

BAB I

PENDAHULUAN

1.1. Latar Belakang

Memahami kebutuhan dan harapan pengguna berdasarkan tingkat kesamaan barang yang dipilih merupakan topik pembahasan yang masih perlu dikembangkan untuk diteliti. Hal ini menjadi lebih menarik ketika suatu kebutuhan atau harapan pengguna dapat diprediksi secara sistematis dengan menggunakan mesin, sekaligus memberikan suatu saran atau biasa disebut rekomendasi kepada pengguna yang dituju. Dalam melakukan prediksi, tentu dibutuhkan metode yang tepat sehingga prediksi yang dihasilkan dapat membantu mesin tersebut dalam memberikan saran tentang barang atau *item* apa yang benar-benar diinginkan oleh pengguna [1].

Pada dasarnya, sistem rekomendasi diciptakan untuk membantu pengguna mengatasi kelebihan informasi dengan membuat model prediksi yang memperkirakan berapa banyak pengguna akan menyukai masing-masing *item* dalam *set* yang besar [1]. Hal ini menjelaskan bahwa dalam kejadian nyata seorang pengguna akan merasa kesulitan dalam memilih produk atau *item* mana yang diinginkan dari banyaknya produk atau *item* yang tersedia. Dengan permasalahan tersebut, maka sistem rekomendasi ini memang diperlukan oleh *user* guna membantu dalam memprioritaskan suatu informasi, sehingga *user* dapat mengurangi waktu pencarian yang menghabiskan banyak waktu dalam membaca informasi yang diperlukan atau disukai [2].

Sistem rekomendasi dapat dibuat dengan berbagai macam pendekatan. Pendekatan yang dimaksud ialah cara atau metode yang digunakan untuk menghasilkan rekomendasi berdasarkan atribut-atribut yang dimiliki oleh *item* yang akan direkomendasikan kepada pengguna. Atribut tersebut dapat berupa konten yang berisi ciri-ciri yang dimiliki oleh setiap *item*. Misalnya, konten yang dimiliki oleh film berupa judul, tahun rilis, jenis atau *genre* film, panjang durasi, dan lain-lain. Setiap *item* yang akan direkomendasikan tersebut juga mempunyai nilai yang diberikan oleh pengguna yang disebut nilai *rating* [2].

Secara umum, pendekatan yang paling banyak digunakan untuk menghasilkan rekomendasi ialah pendekatan *Collaborative Filtering* (CF) [3]. Pendekatan ini mengasumsikan bahwa beberapa orang yang menyukai hal-hal yang sama cenderung juga sama-sama menyukai

terhadap hal-hal yang lainnya[4]. Hal ini diperkuat dengan pendapat yang menyatakan bahwa CF adalah teknik rekomendasi yang paling populer digunakan saat ini dan sebagian besar di antaranya menggunakan kumpulan data *rating* dari *item* yang diberikan oleh pengguna [5].

Secara tradisional, CF dapat diimplementasikan berdasarkan dua model, yaitu *user-based* dan *item-based* [6]. Pada CF berbasis *item*, sistem menggunakan kesamaan antara pola penilaian *item*. Jika dua *item* cenderung memiliki pengguna yang sama-sama menyukai dan tidak menyukai, maka kedua *item* tersebut dianggap serupa dan pengguna diharapkan memiliki preferensi yang serupa untuk *item* serupa [7]. Kelebihan dari teknik ini yaitu sistem tidak memerlukan konten untuk dianalisa, tidak memerlukan banyak waktu untuk pengembangan bahasa, analisa dokumen, pengembangan alat *parsing* dan *stemming* kata [8]. Selain itu model *Item-based CF* menghasilkan kualitas rekomendasi lebih baik jika dibandingkan dengan model *User-based CF* [6, 9].

