

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dalam beberapa tahun terakhir, perkembangan teknologi sangatlah pesat. Hal ini mengakibatkan ukuran data atau informasi juga berukuran besar sehingga pengguna mengalami kesulitan ketika akan mencari sebuah informasi yang sesuai keinginan pengguna [1]. Salah satu contohnya adalah merekomendasikan sebuah film dengan jumlah film yang tidak sedikit. Dengan adanya permasalahan ini, dibutuhkan sebuah sistem rekomendasi. Sistem rekomendasi bertujuan untuk mengidentifikasi preferensi dan menyarankan *item* yang relevan dengan pengguna. Terdapat beberapa *e-commerce* dan media sosial yang telah menerapkan sistem rekomendasi [2].

Pendekatan yang sering digunakan pada Sistem Rekomendasi adalah *Collaborative Filtering* (CF) [3], karena konsep dari pendekatan ini merekomendasikan berdasarkan rating suatu *item* film yang dinilai oleh pengguna aktif dan dibandingkan dengan nilai rating yang dinilai oleh pengguna lainnya sehingga menghasilkan kemiripan pengguna aktif dengan pengguna lainnya. Dengan kata lain, sistem mencari pengguna lain yang memiliki selera yang sama dengan pengguna aktif. Dalam *Collaborative Filtering* (CF) terdapat 2 kategori, yaitu pendekatan *memory-based* dan pendekatan *model-based*. Pada pendekatan *memory-based* juga terdapat 2 kategori [4], yaitu *user-based* dan *item-based*. *User-based* melakukan perhitungan berdasarkan kesamaan antar sesama pengguna. Sedangkan *item-based* melakukan perhitungan berdasarkan kesamaan antar sesama *item*. Pada pendekatan *model-based*, membangun sebuah model yang menggambarkan sebuah perilaku atau kebiasaan pengguna [3]. Terdapat beberapa permasalahan yang ada dalam Sistem Rekomendasi, diantaranya *cold start problem*, *synonym*, *sparsity*, *scalability*, dan lain sebagainya [5].

Dari beberapa permasalahan diatas, peneliti akan menyelesaikan permasalahan *sparsity* pada sistem rekomendasi. *Sparsity* adalah keadaan dimana terdapat banyaknya data atau nilai rating yang kosong [6]. Pada *Collaborative*

Filtering, semakin banyak data yang kosong maka akan semakin sulit juga dalam menentukan prediksi yang akurat pada sebuah item. Karena *Collaborative Filtering* menggunakan tetangga terdekat untuk merekomendasikan sebuah *item*, dan semakin sedikitnya nilai rating akan menimbulkan kesulitan dalam menghitung tetangga terdekat [5].

Untuk mengatasi permasalahan *sparsity*, terdapat beberapa pendekatan CF yang menerapkan *dimensionality reduction* yang dapat digunakan, seperti *Stochastic Gradient Descent* (SGD) [1], *Singular Value Decomposition* (SVD) [7], *Non-Negative Matrix Factorization* (NMF) [8], *Regularized single-element-based NMF* (RSNMF) [9] dan lain sebagainya. Terdapat juga beberapa penelitian yang menerapkan kombinasi antara teknik *dimensionality reduction* dengan *clustering*, seperti *Singular Value Decomposition* (SVD) dan *K-Means* [10], *Stochastic Gradient Descent* (SGD) dan *K-Means* [11], *Non-Negative Matrix Factorization* (NMF) dan *iCluster* [12], *Principal Component Analysis* (PCA) dan *K-Means* [13], dan lain sebagainya. Namun tidak semua pendekatan CF dapat mengatasi permasalahan data *sparsity*, seperti *memory-based* atau yang sering disebut dengan CF tradisional [8][14]. Karena *Memory-Based* melakukan perhitungan berdasarkan kemiripan antara pengguna dan *item*, sehingga menghasilkan rekomendasi yang kurang akurat. Pada *memory-based*, *item-based* lebih diunggulkan dibandingkan dengan *user-based* [15].

Pada penelitian ini, peneliti akan menggunakan pendekatan *Collaborative Filtering* yang menerapkan *Non-Negative Matrix Factorization* (NMF) dan *Fuzzy C-Means*. Dari beberapa penelitian, pendekatan NMF lebih unggul dibandingkan dengan pendekatan lainnya, seperti SVD [8][16] dan PCA [17]. Begitupun juga dengan pendekatan *Fuzzy C-Means* yang lebih unggul dibandingkan dengan dengan pendekatan *K-Means* [18].

