

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi internet dan informasi yang cepat telah menghasilkan pertumbuhan yang sangat meningkat dari informasi yang tersedia selama beberapa tahun terakhir [1]. Sistem rekomendasi adalah sistem yang memprediksi *rating* atau preferensi yang akan diberikan *user* ke suatu *item* [2]. Sistem rekomendasi secara otomatis dapat menganalisis penggunaan data *user* untuk menyaring konten halaman *web*, mengkategorisasi pesan *newsgroup*, dan merekomendasikan informasi. Sistem rekomendasi menganalisis data mengenai produk atau interaksi *user* dan *item* untuk menemukan hubungan antara *item* dan *user*. Hasil yang diterima akan ditampilkan sebagai rekomendasi [3]. Tujuan dari sistem Rekomendasi adalah untuk secara otomatis menghasilkan *item* yang disarankan (film, buku, berita, musik, CD, DVD, halaman *web*) untuk *user* sesuai dengan *rating* historis mereka guna menghemat waktu pencarian mereka secara *online* dengan mengumpulkan data yang bermanfaat [1].

Rekomendasi film adalah aplikasi yang paling banyak digunakan ditambah dengan *platform* multimedia *online* yang bertujuan untuk membantu *user* mengakses film pilihan secara cerdas dari perpustakaan film besar. Banyak pekerjaan telah dilakukan baik di bidang akademik dan industri dalam mengembangkan algoritma dan ekstensi rekomendasi film baru [1].

Beberapa pendekatan metode tradisional sistem rekomendasi yaitu metode *Collaborative Filtering (CF)* dan *Content-Based (CB)* [4][5][6]. Meskipun kedua metode memiliki kelebihan, mereka juga memiliki kelemahan tertentu pada masing-masing setiap metode [6][7]. Secara umum pada pendekatan metode *CF* melaporkan kinerja yang lebih baik daripada pendekatan *CB*, tetapi keberhasilannya bergantung pada jumlah *rating user* yang mencukupi. Sistem seperti itu memiliki kekurangan yaitu keterbatasan membuat rekomendasi yang akurat untuk *user* dengan informasi yang sangat sedikit (*sparsity*) [8]. Masalah ini muncul karena analisis kesamaan tidak cukup akurat. Dalam situasi ini penggunaan pendekatan *CB* muncul sebagai alternatif. Namun demikian, pendekatan ini

memiliki keterbatasannya sendiri. Misalnya, kata kunci yang digunakan untuk mewakili konten *item* mungkin tidak terlalu representatif [8]. Maka dari itu terciptalah metode *Hybrid* untuk mengatasi kekurangan pada kedua metode [9].

Salah satu algoritma yang sering digunakan pada metode *CF* ini yaitu *Item-based* dan *User-based*. *Item-based CF* mempunyai akurasi yang lebih baik dibandingkan *User-based CF* [10][11]. Ada dua proses utama pada *Item-based CF*, yaitu menghitung *item-item similarity* dan menghitung prediksi [10].

Untuk pemodelan metode *CB* dapat menggunakan ketersediaan konten (sering juga disebut dengan fitur, atribut atau karakteristik) sebuah *item* sebagai basis dalam pemberian rekomendasi [7]. Untuk proses ekstraksi fitur, atribut atau karakteristik ini dilakukan dengan menggunakan teknik *text mining* [12]. Setelah itu untuk proses *clustering* item menggunakan algoritma *K-Means Clustering* [12].

Berdasarkan permasalahan diatas solusi yang ditawarkan dengan menutupi kekurangan kedua metode, maka pada penelitian ini dilakukanlah metode *Hybrid*(*CB* dan *CF item-based*) yang dimana metode *text mining* sebagai *text preprocessing* nya. Begitu juga algoritma *K-Means* sebagai *clustering* karena metode ini telah dikenal sangat efisien dan diperhitungkan dalam melakukan *clustering*.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Permasalahan yang sering terjadi dalam sistem rekomendasi ialah keterbatasan informasi (*sparsity*) serta pada metode *CB* dan *CF* memiliki kelemahan masing-masing-masing dalam menghasilkan suatu rekomendasi.

1.2.2 Metode Usulan

Penelitian ini mengusulkan metode *Hybrid (Content Based dan Collaborative filtering item-based)* untuk membangun sebuah sistem rekomendasi.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan dan metode usulan maka pertanyaan penelitian ini adalah berapa hasil evaluasi dari metode *Hybrid (CB dan CF item-based)* untuk sistem rekomendasi film?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Mengacu pada latar belakang serta rumusan masalah maka tujuan dari penelitian ini untuk mengetahui evaluasi dari metode *Hybrid (CB dan CF item-based)* untuk sistem rekomendasi film.

1.3.2 Manfaat Penelitian

Hasil penelitian ini diharapkan bisa menjadi alternatif atau rujukan dalam membangun suatu sistem rekomendasi yang memiliki masalah *sparsity* data dan mengatasi kelemahan pada metode *CB* dan *CF Item-Based* dalam menghasilkan suatu rekomendasi.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah dataset yang digunakan yaitu data *rating* MovieLens 100k dari kaggle (<https://www.kaggle.com/rajmehra03/movielens100k>) dan untuk data judul dan sinopsis semua film dari IMDB (<https://www.imdb.com/>).

1.5. Sistematika Proposal

Sistematika penulisan Proposal ini terdiri dari beberapa bab, yaitu:

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, serta sistematika penulisan laporan sistem rekomendasi film dengan metode *Hybrid(Content Based dan Collaborative Filtering Item-Based)*.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas tentang landasan teori *Content Based* dan *Collaborative Filtering* dan penelitian sebelumnya.

BAB III METODE USULAN

Pada bab ini membahas tentang metode untuk sistem rekomendasi yang diusulkan seperti arsitektur sistem (*Input, Proses, Output*), data penelitian, tahapan penelitian, serta rencana pengujian.

REFERENSI

Berisi tentang daftar referensi yang dijadikan acuan dalam penelitian sistem rekomendasi film dengan metode *Hybrid(Content Based dan Collaborative Filtering Item-Based)*.