

BAB I PENDAHULUAN

1.1 Latar Belakang

Di zaman modern ini, aplikasi sangat cepat berkembang dengan *user* yang mencapai milyaran [1]. Mulai dari aplikasi sederhana sampai dengan aplikasi yang sangat kompleks. Beberapa aplikasi *online* yang paling sering digunakan saat ini adalah media sosial, *market place*, *game*, dan berbagai aplikasi *online* lainnya. Banyak perusahaan mengembangkan aplikasi nya dengan berbagai macam fitur yang berguna untuk menarik *user*. Salah satu fitur aplikasi yang cukup berperan penting dalam sebuah aplikasi adalah rekomendasi. Sistem Rekomendasi merupakan sistem untuk menyarankan sebuah konten sesuai minat *user* [1].

Sistem rekomendasi sangat besar perannya untuk mendapatkan keuntungan yang diinginkan [2]. Baik keuntungan untuk kenyamanan *user* maupun untuk kepentingan bisnis pemilik aplikasi karena sistem rekomendasi merekomendasikan *item* sesuai minat *user*, sehingga hal tersebut memperbesar peluang *user* untuk memilih item yang direkomendasikan. Sistem rekomendasi memiliki beberapa masalah seperti keakuratan dalam merekomendasikan *item*, *Cold-start*, efisiennya proses perhitungan, dan sebagainya. Banyak peneliti masih membahas masalah-masalah tersebut untuk mendapatkan hasil yang lebih baik dalam merekomendasikan *item*.

Berdasarkan masalah yang ada, yang paling menantang adalah *Cold-start* [2]. *Cold-start* merupakan kondisi dimana *user* baru belum memiliki riwayat pencarian [3], penilaian, atau aktifitas lainnya terhadap *item* atau sebaliknya. Kondisi tersebut membuat sistem rekomendasi tidak dapat memprediksi minat *user* terhadap *item*, sehingga kejadian tersebut membuat sistem kesulitan untuk merekomendasikan *item* pada *user* baru. Hasil rekomendasi yang tidak sesuai dengan minat *user* akan berdampak pada kenyamanan *user* saat menggunakan aplikasi. Masalah lebih lanjut dapat mengurangi pemasukan pihak terkait.

Masalah *cold-start* ada pada ketidak mampuan metode tradisional yang ada (Item Based) mengatasi *cold-start* [4]. sudah banyak peneliti yang membahas *cold-start* dengan berbagai metode yang ditawarkan. Metode yang sudah ditawarkan memiliki berbagai kelebihan untuk mengatasai *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. Teknik wawancara menggunakan beberapa pertanyaan singkat kepada *user* baru tentang kesukaannya. Metode lain yang digunakan adalah metode informasi media sosial [6], [7], membandingkan *rating* [5] , personalisasi peringkat dengan *neural network* [8], algoritma monte carlo [4], menggunakan teknik *clustering* [3], [9], dan masih banyak lagi.

Metode tradisional yang ada (*Item Based*) tidak mampu mengatasi masalah *cold-start* [4]. Metode wawancara sangat tidak efisien dan membuat beban yang cukup berat pada sistem. Walaupun teknik semacam ini mampu menjanjikan keakuratan yang cukup baik. Namun, metode ini tidak layak pada kebanyakan sistem rekomendasi yang sebenarnya [5]. Pada pendekatan perbandingan *rating*, mendapatkan hasil yang lebih baik dalam hal akurasi [5]. Namun pada sisi efisiensi komputasi belum lebih baik dari metode perbandingan. Menggunakan pendekatan informasi dari sosial media memberikan hasil rekomendasi yang lebih akurat [7]. Namun, untuk mendapatkan informasi dari sosial media akan lebih sulit. Pendekatan dengan algoritma *Monte carlo* mendapatkan prediksi rating terbaik dari metode perbandingan lain pada kondisi *cold-start* dan *warm start* [4]. Dan personalisasi peringkat sistem saraf mampu mengungguli metode lain pada dua *dataset* asli sistem rekomendasi [8]. Berdasarkan kedua pendekatan ini tidak menggunakan *Clustering*, sehingga untuk memprediksi rating untuk data yang besar, akan memakan waktu lama karena menghitung keseluruhan item.

