

BAB I

PENDAHULUAN

1.1. Latar Belakang

Zaman digital saat ini memberikan segala bentuk informasi yang *user* butuhkan dalam satu media, yakni internet. Internet menyediakan berbagai informasi untuk penggunaannya. Hal ini menyebabkan *information overload*, dimana dibutuhkan sebuah sistem yang dapat memberikan saran berupa rekomendasi sesuai dengan kebutuhan dan harapan *user*. Dalam memberikan sebuah rekomendasi, membutuhkan sebuah model yang tepat agar mendapatkan hasil yang sesuai dengan kebutuhan *user*.

Sistem rekomendasi merupakan sebuah sistem yang secara garis besar dibuat untuk memberikan informasi, berupa rekomendasi kepada *user*. Rekomendasi yang diberikan berhubungan dengan pengambilan keputusan seperti, *film* apa yang akan dilihat, makanan apa yang akan dimakan, buku apa yang akan dibaca, dan musik apa yang akan didengar [1]. Pada hakikatnya, sistem rekomendasi bekerja dengan memprediksi nilai suatu *item* yang kemudian akan di proses menggunakan pendekatan yang sesuai dengan kebutuhan untuk menghasilkan sebuah rekomendasi yang akan di berikan kepada *user*.

Dalam sistem rekomendasi terdapat dua pendekatan utama yakni *Content-based Filtering* (CB) dan *Collaborative Filtering* (CF). Pendekatan dalam sistem rekomendasi merupakan sebuah model yang digunakan dalam memberikan sebuah rekomendasi. CF merupakan pendekatan tradisional yang umum digunakan dalam sistem rekomendasi. Dengan memprediksi nilai suatu *item* yang kemudian diberikan sebagai rekomendasi pada *user* berdasarkan *similaritas user* dengan *user* lainnya atau *similaritas item* dengan *item* lainnya [2].

Dalam pendekatan CF dibagi menjadi dua model, yaitu *item-based* CF dan *user-based* CF [3]. *Item-based* CF merekomendasikan *item* berdasarkan *similaritas ranking* yang diberikan *user target* kepada *item neighbourhood*. *User-based* memberikan rekomendasi berdasarkan hasil prediksi *similaritas user target* dengan *user neighbourhood* berdasarkan pemberian nilai pada suatu *item* atau produk.

Pada penelitian sebelumnya menyebutkan jika pendekatan *item-based* CF memberikan hasil rekomendasi lebih baik yang baik sesuai dengan kebutuhan *user* dibandingkan dengan model *user-based* CF [4].

Namun, penerapan pendekatan tradisional *item-based* CF umumnya tidak memanfaatkan atribut *item* dalam perhitungan hasil rekomendasinya, sehingga menghasilkan rekomendasi yang kurang baik. Dalam hal ini, penerapan pendekatan tradisional *item-based* CF mengabaikan peran penting *genre* sebagai atribut item rekomendasi yang telah terbukti dapat memperbaiki hasil dari rekomendasi[5].

Pemanfaatan *genre* dalam metode sistem rekomendasi dapat dilakukan dengan dua acara, yang pertama adalah dengan menggunakan satu tahap permodelan. Akan tetapi cara ini menghasilkan kualitas rekomendasi tidak terlalu *signifikan*. Cara yang kedua dengan dua tahap permodelan, yaitu tahap *candidate* rekomendasi dan tahap *re-ordering*. Ada beberapa model pendekatan rekomendasi yang dapat dilakukan. Misal, tahap *candidate* rekomendasi menggunakan pendekatan *collaborative filtering*, dan *re ordering* dilakukan *content-based* atau *naive bayes*. Dalam *re-ordering* penggunaan *content-based* susah di *implementasi* kan, sebab diperlukan banyak data dan data tidak selalu ada (*open source*)[6]–[8].

