

BAB I PENDAHULUAN

1.1 Latar Belakang

Teknologi internet dan *mobile* memberikan kemudahan bagi penggunanya dalam mengakses informasi kapanpun dan dimanapun. Hal tersebut juga mempengaruhi gaya hidup masyarakat yang lebih intens dalam melakukan aktivitas secara *online*, mulai dari mengakses media social, menonton film, *e-commerce*, dan berbagai aplikasi lainnya[1]. Seiring berjalannya waktu, volume dan jumlah data akan terus meningkat. Hal ini dapat menimbulkan kesulitan bagi pengguna dan memerlukan lebih banyak waktu untuk memilih konten yang sesuai dengan preferensinya. Berdasarkan permasalahan tersebut maka salah satu solusi yang bisa diterapkan adalah dengan mengimplementasikan sistem rekomendasi.

Sistem rekomendasi merupakan alat yang efisien dalam menyaring informasi yang tersebar secara luas, adanya perubahan kebiasaan penggunaan komputer, tren personalisasi, dan munculnya akses internet. Sistem rekomendasi banyak mendapat perhatian akhir-akhir ini, karena mampu menyarankan berbagai hal kepada pengguna secara efektif. Tujuan dari mengimplementasikan system rekomendasi diantaranya meningkatkan jumlah penjualan, merekomendasikan ke teman, meningkatkan kepuasan pengguna, meningkatkan pendapatan perusahaan, dan lainnya[2]. Metode yang dapat digunakan untuk memberikan rekomendasi yang sesuai dengan minat pengguna adalah *Collaborative Filtering*. *Collaborative Filtering* mampu memberikan rekomendasi dengan tepat, namun menghadapi masalah berupa *scalability*, *sparsity*, dan *cold start*[3]. *Scalability* merupakan kondisi dimana sistem mengalami kegagalan pada kondisi data yang mengalami peningkatan baik jumlah item/pengguna, dan tidak mampu memberikan rekomendasi dengan waktu respon yang wajar. Hal ini disebabkan karena dimensi data yang tinggi atau data dengan ukuran yang sangat besar, sehingga membutuhkan banyak waktu untuk menghitung korelasi antar setiap pasangan pengguna atau pasangan item[2]. *Sparsity* terjadi karena matrik rating sangat jarang, disebabkan karena pengguna sangat sedikit memberikan rating terhadap item yang mereka beli atau lihat[2]. *Cold start* terjadi pada saat pengguna baru atau produk baru, sehingga peminatan pengguna belum di ketahui[4].

Beberapa penelitian telah dilakukan oleh peneliti terdahulu untuk menyelesaikan masalah tersebut, diantaranya yang dilakukan oleh Xiong, C et al, mengusulkan *K-Means clustering* dan *missing value prediction* menggunakan algoritma *similarity*[5]. Vahidy Rodpysh et al, mengusulkan *multi-level singular value decomposition* untuk menyelesaikan masalah *cold start* dan *sparsity*[6]. Hawashin et al, menggunakan teknik *machine learning* dan *user interest* untuk mengatasi masalah *cold start*[7]. Sementara pada penelitian ini penulis mengusulkan pendekatan *clustering* dan teknik *imputation* untuk menyelesaikan masalah *cold start* dan *sparsity*. Algoritma *cluster* yang digunakan adalah *K-Means*, sementara untuk teknik *imputation* menggunakan fungsi *average*, dan dilanjutkan dengan algoritma *KNN*.

1.2 Batasan Masalah

Dataset yang digunakan pada penelitian ini adalah Dataset Movielens 100k yang dikumpulkan oleh Proyek Penelitian GroupLens Universitas Minnesota dan tersedia di situs web <https://grouplens.org/datasets/movielens/100k/>. Dataset ini terdiri dari 100.000 peringkat dalam rentang skala 1-5, yang diberikan oleh 943 pengguna untuk 1682 film. Setiap pengguna setidaknya memberikan peringkat untuk 20 film, dan dataset juga mencakup informasi demografis seperti usia, jenis kelamin, pekerjaan, dan kode pos pengguna. Dataset ini mengandung *sparsity* sebesar 93,7%. Data pada dataset akan dikluster berdasarkan informasi demografi menggunakan algoritma *K-Means* yang diaplikasikan pada *tools* RapidMiner. Selanjutnya akan dilakukan imputasi untuk mengatasi *sparsity* menggunakan dua pendekatan terhadap hasil kluster. Pendekatan pertama yaitu melakukan percobaan imputasi pada data kluster secara menyeluruh, yaitu ketika data kluster masih tercampur antara kluster satu dengan kluster lainnya. Sedangkan pendekatan yang kedua yaitu melakukan percobaan imputasi pada data per masing-masing kluster secara terpisah. Lalu akan dilakukan pengujian menggunakan algoritma *KNN* untuk mengetahui perbandingan akurasi antara masing-masing pendekatan yang telah dilakukan.

1.3 Rumusan Masalah

Berdasarkan latar belakang permasalahan, dapat dirumuskan permasalahan dalam penelitian ini yaitu “Apakah teknik *clustering* dan *imputation* mampu mengatasi masalah *cold start* dan *sparsity* pada *collaborative filtering*?”

1.4 Tujuan Penelitian

Tujuan dari penelitian ini yaitu meneliti peran penerapan algoritma *K-Means* dan *imputation* terhadap tantangan *cold start* dan *sparsity* di dalam lingkup *Collaborative Filtering* beserta kualitas rekomendasi yang dihasilkan.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini yaitu ditemukan peran penerapan algoritma *K-Means* dan *imputation* terhadap tantangan *cold start* dan *sparsity* di dalam lingkup *Collaborative Filtering* beserta kualitas rekomendasi yang dihasilkan.