

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di zaman modern ini dengan jumlah *user online* yang semakin banyak mendorong berbagai perusahaan digital untuk mengembangkan aplikasi mereka. Berbagai aplikasi seperti *social media*, *market place*, *movie*, *music* dan *game* [1]. Salah satu cara yang digunakan untuk menarik *user* yaitu dengan meningkatkan kemampuan aplikasi merekomendasikan konten atau item yang paling sesuai dengan minat *user* [1].

Sistem rekomendasi memiliki potensi untuk mempengaruhi konsumsi suatu produk, persepsi individu dan keputusan [2]. Sistem rekomendasi tradisional digunakan oleh beberapa *e-commerce* untuk menyelesaikan permasalahan personalisasi dengan menyarankan produk yang sesuai dengan minat *user* [3]. Rekomendasi produk kepada *user* target dibuat berdasarkan produk yang pernah dipilih sebelumnya oleh *user*, ketika *user* target merupakan *user* baru maka sistem tidak dapat merekomendasikan produk apa pun karena tidak ada riwayat pembelian produk sebelumnya [3].

Salah satu permasalahan pada sistem rekomendasi adalah *cold-start* yaitu permasalahan rekomendasi untuk *item* yang belum pernah dinilai sebelumnya [4]. Kondisi ini membuat sistem rekomendasi tidak dapat memprediksi minat *user* terhadap *item*, sehingga sulit untuk merekomendasikan *item* untuk target *user* karena hasil rekomendasi yang tidak sesuai dengan minat *user*. Permasalahan ini biasanya terjadi jika target *user* merupakan *user* baru yang belum pernah melakukan penilaian terhadap *item* sebelumnya.

Metode tradisional yang ada (*item based*) tidak mampu mengatasi masalah *cold-start* [3]. Karena *collaborative filtering* murni tidak dapat menyelesaikan permasalahan ini karena tidak ada informasi mengenai karakter penilaian *user* (*user preferences*) sebagai bentuk dasar untuk rekomendasi [4]. Terdapat beberapa cara untuk menangani permasalahan *cold-start*, salah satunya adalah dengan memanfaatkan *content information* untuk melihat kemiripan antara *item* baru

dengan *item* lain yang telah diberikan penilaian [4]. Menggunakan pendekatan informasi dari sosial media [5, 7]. Tetapi, untuk mendapatkan informasi dari sosial media akan lebih sulit. Pendekatan pembandingan *rating* (*Rating Comparison*) mendapatkan hasil rekomendasi yang baik [6], namun pada sisi efisiensi komputasi belum lebih baik karena belum menerapkan *clustering* sehingga untuk proses prediksi *rating* pada data yang besar akan membutuhkan waktu lama.

Metode *clustering* dapat diterapkan untuk mengatasi permasalahan *cold-start* [8, 9]. Salah satunya menerapkan metode *co-clustering* dengan mengelompokkan *user* dan *item* secara bersamaan. Metode *co-clustering* yang diterapkan menggunakan *SCOAL* (*Simultaneous Co-clustering and Learning*) menghasilkan nilai akurasi yang lebih baik [8]. Pengelompokan *user* berdasarkan atribut *user* dapat menggunakan metode *K-Modes clustering*. Metode *K-Modes clustering* digunakan untuk mengelompokkan data dengan nilai kategori seperti jenis kelamin, pekerjaan, alamat dan lainnya karena lebih *simple* dan efisien untuk data ukuran besar [10,11]. Sedangkan pengelompokan *item* dapat dilakukan menggunakan metode *Fuzzy C-Means clustering* untuk mengelompokkan data bertipe numerik. *Fuzzy C-Means clustering* telah terbukti memiliki hasil akurasi lebih baik dibandingkan dengan *K-Means clustering* [13,15].

Sistem rekomendasi yang dibangun menggunakan teknik popularitas *item* 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 popularitas *item* mendapatkan akurasi yang lebih baik [12,14]. *Rating* yang digunakan pada teknik popularitas *item* adalah hasil dari *filter rating* dengan menetapkan batas minimal *rating*. *Filter rating* dilakukan untuk mendapatkan spesifikasi nilai *rating* *item* pada hasil rekomendasi *Top-N*, misal batas minimal *rating* lebih besar sama dengan 3 maka *rating* yang akan digunakan pada teknik popularitas adalah nilai *rating* 3, 4 dan 5.

