BAB I
PENDAHULUAN

1.1. Latar Belakang Masalah

Transformasi ekosistem digital telah menghadirkan tantangan mendasar dalam konsumsi informasi, di mana pengguna dihadapkan pada volume dan keragaman konten yang terus meningkat sementara kapasitas pemahaman untuk memproses informasi tetap terbatas (Eppler & Mengis, 2004). Kondisi ini memicu fenomena *information overload*, yakni situasi ketika kompleksitas dan kuantitas informasi melampaui kemampuan individu atau sistem dalam melakukan kurasi dan pemeringkatan konten secara efektif (Roetzel, 2018). Dalam domain sistem rekomendasi film, *information overload* terwujud sebagai kesulitan sistem dalam merepresentasikan preferensi pengguna secara akurat, khususnya ketika berhadapan dengan ribuan film yang memiliki atribut serupa namun harus dibedakan berdasarkan berbagai dimensi fitur secara bersamaan (Fayyaz et al., 2020).

Sistem rekomendasi konvensional umumnya mengandalkan dua paradigma utama: *content-based filtering* yang bergantung pada metadata eksplisit dan *collaborative filtering* yang memanfaatkan pola interaksi historis antar pengguna dan item. Meskipun kedua pendekatan ini menunjukkan efektivitas pada skala data normal, mereka menghadapi limitasi signifikan dalam konteks kompleksitas tinggi. Masalah *cold-start* untuk item atau pengguna baru, tingkat *sparsity* matriks interaksi yang ekstrem, serta keterbatasan dalam pembentukan representasi item yang hanya mengandalkan subset kecil dari atribut yang tersedia, menjadi hambatan utama dalam meningkatkan akurasi rekomendasi (Rodríguez-Baena et al., 2025; Z. Zhou et al., 2023). Dampaknya, sistem sering kali gagal membedakan relevansi antar film yang memiliki pola interaksi atau metadata serupa, yang pada akhirnya menurunkan kualitas pengalaman pengguna.

Setiap film pada dasarnya menyimpan informasi yang beragam dan melengkapi representasi visual melalui poster yang menangkap estetika dan genre, deskripsi naratif dalam sinopsis yang mengandung plot dan tema, serta metadata terstruktur seperti rating, popularitas, dan informasi aktor yang mencerminkan

konteks sosial. Integrasi berbagai modalitas ini memungkinkan pembentukan representasi item yang jauh lebih kaya dibandingkan pendekatan monomodal (Mu & Wu, 2023). Arsitektur Transformer, dengan mekanisme *self-attention* dan *cross-attention*-nya, telah terbukti efektif dalam menangkap relasi kompleks antar modalitas dan mengekstraksi fitur semantik tingkat tinggi dari data visual maupun tekstual (Sun et al., 2019). Pendekatan multimodal berbasis Transformer memberikan kemampuan bagi sistem untuk memodelkan karakteristik film secara komprehensif, yang menjadi krusial dalam mengatasi *information overload* yang memerlukan pemrosesan fitur dengan dimensi dan kompleksitas tinggi (Liu et al., 2025).

Di sisi lain, LightGCN telah muncul sebagai metode *collaborative filtering* berbasis graf yang efisien dalam memodelkan relasi implisit antara pengguna dan item melalui propagasi sinyal pada struktur jaringan bipartit (X. He et al., 2020). Namun, representasi *node* item dalam LightGCN yang hanya bergantung pada pola interaksi historis belum sepenuhnya memanfaatkan kekayaan informasi atribut yang tersedia pada setiap film. Ketika menghadapi kondisi *sparsity* tinggi atau item dengan interaksi terbatas, *embedding* yang dihasilkan cenderung kurang informatif. Hal ini menciptakan peluang untuk memperkaya proses pembelajaran representasi dengan mengintegrasikan fitur multimodal dan metadata terstruktur ke dalam mekanisme propagasi graf, sebagaimana ditunjukkan oleh penelitian terkini yang mengeksplorasi sinergi antara konten semantik dan struktur kolaboratif (Kim et al., 2023).