Metode *clustering* data, telah terbukti dapat digunakan untuk mengatasi *cold-start* [3], [9]. *Clustering* dua sisi (*user* dan *item*) yang diterapkan menggunakan *SCOAL* menghasilkan hasil akurasi yang bagus [3]. Pengelompokkan dua sisi memiliki hasil lebih baik dibandingkan *Clustering* satu

sisi [3], [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-modes clustering* merupakan metode khusus untuk mengelompokkan data dengan nilai atribut kategori (bukan numerik/angka) dan metode ini sangat efisien untuk data sekala besar [12], [13]. Oleh karena itu, untuk mengelompokkan *user*, akan digunakan metode *K-modes clustering*. Metode ini akan mengelompokkan *user* berdasarkan atribut *user* (Jenis kelamin, umur, pekerjaan, dan kode pos). *Fuzzy C-Means Clustering* telah dibuktikan pada beberapa penelitian dapat memberikan hasil yang lebih baik dibandingkan dengan metode tradisional (*K-Means clustering*) [9], [14]. Oleh karena itu, pengelompokkan *item* akan digunakan metode *Fuzzy C-Means Clustering*. Rekomendasi *Top-N* merupakan aspek penting dalam penelitian ini. Sehingga, untuk melakukan rekomendasi *Top-N* akan dilakukan proses perhitungan berdasarkan popularitas *item* untuk mendapatkan akurasi yang lebih baik [9], [15].

Penggunaan kombinasi *Clustering* dua sisi menggunakan metode *K-modes clustering* untuk *user* dan *Fuzzy C-Means clustering* untuk *item* serta teknik popularitas *item* pada penelitian ini, diharapkan mampu mendapatkan hasil dengan akurasi yang lebih baik dibandingkan metode tradisional yaitu *item-based (IB)*. Hasil kombinasi *Clustering* menggunakan *K-modes clustering* dan *Fuzzy C-Means clustering* serta penggunaan teknik popularitas *item* pada penelitian ini akan dibandingkan dengan hasil dari metode tradisional yaitu *item-based (IB)* untuk mengetahui dari metode mana hasil rekomendasi yang lebih akurat akan didapatkan.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Cold-Start merupakan masalah yang terjadi ketika *user* baru yang belum atau masih sedikit menilai *item*. Metode tradisional tidak efektif mengatasi hal tersebut (*item based*).

1.2.2 Metode Usulan

Untuk mendapatkan kesamaan selera *user* maka akan dilakukan proses *Clustering user*. Dan untuk proses *Clustering*, akan digunakan metode *K-modes clustering* untuk mengelompokkan *user* dan menggunakan metode *Fuzzy C-Means clustering* untuk mengelompokkan *item*. Yang kemudian akan merekomendasikan *Top-N item* berdasarkan popularitas *item*.

1.2.3 Pertanyaan Penelitian

1. Berapa akurasi rekomendasi pada implementasi kombinasi metode *Clustering K-Modes* dan *Fuzzy C-Means* serta teknik popularitas *item* untuk mengatasi masalah *cold-start* pada sistem rekomendasi?
2. Apakah akurasi rekomendasi pada implementasi kombinasi metode *Clustering K-Modes* dan *Fuzzy C-Means* serta teknik popularitas *item* lebih baik dari metode tradisional (*item based*) ?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

1. Untuk mengetahui akurasi rekomendasi pada implementasi kombinasi metode *Clustering K-Modes* dan *Fuzzy C-Means* serta teknik popularitas *item* untuk mengatasi masalah *cold-start* pada sistem rekomendasi.
2. Untuk mengetahui apakah akurasi rekomendasi pada implementasi kombinasi metode *Clustering K-Modes* dan *Fuzzy C-Means* serta teknik popularitas *item* lebih baik dibandingkan metode tradisional (*item based*).

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

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Menggunakan metode *K-Modes clustering* untuk mengelompokkan *user* berdasarkan jenis kelamin, umur, pekerjaan, dan kode pos. 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 sebanyak 100.000 data *rating*, dari 943 user dan 1682 film.

1.5 Sistematika Proposal

Sistematika laporan ini adalah sebagai berikut :

1. BAB I PENDAHULUAN

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

2. BAB II KAJIAN PUSTAKA

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

3. 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.