

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di zaman sekarang ini, teknologi aplikasi sangat cepat berkembang. Mulai dari aplikasi sederhana sampai dengan aplikasi yang sangat kompleks. Banyak aplikasi *online* dan *e-commercial services*, seperti jejaring sosial, rekomendasi produk, yakni film, musik, dan artikel [1]. Banyak perusahaan yang mengembangkan aplikasinya dengan berbagai macam fitur yang berguna untuk menarik perhatian *user*. Salah satu fitur aplikasinya adalah rekomendasi. Sistem Rekomendasi sendiri merupakan sistem untuk menyarankan sebuah konten atau *item* yang sesuai dengan minat *user* [2].

Sistem rekomendasi merekomendasikan *item* sesuai minat *user* untuk kenyamanan pengguna atau *user* dan juga untuk kepentingan bisnis pemilik aplikasi [2]. Sehingga hal tersebut memperbesar peluang untuk *user* dalam memilih *item* yang sesuai [2]. Sistem rekomendasi memiliki beberapa masalah seperti keakuratan dalam merekomendasikan *item*, *cold-start*, efisiensi proses perhitungan, dan sebagainya [3].

Berdasarkan masalah yang ada, yang paling menantang adalah *cold-start* [2]. *Cold-start* merupakan kondisi dimana *user* baru belum memiliki riwayat pencarian, memberi *rating*, atau aktifitas lainnya terhadap *item* atau sebaliknya [3]. Kondisi tersebut membuat sistem rekomendasi tidak dapat memprediksi minat *user* terhadap *item*, sehingga kejadian tersebut membuat sistem sulit untuk merekomendasikan *item* pada *user* baru. Hasilnya, rekomendasi yang tidak sesuai dengan minat *user* berdampak pada kenyamanan *user* saat menggunakan aplikasi.

Metode tradisional yang ada (*Item Based*) tidak mampu mengatasi masalah *cold-start* [4]. Banyak peneliti yang membahas *cold-start* dengan berbagai metode yang ditawarkan. Metode yang sudah ditawarkan memiliki berbagai kelebihan untuk mengatasi *cold-start*. Salah satu teknik yang digunakan adalah wawancara [5]. Teknik wawancara merupakan pertanyaan singkat yang diajukan saat *user* baru menggunakan aplikasi dengan tujuan untuk mengetahui minat *user* pada kategori *item* tertentu. Metode wawancara sangat tidak efisien dan membuat

beban yang berat pada sistem. Walaupun teknik ini mampu menjamin keakuratan yang cukup baik. Namun, metode ini tidak layak pada kebanyakan sistem rekomendasi yang sebenarnya [5]. Metode lain yang digunakan adalah pada pendekatan perbandingan *rating*, mendapatkan hasil yang lebih baik dalam hal akurasi [5]. Namun pada sisi efisiensi komputasi belum lebih baik. Menggunakan pendekatan informasi dari sosial media memberikan hasil rekomendasi yang lebih akurat [6]. Namun, untuk mendapatkan informasi dari sosial media lebih sulit. Pendekatan dengan algoritma *Monte Carlo* mendapatkan prediksi *rating* terbaik dari metode perbandingan lain pada kondisi *cold-start* [4]. Dan pendekatan *Neural Semantic Personalized Ranking* (NSPR) mampu mengungguli metode lain pada dua *dataset* asli sistem rekomendasi [7]. Untuk kedua pendekatan yakni *Monte Carlo* dan NSPR tidak menggunakan *clustering*, sehingga untuk memprediksi *rating* untuk data yang besar, memakan waktu lama karena menghitung keseluruhan *item*. Untuk mengatasi prediksi *rating* data yang besar, dilakukan proses *clustering* data berdasarkan *user* dan *item* untuk mendapatkan akurasi yang lebih baik dan tidak memakan waktu lama, karena tidak menghitung keseluruhan *item*.

