

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Perkembangan teknologi digital telah mengubah cara masyarakat mengakses dan mengonsumsi hiburan, termasuk dalam hal menonton film. Platform *streaming* seperti *Netflix*, *Vidio*, dan *Disney+* menyediakan ribuan judul film dan serial yang dapat diakses dengan mudah. Namun, melimpahnya pilihan ini justru dapat menimbulkan informasi berlebihan (*information overload*) yang menyulitkan pengguna menemukan konten sesuai preferensi, serta memicu kelelahan dalam pengambilan keputusan (*decision fatigue*) akibat terlalu banyak opsi yang tidak terpersonalisasi. Seperti yang diungkapkan oleh Iyengar dan Lepper (2000) dalam studi klasik mereka, dan diperkuat oleh tinjauan Misuraca *et al.* (2020) yang lebih baru, terlalu banyak pilihan justru dapat menurunkan kepuasan dan membebani kognisi pengguna. Temuan ini sejalan dengan penelitian Knijnenburg *et al.* (2012), yang menunjukkan bahwa kompleksitas antarmuka dapat secara signifikan menurunkan kepuasan dan efektivitas pengguna, karena mereka merasa kewalahan dengan banyaknya fitur dan interaksi yang tersedia.

Sebagai solusi atas permasalahan tersebut, berbagai model rekomendasi dikembangkan untuk membantu pengguna menemukan film yang relevan dan sesuai dengan minat mereka. Ricci *et al.* (2015) menjelaskan bahwa model rekomendasi berbasis *Collaborative Filtering* telah terbukti meningkatkan keterlibatan pengguna dengan mempertemukan mereka pada *item* yang relevan berdasarkan pola perilaku dan preferensi pengguna serupa sehingga meningkatkan

interaksi dan penggunaan sistem secara konsisten. Dalam konteks film, model rekomendasi tidak hanya membantu pengguna menemukan film yang menarik, tetapi juga mendukung eksposur film-film *niche* atau spesifik melalui konsep *long-tail recommendation* (Celma, 2010). Lebih jauh, Knijnenburg *et al.* (2012) menunjukkan bahwa efektivitas model dalam sistem rekomendasi dapat berdampak pada penurunan beban kognitif pengguna, misalnya melalui temuan bahwa aktivitas penelusuran (*browsing*) yang lebih sedikit dapat mengindikasikan persepsi efektivitas sistem yang lebih tinggi.

Pendekatan *Collaborative Filtering* sendiri terbagi menjadi dua jenis, yaitu berbasis memori (*memory-based*) dan berbasis model (*model-based*). Pendekatan *memory-based* membandingkan kesamaan antar pengguna atau *item* secara langsung, namun memiliki keterbatasan seperti *sparsity*, *scalability*, dan *cold-start problem*, terutama saat berhadapan dengan data dalam jumlah besar (Adomavicius & Tuzhilin, 2005). *Sparsity* (kelangkaan data) adalah kondisi umum yang mana sebagian besar nilai dalam matriks pengguna-*item* kosong karena hanya sedikit interaksi yang tercatat. *Scalability* (skalabilitas) mengacu pada kemampuan model untuk tetap efisien seiring bertambahnya data. Terakhir, *cold-start problem* berkaitan dengan kesulitan memberikan rekomendasi bagi pengguna atau *item* baru, yang mana menjadi perhatian penting dalam pengembangan model rekomendasi modern. Sebaliknya, pendekatan *model-based* membangun model prediktif menggunakan teknik *deep learning*, yang memungkinkan model menghasilkan rekomendasi yang lebih efisien, akurat, dan mampu mengatasi sebagian besar keterbatasan tersebut (Zhang *et al.*, 2019).

Dalam pendekatan *Collaborative Filtering*, umpan balik pengguna dapat dikategorikan menjadi dua jenis, yaitu umpan balik implisit (*implicit feedback*) dan umpan balik eksplisit (*explicit feedback*) (Hu *et al.*, 2008; Koren & Bell, 2011). Umpan balik implisit merujuk pada data interaksi tidak langsung, seperti jumlah pemutaran lagu, klik, atau waktu menonton, yang tidak menyatakan preferensi secara eksplisit tetapi bisa diinterpretasikan sebagai minat pengguna terhadap *item* tertentu. Menurut Hu *et al.* (2008), meskipun data implisit lebih melimpah, ia mengandung ambiguitas lebih tinggi karena tidak mengandung informasi negatif eksplisit, sehingga membutuhkan pendekatan pemodelan khusus. Sedangkan umpan balik eksplisit berasal dari masukan langsung dari pengguna, seperti rating terhadap film, dan umumnya dianggap lebih akurat dalam merepresentasikan preferensi. Salah satu teknik *model-based* yang populer untuk memanfaatkan data rating eksplisit adalah *matrix factorization* (MF), yang memetakan matriks interaksi pengguna-*item* ke dalam dua matriks *latent* (tersembunyi) yang merepresentasikan karakteristik pengguna dan *item*.