Oleh sebab itu pada penelitian ini penulis mencoba menerapkan dua tahap permodelan *candidate ranking* menggunakan *collaborative filtering*, dan *re-ordering* menggunakan pendekatan *naive bayes*. Berdasarkan penelitian sebelumnya pada penelitian jika penggunaan atribut dalam proses rekomendasi dengan dua tahap akan menghasilkan rekomendasi yang baik dan memperbaiki hasil rekomendasi. Akan tetapi pada penelitian tersebut dilakukan dengan *study* kasus data *implicit* (3 dimensi) dengan data *tagging*, yang menyebutkan jika *re-ordering* yang di lakukan dapat memperbaiki hasil rekomendasi dibandingkan dengan satu proses data (tanpa *re-ordering*) [7][8].

Oleh sebab itu pada penelitian ini penulis mencoba menerapkan atribut berupa *genre* digunakan untuk menghasilkan rekomendasi. Selain menerapkan atribut berupa *genre* pada penelitian ini, menggunakan satu metode dengan dua model rekomendasi, yaitu model tradisional *item-based* CF yang kemudian hasil rekomendasinya akan di *re-ordering* menggunakan model pendekatan *naïve bayes at probabilistic ranking* dengan *study* kasus *movie* atribut *genre* pada data *explicit* (2 dimensi).

Dengan permasalahan seperti dijelaskan diatas, penulis mengharapkan mendapatkan hasil rekomendasi yang sesuai dengan kebutuhan *user*. Dengan menggunakan satu metode, dengan dua model pendekatan untuk menghasilkan sebuah rekomendasi untuk *user* yaitu model tradisional *item-based* CF yang hasilnya di *re-ordering* dengan

model *naive bayes at probabilistic ranking* yang memanfaatkan atribut berupa *genre* pada *film*, untuk memperbaiki hasil rekomendasi.

1.2. Rumusan Masalah

Berdasarkan latar belakang permasalahan diatas, maka dapat ditentukan rumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut :

1. Berapa jumlah *atribute genre* yang dapat mempresentasikan selera seorang *user*?
2. Apakah hasil rekomendasi dari penggunaan atribut *genre* pada model *naive bayes* untuk *probabilistic ranking* lebih baik dari model *item-based CF* tanpa penggunaan atribut *genre*.

1.3 Tujuan

Mengacu pada latar belakang serta perumusan masalah diatas maka tujuan dari penelitan ini adalah :

1. Untuk mengetahui jumlah *attribute genre* yang dapat mempresentasikan selera seorang *user*.
2. Untuk membandingkan hasil rekomendasi dari penggunaan atribut *genre* pada model *naive bayes* untuk *probabilistic ranking* lebih baik dari model *item-based CF* tanpa penggunaan atribut *genre*.

1.4 Manfaat Penelitian

Manfaat dari hasil penelitian ini adalah mengetahui berapa jumlah *atribut genre* yang dapat mempresentasikan selera seorang *user* dengan melakukan *re-ordering* dari hasil prediksi *rating candidate item* dari model tradisional *item-based CF* dengan model pendekatan *naive bayes at probabilistic ranking* dengan .

1.5 Batasan Masalah

Dari perumusan permasalahan diatas maka perlu adanya sebuah batasan masalah sebagai berikut:

1. Sistem rekomendasi menggunakan model *Naive bayes, probabilistic ranking*, dan *Item-based collaborative filltering*.
2. Studi kasus dari penelitian ini menggunakan data dari *MovieLens.org* dengan 100.000 data *rating*, 9.000 *film* dan 700 *user*. Terakhir di *update* pada 10/2016.
3. Data didownload (<https://grouplens.org/datasets/movielens/100k/>).
4. Pengembangan menggunakan *software Matlab2015b*.

5. *Content item* yang digunakan adalah *genre* pada *film*.
6. *Item* yang dijadikan objek rekomendasi adalah *film* yang telah dirating oleh *user* yang didapat dari dataset *MovieLens.org*.
7. Tidak menghitung untuk *cold-start problem* baik pada *user* maupun *item*.
8. Model evaluasi adalah *offline*, yang berarti bahwa daftar rekomendasi akan dihasilkan dengan menghitung data *history user*.