Namun, seperti pada kebanyakan metode rekomendasi lainnya, CF juga memiliki beberapa kekurangan. Kekurangan yang dimaksud seperti tidak dapat memberikan rekomendasi untuk *item* baru karena tidak adanya *rating* pengguna yang menjadi dasar prediksi (*cold-start problem*), dan masalah *sparsity* yaitu jumlah *item* melebihi jumlah pengguna dengan *margin* yang besar sehingga sulit menemukan *item* yang telah di *rating* oleh banyak orang [10].

CF secara tradisional memiliki dua permasalahan utama, yaitu *sparsity* dan *scalability* [3]. Untuk mendapatkan hasil rekomendasi yang lebih baik dengan data *sparse*, maka akan dilakukan pemadatan *rating* dengan mengisi nilai *rating* yang kosong [5]. Setelah nilai matriks *rating* terpenuhi akan dilakukan *clustering* untuk mengurangi cakupan CF berbasis *item* [3].

Sparsity masih tetap menjadi persoalan besar yang mempengaruhi komputasi *item similarity* [3]. Tentunya hal ini perlu diselesaikan dengan lebih teliti. *Sparsity* membuat sistem kesulitan untuk menentukan kesamaan antar pengguna sehingga berpengaruh pada kualitas rekomendasi yang dihasilkan [3]. Hal tersebut juga didukung pendapat yang menyatakan bahwa *sparsity* secara signifikan mempengaruhi metode rekomendasi dengan mengurangi akurasi hasil prediksi [5]. Untuk mengatasi data *sparsity*, dapat diselesaikan dengan menggunakan *SVD*, *Bayesian network*, *BP ANN*, dan lain-lain. Namun metode-metode tersebut tidak mempertimbangkan pengaruh dari atribut *item* [11]. Untuk memprediksi *item rating* dan mengisi matriks *user-item rating* serta

mengurangi masalah kekurangan data, kesamaan antar atribut *item* perlu dipertimbangkan dengan baik [11].

Selain permasalahan *sparsity*, model rekomendasi CF dipengaruhi oleh masalah *scalability* [3]. *Scalability* adalah permasalahan yang ada ketika sistem melakukan komputasi terhadap data dengan ukuran terlalu besar [12]. Untuk mengatasi permasalahan tersebut dapat dilakukan *clustering* untuk kelompok pengguna dengan pola penilaian serupa ke dalam kelompok sehingga dapat mengurangi cakupan lingkungan yang harus dipertimbangkan selama proses rekomendasi [13]. Hal serupa juga dapat dilakukan proses *clustering* dengan mengelompokkan *item* ke dalam beberapa kelas berdasarkan *rating user* terhadap *item* [4].

Dengan uraian yang telah dijelaskan di atas, penelitian ini diharapkan dapat menyelesaikan permasalahan *sparsity* dan *scalability* pada model *Item-based CF* secara tradisional (CF Tradisional) dengan menerapkan *item attribute* dan *Fuzzy C-Means* dalam model *Item-based CF* yang menggunakan metode *clustering* (CF *Clustering*) sehingga didapatkan hasil rekomendasi film yang lebih baik dan akurat.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang dijelaskan di atas, maka dapat ditentukan rumusan masalah sebagai berikut :

1. Apakah perhitungan gabungan similaritas antara *item similarity* dan *item attribute similarity* dapat menyelesaikan masalah *sparsity*?
2. Apakah penerapan *Fuzzy C-Means clustering* dalam model CF dapat menyelesaikan masalah *scalability*?
3. Apakah hasil rekomendasi CF *Clustering* lebih baik jika dibandingkan dengan rekomendasi CF Tradisional [3]?

1.3 Tujuan Penelitian

Mengacu pada rumusan masalah diatas maka tujuan dari penelitian ini yaitu :

1. Untuk menyelesaikan permasalahan *sparsity*.
2. Untuk menyelesaikan permasalahan *scalability*.
3. Untuk mengetahui perbandingan antara CF *Clustering* dan CF Tradisional.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini yaitu agar untuk kedepannya penerapan *item attribute* dan *Fuzzy C-Means clustering* dalam model *Item-based Collaborative Filtering* dapat diimplementasikan pada sistem rekomendasi.