Penerapan teknik popularitas *item* pada hasil *clustering user* dan *item* menggunakan metode *K-Modes* dan *Fuzzy C-Means* pada penelitian ini bertujuan untuk mendapatkan hasil rekomendasi yang lebih baik daripada metode tradisional *item-based* dalam menangani masalah *cold-start*. Teknik popularitas yang akan digunakan terdiri dari tiga variasi yaitu teknik popularitas dengan menghitung hasil penjumlahan (*sum*), nilai tengah (*median*) dan selisih antara nilai maksimal dengan nilai minimal (*range*). Untuk mengetahui hasil rekomendasi yang lebih akurat maka kombinasi metode variasi teknik popularitas *item* dan *clustering* akan dibandingkan dengan metode tradisional *item-based*.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Metode tradisional yaitu *item-based* tidak dapat memberikan rekomendasi *item* kepada *user* baru yang belum pernah atau sedikit memberikan *rating item* (permasalahan *cold-start*).

1.2.2 Metode Usulan

Menerapkan variasi teknik popularitas *item* untuk mendapatkan rekomendasi *Top-N item* berdasarkan hasil *clustering user* dan *item*. Proses *clustering user* dilakukan dengan menerapkan metode *K-Modes*. Proses *clustering item* dilakukan dengan menerapkan metode *Fuzzy C-Means*.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan yang ada maka pertanyaan penelitian pada penelitian ini adalah:

1. 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. Variasi teknik popularitas *item* hasil *cluster user* menggunakan *K-Modes* dan *cluster item* menggunakan *Fuzzy C-Means* manakah yang memiliki akurasi rekomendasi *Top-N* terbaik dalam mengatasi permasalahan *cold-start* ?

3. Bagaimana perbandingan akurasi dari metode yang menerapkan variasi (terbaik) teknik popularitas *item* berdasarkan hasil *cluster K-Modes* dan *Fuzzy C-Means* dengan akurasi metode tradisional (*item-based*) ?

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 variasi teknik popularitas *item* hasil *cluster user* menggunakan *K-Modes* dan *cluster item* menggunakan *Fuzzy C-Means* manakah yang memiliki akurasi rekomendasi *Top-N* terbaik dalam mengatasi permasalahan *cold-start*.
3. Untuk mengetahui perbandingan akurasi metode yang menerapkan variasi (terbaik) teknik popularitas *item* berdasarkan hasil *cluster K-Modes* dan *Fuzzy C-Means* dengan akurasi metode tradisional (*item-based*).

1.3.2 Manfaat

Hasil penelitian ini diharapkan dapat menghasilkan sistem rekomendasi yang mampu merekomendasi *item* dengan akurasi yang lebih baik kepada *user* baru.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Menggunakan metode *K-Modes clustering* untuk mengelompokkan *user* berdasarkan jenis kelamin, umur dan pekerjaan.
2. Menggunakan metode *Fuzzy C-Means* untuk mengelompokkan *item* berdasarkan *rating*.
3. Menggunakan variasi popularitas *item* untuk menentukan *Top-N*.
4. Menggunakan 100.000 data *rating* dari 943 *user* dan 1682 film sebagai *dataset*. Dataset diperoleh dari situs <https://grouplens.org/datasets/movielens/100k/>.

1.5 Sistematika Laporan

Sistematika penulisan laporan skripsi ini adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini akan dibahas mengenai Latar Belakang, Rumusan Masalah, Tujuan dan Manfaat, Batasan Masalah, dan Sistematika Laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini akan dibahas mengenai metode usulan yang dipilih, penelitian sebelumnya, dan contoh penggunaan metode.

BAB III METODE USULAN

Pada bab ini akan dibahas mengenai algoritma metode usulan, penjelasan arsitektur sistem, *dataset* yang digunakan, metode penelitian, dan perkiraan jadwal penelitian.

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, kesimpulan didapatkan dari hasil uji coba yang telah dilakukan dan saran untuk pengembangan penelitian terkait selanjutnya.