahami karakteristik dataset ulasan, dilanjutkan dengan ringkasan efektivitas tahap *preprocessing* teks. Setelah itu, dipaparkan hasil pemodelan topik menggunakan BERTopic beserta visualisasi distribusi aspek produk yang ditemukan. Bab ini juga mendemonstrasikan implementasi *Explicit Factor Model* (EFM) dalam memprediksi rating dan menyusun perankingan produk, yang dievaluasi melalui 18 skenario pengujian menggunakan metrik RMSE dan Precision-K. Bagian akhir membahas kemampuan sistem dalam memberikan justifikasi rekomendasi serta tampilan prototipe dasbor berbasis Streamlit.

BAB V Kesimpulan dan Saran

Bab ini merupakan bagian penutup yang merangkum seluruh hasil penelitian. Kesimpulan disusun untuk menjawab efektivitas integrasi BERTopic dan EFM dalam membangun sistem rekomendasi berbasis aspek pada ulasan produk batik Indonesia. Selain itu, bab ini juga memuat saran untuk pengembangan penelitian selanjutnya, seperti perluasan sumber data ulasan, penerapan metode deep learning, serta penggunaan model transformer khusus Bahasa Indonesia seperti IndoBERT.