Berdasarkan identifikasi gap tersebut, penelitian ini mengembangkan sistem rekomendasi film dengan arsitektur hibrida yang mengintegrasikan pemrosesan multimodal berbasis Transformer dengan pemodelan graf kolaboratif LightGCN. Fitur visual diekstraksi dari poster film menggunakan Vision Transformer (ViT) yang di-*fine-tune*, sementara fitur tekstual dari sinopsis diproses melalui BERT yang juga di-*fine-tune* pada domain perfilman. Informasi *cast* dan genre direpresentasikan melalui *embedding layer* dan proyeksi linear. Keempat representasi difusikan melalui modul *cross-attention* menjadi *embedding* multimodal yang dikombinasikan secara aditif dengan *embedding* kolaboratif item melalui parameter α yang dipelajari secara adaptif sebelum memasuki propagasi

graf LightGCN. Pendekatan ini ditujukan untuk meningkatkan akurasi rekomendasi dalam kondisi *information overload*, mengurangi dampak *cold-start* pada item dengan interaksi terbatas, serta meningkatkan performa metrik rekomendasi top-N seperti Precision@K, Recall@K, dan NDCG@K pada *dataset* MovieLens yang telah diperkaya dengan metadata dari TMDb.

1.2. Perumusan Masalah

Berdasarkan latar belakang tersebut, penelitian ini merumuskan dua fokus penelitian utama sebagai berikut:

a) Bagaimana merancang mekanisme ekstraksi dan integrasi fitur multimodal dari poster, sinopsis, dan metadata film untuk membentuk representasi item yang komprehensif, sekaligus mengintegrasikannya sebagai inisialisasi *embedding* item dalam arsitektur graf bipartit LightGCN sehingga mampu memperkaya propagasi sinyal kolaboratif dan mengatasi masalah *cold-start* lebih efektif dibandingkan pendekatan berbasis rating konvensional tanpa meningkatkan kompleksitas komputasi secara signifikan?
b) Sejauh mana peningkatan performa rekomendasi top-N dari arsitektur hibrida Transformer-LightGCN yang diukur melalui metrik Precision@K, Recall@K, NDCG@K, dan Hit Ratio@K dibandingkan dengan *collaborative filtering klasik (ItemKNN), LightGCN* murni, dan model multimodal berbasis *CNN (VBPR)?*

1.3. Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditetapkan, penelitian ini memiliki dua tujuan utama yang masing-masing dirancang untuk menjawab setiap fokus masalah di atas:

a) Mengembangkan mekanisme ekstraksi fitur multimodal dari poster film menggunakan Vision Transformer, sinopsis menggunakan BERT, serta informasi pemeran dan genre melalui *embedding layer*, yang seluruhnya difusikan melalui modul *cross-attention* untuk membentuk representasi item tunggal yang komprehensif, kemudian mengintegrasikan representasi tersebut

sebagai inisialisasi *embedding* item pada graf bipartit LightGCN agar sinyal konten dapat langsung memperkaya proses propagasi sinyal kolaboratif sekaligus mengatasi keterbatasan *cold-start* tanpa menambah beban komputasi secara signifikan.

b) Mengevaluasi performa model hibrida Transformer-LightGCN yang diusulkan secara komprehensif menggunakan metrik rekomendasi top-N, yaitu Precision@K, Recall@K, NDCG@K, dan Hit Ratio@K pada *dataset* MovieLens yang diperkaya metadata TMDb, serta membandingkan hasilnya dengan tiga metode *baseline* representatif, yaitu ItemKNN, LightGCN murni, dan VBPR.

1.3.2 Manfaat Penelitian

Penelitian ini memberikan manfaat baik secara teoretis maupun praktis yang selaras dengan fokus penelitian:

a) Manfaat terkait pengembangan mekanisme multimodal dan integrasi graf: Secara teoretis, penelitian ini berkontribusi pada pengembangan kerangka kerja hibrida yang mengintegrasikan pembelajaran representasi multimodal dengan pemodelan graf kolaboratif, memperluas pemahaman tentang bagaimana sinyal konten dan sinyal kolaboratif dapat digabungkan secara efektif untuk mengatasi masalah *cold-start* dalam kondisi *information overload*. Secara praktis, mekanisme ekstraksi dan fusi multimodal yang dikembangkan dapat diadopsi oleh platform streaming digital dan layanan video-on-demand untuk memperkaya representasi item dengan memanfaatkan berbagai sumber informasi yang tersedia, sementara integrasi dengan LightGCN memungkinkan implementasi yang efisien tanpa memerlukan infrastruktur komputasi yang sangat besar.

b) Manfaat terkait evaluasi performa dan perbandingan model: Secara teoretis, evaluasi komprehensif dengan studi ablasi dan stratifikasi popularitas memberikan bukti empiris mengenai kontribusi spesifik setiap komponen arsitektur serta efektivitas pendekatan hibrida dibandingkan metode baseline klasik dan state-of-the-art. Secara praktis, hasil perbandingan dengan ItemKNN, LightGCN murni, dan VBPR memberikan panduan bagi praktisi

dalam memilih pendekatan rekomendasi yang sesuai dengan karakteristik *dataset* dan kebutuhan bisnis, sementara pipeline evaluasi yang dapat direplikasi memfasilitasi pengembangan dan validasi sistem rekomendasi pada *dataset* publik yang terstandarisasi. Bagi peneliti di bidang sistem rekomendasi, penelitian ini menyediakan arsitektur hibrida dan pipeline evaluasi yang dapat direplikasi sebagai referensi pengembangan sistem serupa pada domain konten yang berbeda.