Dengan adanya penelitian ini, diharapkan dapat mengatasi permasalahan data *Sparsity*, sehingga dapat memberikan hasil rekomendasi sesuai yang diharapkan oleh pengguna dengan akurasi yang tinggi. Selain itu, penelitian ini dapat dijadikan sebagai rujukan dalam membangun sebuah sistem rekomendasi yang memiliki permasalahan data *sparsity*.

1.2. Perumusan Masalah

1.2.1 Permasalahan

Salah satu solusi dalam mengatasi data *sparsity* adalah dengan menggunakan pendekatan CF yang menerapkan teknik *dimensionality reduction*. Dari beberapa penelitian sebelumnya, pendekatan CF yang menerapkan PCA dan *K-Means* dapat mengatasi permasalahan data *sparsity*, namun hasilnya kurang optimal. Tidak semua CF dapat digunakan dalam mengatasi data *sparsity*, seperti *memory-based*. *Memory-based* dibagi 2 yaitu *user-based* dan *item-based*, dimana *item-based* lebih diunggulkan. Maka dari permasalahan ini dibutuhkan pendekatan yang mengatasi permasalahan data *sparsity* untuk meningkatkan akurasi pada sistem rekomendasi.

1.2.2 Metode Usulan

Pendekatan yang diusulkan untuk mengatasi permasalahan data *sparsity* yaitu dengan menggunakan pendekatan *Collaborative Filtering* yang menerapkan *Non-Negative Matrix Factorization* (NMF) dan *Fuzzy C-Means* (FCM).

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan dan usulan solusi atau pendekatan maka didapatkan rumusan masalah :

1. Berapa akurasi rekomendasi yang dihasilkan oleh pendekatan CF yang menerapkan NMF dan FCM dalam mengatasi data *sparsity* ?
2. Apakah akurasi rekomendasi yang dihasilkan lebih baik daripada pendekatan *Item-Based Collaborative Filtering* (IBCF) dalam mengatasi data *sparsity* ?

1.3. Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Mengacu pada latar belakang serta rumusan masalah diatas maka berikut tujuan dari penelitian ini :

1. Untuk mengetahui hasil akurasi dari pendekatan CF yang menerapkan NMF dan FCM dalam mengatasi *sparsity* data.
2. Untuk mengetahui hasil akurasi dan pendekatan manakah yang lebih baik ketika dibandingkan dengan *Item-Based Collaborative Filtering* (IBCF) dalam mengatasi *sparsity* data.

1.3.2 Manfaat

Penelitian ini dapat menjadi rujukan dalam membangun sistem rekomendasi yang memiliki masalah dengan data *sparsity*.

1.4. Batasan masalah

Batasan masalah pada penelitian ini adalah dataset yang digunakan yaitu dataset *explicit* dari MovieLens 100K (<https://grouplens.org/datasets/movielens/>).

1.5. Sistematika Proposal

Sistematika penulisan laporan Proposal ini terdiri dari beberapa bab, yaitu :

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang, perumusan masalah, tujuan dan manfaat, batasan masalah, serta sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas tentang kajian pustaka yang berisi penjelasan sistem rekomendasi, *Non-Negative Matrix Factorization* (NMF), *Fuzzy C-Means Clustering* (FCM), *Pearson Coeficient Corelation* (PCC), pendekatan evaluasi, metrik evaluasi serta penelitian terkait.

BAB III METODE USULAN

Pada bab ini membahas tentang arsitektur rancangan sistem, *datasets*, skenario pengujian, alur kinerja pengujian sistem serta contoh implementasi.

BAB IV IMPLEMENTASI METODE

Pada bab ini membahas tentang tahapan pembuatan program, hasil dari uji coba, hasil perbandingan dengan pendekatan lain, serta hasil tampilan sistem.

BAB V PENUTUP

Pada bab ini membahas hasil kesimpulan dari hasil uji coba implementasi metode serta saran untuk mengembangkan sistem dengan penelitian terkait.

REFERENSI

Berisi tentang jurnal-jurnal atau referensi yang dijadikan sebagai acuan dalam penelitian ini.