1 Pendahuluan

1.1 Latar Belakang

Jumlah konten dan informasi yang ada di internet saat ini sangatlah besar. Online ensiklopedia seperti Wikipedia berisi lebih dari 1,1 juta artikel, Flickr - sebuah media penyimpanan foto online memiliki sekitar 130 juta foto, Tehnorati memiliki lebih dari 41 juta blog dan 2,5 milyar *link* dalam indeksnya[8]. Hal seperti sering disebut sebagai *information overload*, dimana terdapat terlalu banyak informasi yang diterima namun belum dapat untuk diproses lebih lanjut. Untuk mengatasi *information overload* seperti ini, diperlukan suatu alat yang dapat membantu untuk menentukan atau mencari suatu informasi dari banyaknya informasi yang ditawarkan. *Recommender system* menjadi salah satu pilihan terbaik, sistem ini mampu memberikan saran atau rekomendasi kepada *user* untuk membantu menentukan suatu pilihan berdasarkan *preference*, prilaku *user*, dan informasi tentang *item* itu sendiri[3]. *Recommender system* diimplementasikan baik untuk komersial dan non-profit website untuk memprediksi kecenderungan *user*[2]. Salah satu contohnya adalah movielens.org yang mengumpulkan profile dari *user* untuk memprediksi film apa yang sebaiknya ditonton *user* tersebut.

Teknik *collaborative filtering* menjadi teknik paling sukses dalam memprediksi kecenderungan *user* dalam recomm*endation system[*5]. *Collaborative filtering* bertujuan mencari keterhubungan antar *user* dan mencari kesamaannya untuk seterusnya memberikan rekomendasi paling sesuai kepada seorang *user* berdasarkan profil yang mirip dengan *user* lainnya. Teknik *collaborative filtering* bekerja dengan mengumpulkan feedback dari *user* dalam dalam suatu form untuk memberi rating pada domain tertentu dan mengeksploitasi persamaan dan perbedaan antara beberapa profil *user* dalam menentukan bagaimana merekomendasi suatu *item*. Metode ini dapat dilakukan pada domain dimana tidak terdapat cukup konten yang berasosiasi dengan *item*-nya, atau dimana kontennya sulit untuk dianalisa melalui komputer – misalnya ide, opini, tanggapan, dan lain-lain. Metode *collaborative filtering* juga mempunyai kemampuan untuk memberikan rekomendasi-rekomendasi terhadap produk atau *item* yang dinginkan *user* walaupun *item* tersebut tidak berisi konten dari profil *user*. Karena alasan inilah, sistem *collaborative filtering* telah cukup sukses digunakan untuk membangun *recommender system* dalam domain yang beragam[3].

Terdapat berbagai algoritma untuk menentukan kemiripan antar *user* dalam metode *collaborative filtering*. Salah satunya adalah algoritma dengan pendekatan *Nearest Neighborhood* - KNN, pendekatan *Nearest Neighborhood* telah menjanjikan prediksi yang relatif akurat dan mudah untuk diimplementasikan[9]. Walaupun KNN telah secara global dipakai dalam *Collaborative filtering*, terdapat masalah dalam penanganan data yang besar karena diperlukan biaya komputasi yang mahal[9]. Banyak peneliti yang menganggap teknik *clustering* mempercepat proses komputasi dan memperbaiki akurasi dari rekomendasi yang dihasilkan[9]. Oleh karena itu, pada tugas akhir ini algoritma yang akan digunakan adalah algoritma KNN dengan terlebih dahulu melakukan teknik *clustering*, yang disebut ClustKNN. Tugas akhir ini akan menganalisa ketepatan rekomendasi yang dihasilkan dengan algoritma ClustKNN pada *recommender system* berbasis *user-based collaborative filtering* dengan menghitung *error rate* yang dihasilkan berdasarkan *Pearson's correlation coefficient* dengan parameter jumlah *cluster*, ukuran *neighborhood*, serta perbandingan *training set* dan *test set*.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka masalah yang dapat dirumuskan diantaranya:

1. Bagaimana menerapkan *recommender system* berbasis *collaborative filtering* dengan menggunakan algoritma *ClustKNN*?
2. Bagaimana ketepatan hasil prediksi yang dihasilkan berdasarkan banyaknya *training set dan test set*, jumlah *cluster*, serta ukuran *neighbor* dengan algoritma *ClustKNN* yang diterapkan pada *collaborative filtering*?
3. Bagaimana pengaruh algoritma *ClustKNN* terhadap akurasi prediksi rating yang dihasilkan terhadap algoritma *user-based KNN*?

