

BAB I PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi dan informasi berkembang sangat cepat dengan banyaknya aplikasi atau *item* yang ada saat ini. Semakin banyaknya informasi produk yang ada di internet menghadirkan tantangan baik pembeli maupun pebisnis online dalam lingkungan *e-commerce*. Dengan pertumbuhan pengguna *online* dan koleksi *item* yang sangat banyak, telah menciptakan sebuah permintaan rekomendasi yang efisien untuk memandu *user* dalam memilih suatu *item* [1].

Sistem rekomendasi mempunyai peranan yang sangat penting untuk membantu memberikan referensi peringkat produk maupun *item* kepada *user*. Keakuratan rekomendasi sangat penting untuk *item* yang akan direkomendasikan kepada *user*. Jika perekomendasian *item* mengalami kesalahan dalam keakuratan, maka akan berdampak pada peringkat *item*, kesalahan rekomendasi, *cold-start*, tidak efisiennya perhitungan, dan sebagainya [2].

Cold-start merupakan masalah yang terjadi ketika sistem belum cukup memiliki preferensi untuk menghasilkan rekomendasi *item* baru yang efisien bagi *user* baru. Masalah *cold-start* yang kedua yaitu sistem basis data yang sangat terbatas atau tidak adanya preferensi tentang *item* pada pengguna [3].

Banyak peneliti yang telah membahas mengenai masalah *cold-start* dan menyelesaikannya dengan berbagai metode. Masalah *cold-start* tidak dapat diselesaikan hanya dengan metode tradisional (*item based*) [3]. Karena metode tradisional memiliki nilai keakuratan yang kurang baik [2]. Metode yang digunakan dalam menangani masalah *cold-start* adalah metode *enhanced rating based* [4], *regresi model* [2], *Collaborative error-reflected models* [19] dan sebagainya.

Metode tradisional (*Item Based*) tidak mampu mengatasi masalah *cold-start* [3]. Ada juga beberapa metode yang dilakukan untuk mengatasi masalah *cold-start*. Metode *enhanced rating based* mendapatkan hasil yang lebih efektif namun, untuk mendapatkan informasi dari media social akan lebih sulit [4]. Untuk metode *regresi model* [2] dan *Collaborative error-reflected models* [19] mendapatkan prediksi rating yang lebih baik dari metode pembandingan yang lain.

Dua metode ini tidak menggunakan *Clustering*, sehingga memprediksi rating data yang besar, akan memakan waktu yang cukup lama.

Metode *clustering* data telah terbukti dapat menyelesaikan masalah *cold-start* [5]. *Clustering user* dapat digunakan menggunakan atribut *user*. Metode *K-Modes clustering* merupakan metode untuk mengelompokkan nilai atribut kategori *user*. Metode ini mengelompokkan *user* berdasarkan atribut *user* seperti jenis kelamin, umur, dan pekerjaan [6]. *Fuzzy C-Means* menghasilkan pengelompokan yang lebih baik dibandingkan dengan *K-Means Clustering*. *Fuzzy C-Means* digunakan karena mengelompokkan nilai atribut kategori *numeric* pada *item* [7],[13]. Oleh karena itu, untuk pengelompokan *item* menggunakan *Fuzzy C-Means*. Sistem rekomendasi yang dibangun menggunakan teknik *hybrid popularitas model* berdasarkan hasil *cluster user* dan *cluster item* memberikan hasil rekomendasi *item* kepada *user* target dengan mengambil *item* yang sering dipilih oleh *user* lain atau *item* yang lebih populer diantara *item* lainnya. *User* target merupakan *user* baru yang tidak memiliki riwayat *rating* sebelumnya. Teknik ini dapat mengatasi permasalahan *cold-start* karena tidak memerlukan *rating* dari *item* yang dinilai sebelumnya oleh *user* target, sehingga rekomendasi *Top-N* berdasarkan *hybrid popularitas model* mendapatkan akurasi yang lebih baik [8]. *Rating* yang digunakan pada *hybrid popularitas model* adalah hasil dari filter *rating* yang dilakukan dengan menerapkan batas minimal *rating*. Misal batas minimal *rating* sama dengan 2, maka *rating* yang akan digunakan pada *hybrid popularitas model* adalah nilai *rating* 2, 3, 4 dan 5, sedangkan nilai *rating* sama dengan 1 diganti dengan nilai 0. Filter *rating* diperlukan dalam penelitian ini dikarenakan dalam menentukan *Top-N item* dipengaruhi oleh *rating item* yang telah diberikan oleh *user*.