1.6 Model Penelitian

Secara umum, penelitian dilakukan dengan beberapa tahapan, yaitu:

a. Studi Pustaka

Pada tahap ini merupakan tahap dimana pengumpulan beberapa informasi yang dibutuhkan serta mempelajari informasi yang diperoleh untuk menyelesaikan pengerjaan tugas akhir ini. Informasi yang dikumpulkan meliputi model sistem rekomendasi, *collaborative filtering*, *naive bayes*, dan *probabilistic ranking*. Informasi terkait diperoleh melalui internet, *journal*, buku, *literatur* dan *referensi* lainnya.

b. Mempersiapkan Data

Pada tahap ini, persiapan dibagi menjadi 2 bentuk :

1. Mendapatkan *dataset* yang digunakan dalam penelitian.
2. *Dataset* yang di dapat, harus dibagi menjadi 2. Pertama sebagai data *training* (pelatihan), yang kedua adalah sebagai data *testing* (pengujian).

c. Analisa dan Perancangan

Setelah mendapatkan data yang dibutuhkan, selanjutnya penulis melakukan analisa guna menyusun rancangan sistem yang akan dibuat. Perancangan di susun sesuai dengan proses analisa untuk mendapatkan hasil sesuai yang diharapkan.

d. Membuat dan Menyusun Fungsi Program

Tahap ini membuat setiap fungsi-fungsi yang digunakan untuk penelitian. Fungsi ini meliputi, fungsi *user-genre*, *genre-item*, *user-user*, *mean center rating* dan masih banyak lainnya. Membuat fungsi ini akan mempermudah dalam pembuatan program untuk penelitian. Setelah fungsi program telah dibuat, setiap fungsi disusun sesuai dengan model yang digunakan. Agar menjadi sebuah sistem yang utuh dalam membentuk sistem rekomendasi.

e. Uji Coba Sistem

Pada tahap ini sistem akan dilakukan proses pelatihan data dengan menggunakan model *K-Fold Cross Validation*. Dataset dibagi menjadi lima *fold* yang kemudian dilakukan proses analisa dan evaluasi.

f. Analisa dan Evaluasi

Analisa dari uji coba sistem akan di evaluasi dengan metode NDCG. Evaluasi dilakukan untuk mengetahui hasil akurasi rekomendasi dari model yang dibuat.

g. Dokumentasi

Dokumentasi dilakukan sebagai penyusunan laporan tugas akhir yang berlangsung mulai dari tahap studi pustaka hingga tahap evaluasi.

1.7 Sistematika Penulisan

Adapun sistematika pembahasan tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini berisikan latar belakang dari model yang digunakan pada Sistem Rekomendasi, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, model penelitian dan sistematika penulisan.

BAB II STUDI PUSTAKA

Pada bab ini dijelaskan mengenai materi-materi yang menjadi landasan dalam penyusunan penelitian ini. Materi-materi tersebut mengenai Sistem Rekomendasi dan bagian-bagiannya, *Item-based Collaborative Filtering*, *dataset MovieLens.org*, *Naive bayes* pada *probabilistic ranking*, dan model Evaluasi.

BAB III PERANCANGAN SISTEM

Pada bab ini akan menjelaskan rancangan dari seluruh sistem yang akan dibuat, mulai kebutuhan *software* dan *hardware* yang dibutuhkan, rancangan model Sistem Rekomendasi, serta alur dari sistem yang akan dikembangkan.

BAB IV IMPLEMENTASI SISTEM

Pada bab ini akan dijelaskan implementasi sistem dari rancangan yang telah dibuat pada bab sebelumnya, dimana sistem akan diuji coba berdasarkan model dan skenario yang telah ditentukan.

BAB V PENUTUP

Pada bab ini berisi kesimpulan dari hasil uji coba serta saran yang dapat digunakan untuk pengembangan selanjutnya.

~Halaman Ini Sengaja Dikosongkan~