Batasan masalah dalam pembuatan tugas akhir ini adalah:

1. Data set untuk training dan *test set* adalah data film yang telah dirating berasal dari movielens.org.
2. Produk yang akan direkomendasikan adalah film.
3. Algoritma yang akan digunakan adalah pendekatan algoritma KNN dengan *clustering* menggunakan K-Means.

1.3 Tujuan

Berdasarkan perumusan masalah di atas, maka tujuan yang ingin dicapai dalam pengerjaan Tugas Akhir ini adalah sebagai berikut:

1. Membangun *recommender system* berbasis *user-based collaborative filtering* yang mengimplementasikan algoritma *ClustKNN*.
2. Menganalisis ketepatan hasil rekomendasi yang dihasilkan berdasarkan banyaknya *training set dan test set*, jumlah *cluster*, serta ukuran *neighbor* dengan algoritma *ClustKN* yang diterapkan pada *collaborative filtering*..
3. Menganalisis pengaruh algoritma *ClustKNN* berdasarkan akurasi prediksi rating yang dihasilkan terhadap algoritma *user-based KNN*.

1.4 Metodologi Penyelesaian Masalah

Metodologi yang digunakan dalam memecahkan masalah di atas adalah dengan menggunakan langkah-langkah berikut:

1. Studi Literatur
   Pencarian referensi dan sumber-sumber yang berhubungan dengan *Recommender system*, *Collaborative filtering* dan algoritma ClustKNN dalam menyelesaikan tugas akhir ini.
2. Pengumpulan data
   Mengumpulkan data-data produk yang diperlukan sebagai *training set* dan *test set*.
3. Perancangan Sistem
   Analisis dan perancangan terhadap *recommender system* yang akan dibangun.
4. Implementasi Sistem
   Tahap pembangunan *recommender system* dengan melakukan pembangunan terhadap *database*, pembuatan algoritma, melakukan implementasi algoritma pada *collaborative filtering engine*, serta pembuatan antar muka untuk aplikasi *client*.
5. Pengujian Sistem
   Melakukan pengujian dan analisis terhadap perangkat lunak yang dibangun. Melakukan analisis dari implementasi sistem dan pengujian hasil. *Mean Absolute Error* (MAE) digunakan untuk menghitung akurasi dari rekomendasi yang dihasilkan.
6. Pengambilan Kesimpulan dan penyusunan laporan Tugas Akhir.
   Memberikan kesimpulan terhadap hasil analisis dan penyusunan laporan.

1.5 Sistematika Penulisan

Tugas akhir ini disusun dengan sistematika penulisan sebagai berikut:

BAB I **PENDAHULUAN**

Berisi pemaparan mengenai latar belakang permasalahan, tujuan yang

ingin dicapai dengan adanya penelitian ini, perumusan masalah, batasan masalah, metodologi tugas akhir, dan sistematika penulisan.

BAB II **LANDASAN TEORI**

Berisi uraian mengenai landasan teori yang akan digunakan, meliputi teori tentang algoritma *clustKNN* dan teori-teori lain yang berkaitan dengan penelitian tugas akhir ini

BAB III **ANALISIS DAN PERANCANGAN SISTEM**

Berisi tentang analisa dan perancangan terhadap *recommender system* yang akan dibangun.

BAB IV **ANALISIS DAN PENGUJIAN SISTEM**

Berisi implementasi dari hasil analisa dan perancangan sistem yang dibuat, serta pengujian sistem.

BAB V **KESIMPULAN DAN SARAN**

Berisi kesimpulan dan saran-saran untuk pengembangan lebih lanjut terhadap hasil penelitian ini.