

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Film sudah menjadi salah satu media hiburan yang populer dikalangan masyarakat. Sejak tahun 1874 sampai 2015, sebanyak 3,361,741 judul film telah dikeluarkan oleh industri perfilman (imdb.com, 2016). Banyaknya judul-judul film yang telah beredar membuat masyarakat sulit untuk menemukan film yang mereka inginkan. Data-data *rating* film yang terdapat dalam suatu *website* dapat diolah dan dimanfaatkan untuk merekomendasikan film kepada *user* lain. Pertimbangannya adalah menemukan film berdasarkan hubungan antara satu film dan film lainnya yang sudah diberi *rating* oleh *user* untuk dijadikan rekomendasi kepada *user* lain. Oleh karena itu, diperlukan suatu sistem yang dapat merekomendasikan film kepada *user*.

Sistem rekomendasi adalah suatu mekanisme yang dapat memberikan suatu informasi atau rekomendasi sesuai dengan kesukaan *user* berdasarkan informasi yang diperoleh dari *user* (Karypis, 2001). Oleh karena itu, diperlukan model rekomendasi yang tepat agar rekomendasi yang diberikan oleh sistem sesuai dengan kesukaan *user*, serta mempermudah *user* mengambil keputusan dalam menentukan item (film) yang akan dipilih (McGinty, 2006). Salah satu metode rekomendasi yang digunakan dalam sistem rekomendasi adalah *Collaborative filtering*. *Collaborative filtering* menghubungkan setiap *user* dengan kesukaan yang sama terhadap suatu item (film) berdasarkan *rating* yang diberikan *user*. Untuk meningkatkan keakurasian hubungan antara *user* dengan kesukaan yang sama terhadap suatu item (film) digunakan algoritma *clustering* (Gupta,2009).

*Clustering* adalah mengelompokkan item data kedalam sejumlah kecil grup sedemikian sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial (Garcia-Molina, Hector., 2002). Pada penelitian Smrity Gupta (2009) telah mencoba menggabungkan *Collaborative filtering* dengan algoritma *k-means* yang menghasilkan sistem rekomendasi yang efisien untuk pemrosesan data dalam jumlah besar dan akurasi yang tinggi. *Bisecting K-Means* merupakan algoritma yang lebih baik dibandingkan algoritma *K-Means* karena memproduksi *cluster* yang seragam dan tidak memproduksi

*cluster* kosong, tingkat keakurasian yang baik dan lebih efisien ketika jumlah *cluster* meningkat (Patil,R, Khaan,2015). Pada penelitian ini, peneliti mencoba mengkolaborasi *Collaborative filtering* dengan *Bisecting K-Means* dengan tujuan menghasilkan rekomendasi yang lebih akurat menggunakan pengelompokan (*clustering*) yang lebih baik dari *K-Means*. Berdasarkan uraian di atas, maka akan diangkat penelitian “**Sistem Rekomendasi Film menggunakan Collaborative Filtering dan Bisecting K-Means**” sebagai judul tugas akhir.

## 1.2 Rumusan Masalah

Berdasarkan pada penjelasan latar belakang sebelumnya, rumusan masalah pada tugas akhir ini adalah: banyaknya judul-judul film yang beredar maka diperlukan teknik pengelompokan terhadap judul film dan pada penelitian sebelumnya hasil algoritma *K-Means* tergantung pada *centroid* awal yang acak, *cluster* yang dibentuk tidak seimbang dan tidak sama ukurannya dan mungkin dapat menghasilkan *cluster* yang kosong.

## 1.3 Tujuan dan Manfaat

Tujuan dari penulisan tugas akhir ini adalah mengembangkan website untuk merekomendasikan film dengan menerapkan algoritma *Collaborative Filtering* dan *Bisecting K-Means* berdasarkan nilai *rating* dari *user*.

Manfaat dari penulisan tugas akhir ini adalah sebagai berikut:

1. Membantu *user* untuk mendapatkan rekomendasi film yang sesuai dengan kesukaan *user*.
2. Menghasilkan tingkat rekomendasi yang lebih baik keakurasiannya dari pada sistem yang sebelumnya.
3. Sistem rekomendasi dapat digunakan sebagai bahan referensi untuk penelitian dan pengembangan lebih lanjut di bidang yang berkaitan.

## 1.4 Batasan Masalah

Adapun batasan masalah dalam tugas akhir ini adalah:

1. Hasil rekomendasi berupa daftar film yang sudah dihitung nilai *similarity* dan nilai prediksinya menggunakan *Collaborative Filtering*.
2. Dataset yang digunakan adalah MovieLens<sup>1</sup> dengan karakteristik berupa 100 ribu *rating* dari 668 *user* untuk 10.329 judul film, dimana film dikelompokkan berdasarkan 18 *genre* (*drama, comedy, short, dokumentary, talk-show, family, news, romance, animation, music, reality –Tv, crime, action, game-Show, adventure, thrailler, mystery, fantasy, sci-fi, adult, sport, horror, history, biography, western, war, film noir*). *User* dapat memberikan *rating* (skala 1-5) dengan rincian skala 1 adalah yang paling buruk dan skala 5 adalah yang paling bagus, terhadap film jika telah melakukan registrasi.

### 1.5 Metodologi Pengembangan Sistem

Pada sistem ini pengembangan dilakukan dengan menggunakan metode *waterfall*. Tahapan-tahapan yang dilakukan dalam sistem rekomendasi ini adalah sebagai berikut:

1. Studi Pustaka
  - a. Melakukan kajian literatur mengenai metode rekomendasi dari buku, jurnal dan data-data dari internet.
  - b. Melakukan pengumpulan data berupa nilai *rating*, judul film dan *user* dari *website* MovieLens<sup>1</sup>.
2. Analisis

Tahap ini menganalisis kebutuhan yang diperlukan sistem baik secara fungsional menggunakan UML *activity diagram* dan non-fungsional menggunakan kerangka PIECES.

- a. Desain

Tahap ini merancang tampilan antarmuka *website* rekomendasi *film* dengan menggunakan *tools balsamiq* dan merancang desain data menggunakan *Entity Relationship Diagram (ERD)*.

- b. Implementasi

Tahap ini mengimplementasikan rancangan tampilan antarmuka menggunakan bahasa pemrograman PHP, HTML dan MySQL.

### c. Pengujian

Pengujian dilakukan dengan data set dari MovieLens<sup>1</sup>, yaitu data berisi 100,000 *rating* (skala 1-5) , 10,329 film dengan 668 *users*. Untuk menguji sistem dilakukan dengan menguji hasil rekomendasi berdasarkan prediksi *rating*. Pengujian akan dihitung tingkat keakurasiannya menggunakan *Mean Absolute Error* (MAE) berdasarkan parameter *Neighbourhood Size* (NS), kemudian dilakukan analisis hasil dan penarikan kesimpulan. Untuk menguji sistem dilakukan dengan melakukan penyebaran kuisioner terhadap 20 orang responden dengan rentan umur 10 tahun sampai diatas 34 tahun, pertanyaan terdiri atas 10 pertanyaan berupa *linkedlist* dan 1 pertanyaan berupa saran terhadap tampilan dan fitur dari sistem. Penyebaran kuisioner dilakukan melalui *google form* kemudian tanggapan responden dibuat secara grafik untuk mengetahui tingkat kepuasan dan penggunaan terhadap sistem.

UNIVERSITAS  
MIKROSKIL