

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan terus terjadi pada dunia teknologi, yang berdampak besar pada bidang komunikasi dan informasi. Hal ini dibuktikan dengan terus meningkatnya penggunaan internet pada kehidupan di masyarakat. Menurut survey yang telah dilakukan oleh Asosiasi Penyelenggara Jaringan Internet Indonesia (APJII). Perkembangan internet di Indonesia saat ini menunjukkan dampak positif. Ini tampak dari jumlah pengguna internet yang semakin berkembang dari tahun ke tahun. Dimana hasil survei APJII pada tahun 2017 mencatat adanya peningkatan pengguna internet di Indonesia dari 132,7 juta pada 2016, melonjak menjadi 143,26 juta jiwa dari total 262 juta penduduk Indonesia[1].

Selain internet dunia perfilman juga mengalami perkembangan pesat. Peningkatan ini diikuti oleh meningkatnya minat menonton film oleh masyarakat. Namun dengan berkembangnya dunia perfilman, jumlah film yang beredar pun semakin banyak, sehingga masyarakat memerlukan informasi yang dapat membantu dalam menentukan film yang diinginkan. Berbagai cara telah dilakukan untuk mempermudah seorang *user* dalam menetapkan suatu pilihan salah satunya adalah rekomendasi sistem.

Rekomendasi sistem adalah alat atau teknik perangkat lunak yang menyediakan sugesti mengenai suatu *item* kepada pengguna atau *user*. Sugesti

terkait dengan berbagai proses pengambilan keputusan, seperti barang apa yang akan dibeli, music apa yang ingin didengar , atau film apa yang ingin ditonton.

Secara teknik rekomendasi sistem terbagi dari empat kategori yaitu[2] : 1) Collaborative Filtering ; 2) Content Based Filtering ; 3) Demographic Filtering 4) Knowledge Based Filtering .

Terkadang satu teknik saja tidak cukup untuk menyelesaikan permasalahan di rekomendasi sistem seperti *cold-start*, *sparsity*, *diversity* dan lainnya. Maka dibutuhkan kombinasi dari beberapa metode seperti[2] CF-X, CF-CBF, CF-CBF-X, IICF-UUCF, CBF-X. Penggabungan IICF-UUCF telah dilakukan oleh beberapa peneliti seperti penelitian yang dilakukan oleh Jiangnan Xu dan temannya yang menggunakan metode topic model based *Collaborative Filtering* (TMCF), Ramadhanuz A-Djamil dan temannya yang menggunakan algoritma *Item-Based Clustering Hybrid Method* (ICHM) berbasis model yang menggabungkan pendekatan CF dan CBF yang diklaim dapat mengatasi masalah *cold-start*. penelitian yang dilakukan Jiangnan Xu dan temannya tidak dapat membuktikan bahwa model yang telah mereka buat dapat menyelesaikan permasalahan rekomendasi system terutama tentang keefektifan dan efisiensi model yang telah dibuat. Serta mereka hanya membandingkan model yang mereka buat dengan model CF tradisional saja. Sedangkan penelitian yang dilakukan Ramadhanuz A Djamil dan temannya telah berhasil dalam membuktikan bahwa dengan menambahkan metode clustering pada *Hybrid method* dapat mengatasi masalah *cold-start* pada *item*

baru yang tidak ada hasil *ratingnya*, akan tetapi penelitian ini tidak memberikan tingkat akurasi yang detail terhadap pengaruh lain seperti permasalahan *sparsity* data.

Beberapa poin pembeda penelitian yang dilakukan dengan penelitian sebelumnya adalah :

- a) Peneliti menggunakan tiga *dataset movie* yang berbeda dalam menguji algoritma.
- b) Peneliti menggunakan pengujian *AUC Score*, *recall*, dan *precision at k* sebagai penentuan akurasi
- c) Peneliti menggunakan Algoritma *Matrix factorization* dengan bantuan *LightFM*

1.2 Rumusan Penelitian

Mengetahui performa penggabungan teknik *Collaborative Filtering* dan *Content-based filtering* menggunakan *Matrix factorization* dengan bantuan *LightFM* pada *dataset* *movielens*.

1.3 Batasan Penelitian

Batasan masalah dalam penelitian ini adalah:

1. Algoritma yang digunakan adalah *matrix factorization* dengan metode *Hybrid Recommendation* berbasis model dengan bantuan *library* *LightFM*
2. Sistem rekomendasi ini hanya sebagai sistem pemberian rekomendasi dalam menentukan film apa yang akan ditonton, dimana hasil akhir tetap menjadi keputusan dari pengguna.

3. Kriteria yang digunakan sebagai dasar penelitian adalah *rating* dari penilaian para pengguna
4. Nilai *rating* yang digunakan adalah skala 1 samapai 5
5. Pembuatan sistem rekomedasi ini menggunakan data public film dari *dataset* movielens 100k sebagai data *training* dan *testing*, serta *dataset* Kaggle.
6. Dalam menentukan rekomendasi ini peneliti menggunakan semua daftar *genre* yang ada di *dataset*
7. *Output* nantinya berupa hasil rekomendasi film.

1.4 Maksud dan Tujuan Penelitian

Tujuan dari penelitian ini adalah

1. untuk menggabungkan metode *Collaborative Filtering* dan *Content-based filtering* untuk membuat sistem rekomendasi film.
2. Mengetahui performa penggabungan teknik *Collaborative Filtering* dan *Content-based filtering* menggunakan *Matrix factorization* dengan bantuan LightFM pada *dataset* movielens.

1.5 Metode Penelitian

Metode penelitian yang dilakukan adalah :

1. Pengumpulan data

Mencari jurnal terkait dan mengumpulkan data yang dilakukan dengan mendownload file pada *website* grouplens. Data yang digunakan adalah

100.000 *ratings* dari 1000 *users* di 1700 film sebagai data *training* dan *testing*. Lalu untuk pengujiannya data yang digunakan adalah data 25.000.000 *rating* dari 162.000 *user* dan 62.000 film, serta *dataset* Kaggle yang berjumlah 26.000.000 *rating* dari 270.000 *user* dan 45.000 film.

2. Analisis

Melakukan analisa terhadap syarat-syarat yang dibutuhkan dalam penelitian

1) Analisis masalah

Melakukan pemetaan terhadap masalah yang akan diselesaikan melalui penelitian ini.

2) Analisis kebutuhan data

Memetakan kebutuhan data yang akan di gunakan pada penelitian ini, yaitu data film dari judul , *genre* , *rating* , *user* .

3) Analisis kebutuhan penelitian

Memetakan kebutuhan yang diperlukan selama penelitian, yaitu kebutuhan fungsional dan kebutuhan non fungsional.

3. Perancangan

Pada tahap ini dilakukan perancangan metode *Hybrid matrix factorization* dengan menggunakan *library* LightFM yang akan diimplementasikan untuk mendapatkan sebuah model yang mampu menghasilkan rekomendasi film.

4. Implementasi

Pada tahap ini akan dilakukan penulisan *code program*.

5. Pengujian

Pada tahap ini akan dilakukan perhitungan akurasi dari data yang digunakan. Metode yang digunakan adalah dengan menghitung *Precision at k*, *Recall at k* dan *ROC AUC score* dengan kombinasi Adagrad dan Adadelata terhadap model dan data yang digunakan. Hasil pengujian adalah jumlah *score* dimana *score* yang mendekati 1.0 menandakan hasil yang lebih baik.