Salah satu algoritma utama dalam implementasi *matrix factorization* adalah *Alternating Least Squares* (ALS). ALS bekerja secara iteratif dengan memperbarui matriks fitur pengguna dan *item* secara bergantian untuk meminimalkan kesalahan prediksi rating. Zhou *et al.* (2008) menunjukkan bahwa ALS memiliki akurasi prediksi yang lebih tinggi dibandingkan metode lain seperti *Stochastic Gradient Descent* (SGD), serta lebih efisien dalam menangani data berskala besar. Koren (2008) menambahkan bahwa *matrix factorization*, seperti SVD dan ALS, secara efektif mengurangi *sparsity* pada matriks interaksi pengguna-*item* dengan memproyeksikan data ke ruang laten berdimensi rendah, sehingga menghasilkan

representasi yang lebih padat dan memungkinkan estimasi rating pada data yang hilang. Selain itu, Priyati *et al.* (2022) menunjukkan bahwa ALS memiliki performa kompetitif dalam model rekomendasi berbasis rating, terutama jika dibandingkan dengan algoritma lain, seperti ALS-WR, SVD, dan SVD++.

Penelitian ini bertujuan untuk menghasilkan model rekomendasi film berbasis rating menggunakan algoritma *Alternating Least Squares* (ALS). Pemilihan ALS didasarkan pada kemampuannya dalam menangani masalah *sparsity* dan skalabilitas pada data rating eksplisit, serta efisiensinya dalam membangun model yang dapat dioperasikan pada skala data besar. Data rating film dalam penelitian ini diperoleh dari MovieLens, salah satu platform informasi film terbesar yang menyediakan rating eksplisit dari jutaan pengguna di seluruh dunia, sehingga menjadi dasar yang relevan dan representatif untuk membangun model. Melalui pendekatan ini, model rekomendasi akan mempelajari pola preferensi pengguna untuk menghasilkan prediksi rating yang akurat, yang kemudian dapat digunakan dalam model aplikasi rekomendasi film yang personal dan adaptif. Diharapkan, model ini dapat meningkatkan pengalaman pengguna dalam menjelajahi konten film, mengurangi beban dalam pengambilan keputusan, serta memberikan kontribusi dalam pengembangan teknologi rekomendasi di industri hiburan digital.

## **1.2 Rumusan Masalah**

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

1. Bagaimana algoritma *Alternating Least Squares* (ALS) dapat diimplementasikan untuk menghasilkan model rekomendasi film berbasis rating pengguna?
2. Bagaimana evaluasi kinerja model rekomendasi film yang dibangun menggunakan algoritma *Alternating Least Squares* (ALS) berdasarkan data rating dilihat dari metrik RMSE dan MAE?
3. Bagaimana analisis hasil rekomendasi film yang dihasilkan dari implementasi algoritma *Alternating Least Squares* (ALS)?

### 1.3 Batasan Masalah

Terdapat batasan masalah pada penelitian ini yaitu:

1. Penelitian ini hanya menggunakan data rating eksplisit sebagai indikator preferensi pengguna terhadap *item*, yang mana dalam konteks penelitian ini merujuk pada film.
2. *Dataset* yang digunakan merupakan data versi kecil dari data MovieLens yang telah melalui proses pra-pemrosesan dan pembersihan (*data cleaning*) untuk menghilangkan *noise* dan data yang tidak relevan.
3. Model rekomendasi dibangun menggunakan metode *Collaborative Filtering* dengan pendekatan *matrix factorization*, yaitu algoritma *Alternating Least Squares* (ALS).
4. Evaluasi kinerja model dilakukan menggunakan dua metrik evaluasi, yaitu *Root Mean Square Error* (RMSE) dan *Mean Absolute Error* (MAE).

### 1.4 Tujuan Penelitian

Adapun tujuan yang dilakukan dalam penelitian skripsi ini adalah:

1. Menghasilkan model rekomendasi film berbasis rating menggunakan algoritma *Alternating Least Squares* (ALS).
2. Mengukur kinerja model rekomendasi menggunakan dua metrik evaluasi, yaitu *Root Mean Square Error* (RMSE) dan *Mean Absolute Error* (MAE), untuk menilai ketepatan prediksi.
3. Menganalisis hasil rekomendasi film yang dihasilkan dari implementasi algoritma *Alternating Least Squares* (ALS).