Jadi untuk menyelesaikan permasalahan *cold-start*, penelitian ini mengkombinasikan penerapan metode *k-modes clustering*, *fuzzy c-means clustering* dan *hybrid popularitas model*.

1.2. Perumusan Masalah

Berdasarkan pada uraian latar belakang di atas maka dapat dipaparkan rumusan masalah sebagai berikut:

1.2.1. Permasalahan

Cold-start merupakan suatu masalah dimana sistem sulit untuk merekomendasikan suatu *item* kepada *user* baru karena *user* baru belum memberikan *rating* pada suatu *item*.

1.2.2. Usulan Solusi

Untuk mendapatkan kesamaan selera *user* maka dilakukan proses *Clustering user* yang mana akan menggunakan metode *K-modes clustering*. Untuk mendapatkan pengelompokan *item* maka dilakukan proses *Clustering item* yang mana menggunakan metode *Fuzzy C-Means clustering*. Dan untuk merekomendasikan *Top-N item* menggunakan metode *hybrid* popularitas model.

1.2.3. Pertanyaan penelitian

Dari permasalahan yang ada terdapat beberapa pertanyaan penelitian yaitu:

1. Apakah *filter rating* diperlukan dalam penerapan *hybrid* popularity model berdasarkan hasil *cluster user* menggunakan *K-Modes* dan *cluster item* menggunakan *Fuzzy C-Means* dalam mengatasi permasalahan *cold-start*?
2. Bagaimana akurasi terbaik dalam teknik *hybrid* popularity model dapat diperoleh berdasarkan hasil *cluster user* menggunakan *K-Modes* dan *cluster item* menggunakan *Fuzzy C-Means*?

1.3. Tujuan dan Manfaat

1.3.1. Tujuan Penelitian

Ada pun tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui apakah *filter rating* diperlukan dalam penerapan teknik popularitas *item* berdasarkan hasil *cluster user* menggunakan *K-Modes* dan *cluster item* menggunakan *Fuzzy C-Means* dalam mengatasi permasalahan *cold-start*.
2. Untuk mengetahui bagaimana akurasi terbaik dalam teknik *hybrid* popularity model dapat diperoleh berdasarkan hasil *cluster user* menggunakan *K-Modes* dan *cluster item* menggunakan *Fuzzy C-Means*.

1.3.2. Manfaat

Dihasilkan kombinasi metode yang dapat merekomendasikan *item* yang lebih akurat untuk *user* baru dibandingkan metode tradisional (*item based*). Sehingga dapat meningkatkan kemudahan bagi *user* baru.

1.4. Batasan Masalah

Pada proposal skripsi ini terdapat beberapa batasan masalah yang diambil yaitu:

1. Menggunakan metode *K-Modes clustering* untuk mengelompokkan *user* berdasarkan jenis kelamin, umur, dan pekerjaan. Dan menggunakan *Fuzzy C-Means Clustering* untuk mengelompokkan *item* berdasarkan nilai *rating*.
2. Menggunakan *hybrid* popularitas model untuk menentukan *top-N*.
3. *Dataset* yang digunakan adalah sebanyak 100.000 data *rating*, dari 943 *user* dan 1682 film. *Dataset* diambil dari website <https://grouplens.org/datasets/movielens/100k/>.

1.5. Sistematika Laporan

Sistematika penulisan laporan skripsi terdiri dari beberapa bab yaitu :

BAB I PENDAHULUAN

Pada bab ini membahas tentang Latar Belakang, Rumusan Masalah, Tujuan dan Manfaat, Batasan Masalah, dan Sistematika Proposal.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas tentang metode usulan yang dipilih dan penelitian sebelumnya.

BAB III METODE USULAN

Pada bab ini membahas tentang langkah-langkah penelitian, penjelasan arsitektur sistem, perhitungan manual, pengujian sistem, dan skenario uji coba.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisi tentang hasil implementasi sesuai arsitektur sistem dan hasil uji coba sesuai dengan skenario uji coba.

BAB V KESIMPULAN DAN SARAN

Pada bab ini berisi tentang kesimpulan dan saran dari seluruh penelitian yang telah dilakukan. Kesimpulan didapatkan dari penelitian berdasarkan pertanyaan penelitian dan saran berisi pengembangan penelitian selanjutnya.