

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sistem rekomendasi adalah subkelas dari sistem penyaringan informasi yang digunakan untuk memprediksi peringkat atau nilai yang akan diberikan kepada pengguna pada sebuah barang. Saat ini *Recommender system* (RS) banyak digunakan untuk aplikasi *online*, seperti sosial media, rekomendasi sebuah produk pada *e-commerce* dimana sistem akan memberikan prediksi atau penawaran kepada pelanggan. Teknologi ini sudah banyak diadopsi oleh beberapa perusahaan besar dunia seperti Amazon, Netflix, dan eBay untuk memberikan rekomendasi produk kepada pengguna. Sistem rekomendasi menjadi teknologi yang penting untuk *e-commerce* karena dengan fungsi untuk memberikan referensi produk yang disukai oleh pelanggan, membuat strategi marketing dan meningkatkan loyalitas pelanggan [1].

Menurut jenis data yang dikumpulkan dan cara menggunakannya dalam sebuah sistem rekomendasi, maka diklasifikasikan menjadi beberapa pendekatan yaitu *content-based* (CB), *collaborative filtering* (CF), dan *hybrid* atau gabungan [2].

Content-based filtering (CB) banyak digunakan pada sistem rekomendasi dengan memanfaatkan konten item kemudian membuat fitur agar sesuai dengan pengguna, namun pendekatan ini memiliki satu kelemahan besar yaitu RS perlu mempelajari kecenderungan *user* untuk beberapa jenis item dan menggunakannya untuk item yang lain. *Collaborative filtering* (CF) merupakan pendekatan paling

populer untuk *design* sistem rekomendasi dengan memanfaatkan jumlah data yang didapatkan dari rating yang diberikan oleh pengguna sebelumnya dan kemudian memberikan prediksi item yang akan disukai oleh pengguna. Pendekatan ini bergantung antara pengguna dan item yang kemudian akan diubah menjadi sebuah matriks [3].

Collaborative Filtering terbagi menjadi 2 pendekatan yaitu *memory-based* dan *model-based*. *Model-based* menjadi pendekatan yang lebih baik dibandingkan dengan *memory-based* dalam hal mengatasi data *sparsity* [4]. Contoh algoritma yang termasuk dalam model base antara lain *Singular Value Decomposition* (SVD), *Non-negative Matrix Factorization*(NMF), *Alternating Least Square* (ALS), SVD++ dan beberapa algoritma lain.

Pada penelitian ini peneliti akan membandingkan tiga algoritma yang termasuk dalam *model-based* pada beberapa dataset antara lain *MovieLens* dan *Book Crossing* untuk mengidentifikasi algoritma yang memiliki akurasi paling tinggi dalam mengatasi *sparsity*.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, diperoleh rumusan masalah yaitu diantara algoritma yang termasuk dalam *model-based* (SVD, NMF, ALS) manakah algoritma yang memiliki akurasi paling tinggi dalam mengatasi masalah data *sparsity* pada data seperti *MovieLens* dan *Book Crossing*.

1.3 Batasan Masalah

Pada penelitian ini akan mencari akurasi dan efektifitas dari beberapa algoritma pada pendekatan *model-based* yaitu SVD, NMF dan ALS dalam menghadapi data sparsity pada dataset *Movielens* dan *Book Crossing*. Metode pengujian akan menggunakan *Root Mean Square Error* (RMSE).

1.4 Maksud dan Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maksud dan tujuan dilakukannya penelitian ini adalah membandingkan beberapa algoritma yang termasuk dalam *model-based* dan mencari algoritma manakah yang memiliki akurasi paling baik dalam mengatasi data sparsity pada dataset *movielens* dan *book crossing*

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah dapat mengetahui algoritma *model-based* yang memiliki akurasi terbaik yang diterapkan pada dataset publik.

1.6 Metode Penelitian

Berikut metode penelitian yang akan digunakan dalam proses pengerjaan penelitian ini dijelaskan sebagai berikut

1.6.1 Pengumpulan Data

Mencari jurnal terkait kemudian mendownload dataset publik yaitu *Movielens* dan *Book Crossing* yang sudah tersedia di beberapa website seperti *kaggle* dan *groupLens*. Dataset yang diambil tidak hanya satu karena untuk menghasilkan analisis yang lebih baik.

1.6.2 Preprocessing

Preprocessing yang akan dilakukan pada penelitian ini guna membersihkan data rating yang tidak jelas seperti pengguna yang tidak jelas dan rating yang bernilai nol. Dengan *preprocessing* data diharapkan hasil yang akan diperoleh menjadi lebih tinggi dan lebih *valid*.

1.6.3 Analisis Data

Analisis data digunakan sebagai pembandingan dengan data lain sehingga kesimpulan dari hasil yang diperoleh setiap algoritma lebih mendasar.

1.6.4 Pengujian Algoritma

Setelah menganalisis data maka selanjutnya akan dilakukan proses pengujian. Dari algoritma yang sudah disebutkan sebelumnya akan dilihat algoritma mana yang memiliki akurasi dan efektifitas paling baik dalam menangani masalah data sparsity. Pengujian yang akan dilakukan akan menggunakan RMSE.

1.7 Sistematika Penulisan

Untuk mengetahui uraian singkat yang memuat gambaran singkat secara keseluruhan isi masing-masing bab, maka dibuat sistematika sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini menjelaskan tentang dasar penelitian, yang berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan dalam pembuatan laporan ini.

BAB II LANDASAN TEORI

Pada bab ini berisi kajian pustaka yang mirip dengan penelitian ini. Pada bab ini juga berisi tentang landasan teori yang mendukung dalam penelitian ini.

BAB III PERANCANGAN SISTEM

Pada bab ini berisi tentang alur dari penelitian yang berupa perancangan perangkat dan bahan yang akan digunakan.

BAB VI PEMBAHASAN

Pada bab ini berisi tentang hasil dari tahapan penelitian yang dilakukan secara menyeluruh termasuk hasil dari pengujian.

BAB V PENUTUP

Bagian terakhir dari penelitian yang berisi tentang kesimpulan dan saran untuk memperbaiki kekurangan yang ada pada penelitian.