1.4. Kontribusi dan Orisinalitas Penelitian

Penelitian ini memiliki unsur kebaruan dan memberikan kontribusi sebagai berikut:

- Arsitektur hibrida *end-to-end* yang mengintegrasikan fusi multimodal berbasis Transformer dengan *embedding* kolaboratif item melalui mekanisme aditif berparameter α pada graf LightGCN.
- Pendekatan ini memungkinkan sinyal konten dan sinyal kolaboratif dipelajari secara bersamaan dalam satu kerangka terintegrasi, berbeda dari penelitian sebelumnya yang umumnya memisahkan proses pemodelan konten dan pemodelan graf.
- Pipeline lengkap dan dapat direplikasi untuk integrasi *dataset* MovieLens dengan metadata TMDb, mencakup pemetaan otomatis, pengambilan data melalui API, dan preprocessing multimodal terstandarisasi. Pipeline ini memungkinkan replikasi eksperimen secara sistematis tanpa memerlukan pengumpulan data manual yang kompleks.
- Evaluasi komprehensif dengan studi ablasi terperinci pada *dataset* MovieLens varian 100K dan 1M yang diperkaya dengan informasi multimodal. Evaluasi ini mengukur kontribusi spesifik setiap komponen arsitektur terhadap peningkatan performa rekomendasi top-N, memberikan bukti empiris mengenai efektivitas integrasi representasi konten dan graf kolaboratif.

1.5. Batasan Penelitian

Batasan dalam penelitian ini adalah sebagai berikut:

- *Dataset* yang digunakan terbatas pada MovieLens 100K dan MovieLens 1M yang diperkaya metadata dari TMDb.
- Modalitas informasi film yang diintegrasikan terdiri dari empat sumber: fitur visual dari poster melalui ViT, fitur tekstual dari sinopsis resmi TMDb melalui BERT, informasi lima aktor utama, dan genre.
- Seluruh deskripsi tekstual film menggunakan bahasa Inggris sesuai yang disediakan oleh TMDb.
- Representasi multimodal diterapkan pada sisi item (film), sedangkan *embedding* pengguna dipelajari secara implisit dari pola interaksi historis melalui komponen *collaborative filtering*.
- Evaluasi performa dilakukan secara *offline* menggunakan metrik rekomendasi *top-N* dengan protokol *sampled evaluation* 99 item negatif.
- Aplikasi di-*deploy* ke server publik melalui Streamlit Cloud untuk keperluan validasi pengguna.
- Proses pelatihan, evaluasi, dan demonstrasi sistem menggunakan *user ID* yang sudah tersedia dalam *dataset* MovieLens. Validasi pengguna dilakukan melalui kuesioner terhadap responden yang mencoba sistem secara langsung.
- Evaluasi *cold-start* dilakukan berdasarkan simulasi terhadap item dengan kurang dari 20 interaksi dalam *dataset*, bukan terhadap item baru yang ditambahkan secara *real-time* ke dalam sistem.

1.6. Sistematika Penulisan

Skripsi ini terdiri dari lima bab. Bab pertama merupakan pendahuluan yang membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, kontribusi dan orisinalitas, batasan penelitian, serta sistematika penulisan. Bab kedua menyajikan kajian pustaka yang mencakup landasan teori sistem rekomendasi, pembelajaran multimodal, arsitektur Transformer, dan LightGCN, serta mengulas penelitian terkait untuk mengidentifikasi posisi dan kebaruan penelitian. Bab ketiga menjelaskan metodologi penelitian, meliputi

pengumpulan data dari MovieLens dan TMDb, pengolahan data multimodal, perancangan arsitektur hibrida Transformer-LightGCN, serta rancangan evaluasi. Bab keempat menyajikan hasil eksperimen, analisis mendalam, implementasi antarmuka sistem, validasi pengguna, dan pembahasan temuan penelitian. Bab kelima berisi kesimpulan dan saran pengembangan lebih lanjut.