1.5 Batasan Masalah

Pada penelitian tentang sistem rekomendasi ini terdapat beberapa batasan masalah yang diambil, yaitu :

- a. Data yang digunakan pada penelitian ini adalah kumpulan data *MovieLens* (<https://grouplens.org/datasets/movielens/100k/>).
- b. Teknik *clustering* yang digunakan adalah *Fuzzy C-means Clustering*.
- c. Sistem rekomendasi ini menggunakan data historikal sebagai *dataset*.
- d. Sistem rekomendasi ini melakukan pengujian secara *offline* menggunakan *K-Fold Cross Validation* dengan $k = 5$.

1.6 Metodologi Penelitian

Dalam proposal penelitian ini terdapat beberapa tahapan yang akan dilakukan dalam perancangan sistem rekomendasi, di antaranya :

a. Studi Pustaka

Pada tahap pertama ini akan dilakukan pencarian dan pengumpulan informasi yang diperlukan guna membantu pengerjaan penelitian tentang sistem rekomendasi. Data yang dikumpulkan di antaranya berupa materi dan referensi tentang sistem rekomendasi, metode *Collaborative Filtering*, *similarity measure*, metode *clustering*, dan penelitian terkait. Selain itu, dataset yang akan digunakan untuk riset juga dikumpulkan dan dipelajari.

b. Analisa dan Perancangan

Analisa sistem diperoleh setelah mempelajari data-data yang telah terkumpul. Selanjutnya, sistem dirancang menurut urutan dari proses analisa data sehingga diperoleh hasil yang diinginkan.

c. Implementasi Sistem

Setelah rancangan sistem selesai dibuat, maka langkah selanjutnya yaitu mengimplementasikan rancangan tersebut menjadi sebuah program. Proses implementasi menggunakan aplikasi *Matlab2015a*.

d. Uji Coba Sistem

Pada tahap pengujian sistem akan dilakukan proses pelatihan data dengan menggunakan model *K-Fold Cross Validation*. Dataset dibagi menjadi sejumlah k yang kemudian dilakukan proses analisa dan evaluasi.

e. Analisa dan Evaluasi

Proses analisa dan evaluasi dari data pelatihan dilakukan dengan evaluasi *Recall*, *Presicion*, *F1-Score*, dan *Mean Absolute Error (MAE)* . Evaluasi tersebut digunakan untuk mendapatkan nilai evaluasi rekomendasi sehingga tingkat akurasi hasil rekomendasi dapat diketahui.

f. Dokumentasi

Dokumentasi dilakukan mulai dari tahap studi pustaka hingga tahap evaluasi.

1.7 Sistematika Penulisan

Adapun sistematika pembahasan tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisikan latar belakang dari penelitian yang akan dilakukan, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II STUDI PUSTAKA

Pada bab ini dijelaskan mengenai materi-materi yang menjadi landasan dalam penyusunan penelitian ini. Materi-materi tersebut mengenai sistem rekomendasi, *dataset MovieLens*, *Collaborative Filtering*, *Fuzzy C-Means Clustering*, dan lain-lain.

BAB III PERANCANGAN SISTEM

Pada bab ini akan menjelaskan rancangan dari seluruh sistem yang akan dibuat, mulai kebutuhan *software* dan *hardware* yang dibutuhkan, rancangan sistem rekomendasi dengan model *Collaborative Fitering*, model pengujian serta alur dari sistem yang akan dikembangkan.

BAB IV IMPLEMENTASI SISTEM

Pada bab ini akan dijelaskan implementasi sistem dari rancangan sistem rekomendasi yang telah dibuat pada bab sebelumnya. Kemudian sistem akan diuji coba berdasarkan model dan skenario pengujian yang telah ditentukan.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dari hasil uji coba serta saran yang dapat digunakan untuk pengembangan sistem rekomendasi selanjutnya.