Metode *clustering* data, telah terbukti dapat digunakan untuk mengatasi *cold-start* [8,9]. *Clustering* dua sisi (*user* dan *item*) yang diterapkan menggunakan *Simultaneous Co-clustering And Learning* (SCOAL) memperoleh hasil akurasi yang bagus [8]. Pengelompokan dua sisi memiliki hasil lebih baik dibandingkan *clustering* satu sisi [8,10]. SCOAL menggunakan *clustering* dengan nilai *rating*. *clustering* dengan menggunakan *rating* tidak efektif pada *user* yang tidak pernah memberi nilai *rating* atau data *sparse* [11]. Alternatif lain, *clustering user* dapat dilakukan dengan menggunakan atribut *user*. Metode *K-Approximate Modal Haplotypes* (K-AMH) merupakan salah satu alternatif untuk mengelompokkan data dengan nilai atribut kategori (bukan numerik/angka) seperti jenis kelamin, umur, dan pekerjaan. Metode ini sangat efisien untuk data yang mirip atau sama [12]. Pengelompokan *user* berdasarkan atribut dapat menggunakan metode K-AMH. Sedangkan pengelompokan *item* menggunakan metode *Fuzzy C Means* untuk mengelompokkan data bertipe numerik, karena telah dibuktikan pada beberapa penelitian dapat memberikan hasil yang lebih baik dibandingkan dengan

metode tradisional (*K-Means clustering*) [9]. Sistem rekomendasi yang dibangun dari hasil *cluster user* dan *cluster item* akan dikombinasikan dengan teknik popularitas *item*. Teknik popularitas *item* adalah teknik yang digunakan untuk merekomendasikan *item* dengan mengambil *item* yang sering dipilih oleh *user* lain atau *item* yang lebih populer diantara *item* lainnya, teknik ini dapat mengatasi permasalahan *cold-start* karena tidak memerlukan *rating* dari *item* yang dipilih sebelumnya oleh *user* target, sehingga, untuk melakukan rekomendasi *Top-N* dilakukan proses perhitungan berdasarkan popularitas *item* untuk mendapatkan akurasi yang lebih baik [9], [13].

Jadi penelitian ini mengimplementasikan kombinasi *clustering* metode *K-AMH* untuk *user* dan *Fuzzy C Means* untuk *item*, serta teknik popularitas *item* untuk melakukan rekomendasi *Top-N* untuk mengatasi masalah *Cold-Start* pada sistem rekomendasi.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Sistem rekomendasi tidak dapat memberikan rekomendasi *item* kepada *user* baru yang belum pernah atau sedikit memberikan *rating item* (*cold-start*).

1.2.2 Metode Usulan

Menerapkan *clustering* dua sisi pada *user* dan *item* serta teknik popularitas *item* untuk mendapatkan rekomendasi *Top-N item* berdasarkan hasil *clustering user* dan *item*. Pada proses *clustering*, digunakan metode *K-AMH* untuk mengelompokkan *user* dan menggunakan metode *Fuzzy C Means* untuk mengelompokkan *item*.

1.2.3 Pertanyaan Penelitian

Bagaimana akurasi terbaik dari implementasi kombinasi metode *Clustering K-AMH* dan *Fuzzy C Means* serta teknik popularitas *item* untuk mengatasi masalah *cold-start* pada sistem rekomendasi dapat diperoleh?

1.3 Tujuan dan Manfaat

1.3.2 Tujuan Penelitian

Untuk mengetahui bagaimana memperoleh akurasi terbaik dari implementasi kombinasi metode *Clustering K-AMH* dan *Fuzzy C Means* serta teknik popularitas *item* untuk mengatasi masalah *Cold-Start* pada sistem rekomendasi.

1.3.3 Manfaat Penelitian

Dihasilkan kombinasi metode yang dapat merekomendasikan *item* yang lebih akurat untuk *user* baru. Sehingga dapat meningkatkan kemudahan bagi *user* baru.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

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

1.5. Sistematika Laporan

Sistematika laporan ini adalah sebagai berikut :

BAB I PENDAHULUAN

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

BAB II KAJIAN PUSTAKA

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

BAB III METODE USULAN

Pada bab ini dibahas mengenai penjelasan arsitektur sistem, *dataset* yang digunakan, perhitungan metode penelitian, evaluasi dan skenario pengujian.

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.