

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi informasi memberikan kemudahan baik dalam hal mendapatkan berbagai informasi, dan juga menyebarkan informasi, serta adanya teknologi informasi bisa dimanfaatkan untuk meningkatkan layanan. Layanan yang optimal merupakan hal yang sangat penting bagi perusahaan maupun instansi, sehingga banyak media ataupun teknik digunakan untuk mewujudkan hal tersebut. Salah satu media adalah sistem on-line seperti pada e-commerce. E-commerce akan memberikan rekomendasi produk ataupun jasa sesuai dengan apa yang menjadi peminatannya, salah satu pendekatan yang digunakan adalah sistem rekomendasi. Sistem rekomendasi bekerja dengan cara mencari informasi paling relevan yang diambil dari sejumlah besar data untuk menghasilkan rekomendasi yang sesuai dengan minat user. Sistem rekomendasi bertujuan mengurangi upaya dan waktu pengguna dalam mencari informasi yang sesuai dengan peminatannya[1]. Rekomendasi paling relevan bisa dilakukan dengan content-based filtering[2], [3], [4], demographic filtering[5], [6], collaborative filtering[7], [8], dan hybrid filtering[2],[9], [10].

Collaborative filtering (CF) merupakan metode yang sukses dan banyak digunakan dalam sistem rekomendasi[11]. Metode tersebut bekerja dengan cara menganalisis pola data rating untuk membuat prediksi. Selain itu metode CF merupakan metode yang sederhana dan efisien[8]. Collaborative filtering selanjutnya dibagi menjadi dua kategori yaitu memory-based approach dan model-based approach[1]. Pada memory-based approach umumnya akan berkinerja baik dengan data yang padat, sedangkan pada data yang jarang (sparsity) maka pendekatan tersebut kurang bisa diandalkan. Sparsity merupakan kondisi dimana data yang belum terisi penuh atau jarang, salah satu penyebabnya adalah adanya item baru yang kurang diinginkan oleh pengguna atau karena minimnya informasi sehingga pengguna belum bisa memberikan rating terhadap item tersebut. Selain itu adanya pengguna baru, yang tidak menyukai item tersebut atau membiarkan item tanpa memberikan rating. Jika data rating kondisinya jarang maka sulit untuk menentukan kesamaan antara pengguna, sehingga kualitas rekomendasi rendah[12].

Beberapa penelitian telah dilakukan untuk mengatasi masalah sparsity diantaranya oleh Alhijawi et al. yang mengusulkan metode berbasis statistic dengan memanfaatkan matrik user-item rating dan matrik item-fitur untuk membangun matrik User Interest Print (UIP). UIP merupakan matrik yang telah terisi (padat) yang menyimpan data tingkat kepuasan pengguna terhadap item's semantic feature[13]. Selanjutnya Yu, Peng mengusulkan algoritma item collaborative filtering berdasarkan kesamaan atribut[14]. Hussien Mohamed et al. menggunakan association rule dengan cluster based untuk menyelesaikan masalah sparsity dan meningkatkan akurasi[12]. Sementara pada penelitian kami, menggunakan imputing missing value untuk mengatasi masalah sparsity pada collaborative filtering. Data yang banyak kosong akan di isi dengan beberapa teknik imputation missing value

yaitu mean, min, max dan KNN Imputer untuk menggantikan rating yang bernilai kosong (NaN). Dengan imputation missing value maka masalah sparsity bisa diatasi sehingga kualitas rekomendasi bisa ditingkatkan.

1.2 Identifikasi Masalah

Sparsity dalam konteks collaborative filtering adalah masalah yang sering dihadapi. Ini merupakan situasi di mana sebagian besar data rating pengguna bernilai nol, karena pengguna umumnya hanya memberikan peringkat untuk sejumlah kecil item. Pengguna cenderung memberikan rating hanya pada sebagian item dari semua item yang tersedia sehingga membuat sulit untuk membangun model yang akurat karena kekurangan informasi rating untuk semua item. Beberapa pengguna mungkin aktif memberikan rating pada banyak item, sementara yang lain hanya memberikan peringkat pada sebagian kecil item itu mengakibatkan pengguna yang aktif dapat memiliki pengaruh yang lebih besar pada item, sedangkan pengguna yang pasif data yang didapat diabaikan

1.3 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, maka dapat dirumuskan beberapa permasalahan sebagai berikut

1. Bagaimana imputasi missing value dapat digunakan sebagai solusi untuk mengatasi masalah sparsity pada Collaborative Filtering?
2. Bagaimana penerapan imputasi dapat meningkatkan akurasi Root Mean Square Error yang bagus pada Collaborative filtering.

1.4 Batasan Masalah

Beberapa batasan masalah pada penelitian ini adalah:

1. Penelitian hanya difokuskan pada imputation data
2. Data yang digunakan dalam penelitian adalah data movielens yang didapat dari website kagle
3. Objek dalam penelitian ini adalah missing value rating movie
4. Metode yang digunakan yaitu mean, max, min dan KNN Imputer untuk imputation dan untuk menguji performa menggunakan RMSE (root mean square error)
5. Tools yang digunakan untuk uji coba yaitu Jupyter Notebook Anaconda

1.5 Tujuan Penelitian

Berdasarkan perumusan masalah di atas maka peneliti memiliki tujuan penelitian yaitu menggunakan imputing missing value untuk mengatasi masalah sparsity pada collaborative filtering

1.6 Manfaat Penelitian

Manfaat dari penelitian ini yaitu ditemukan metode imputation missing value

yang bagus untuk mengatasi sparsity pada collaborative filtering

1.7 Sistematikan Penulisan

Berikut uraian singkat mengenai sistematika penulisan pada masing – masing bab yang digunakan oleh penulis untuk mempermudah dalam penyusunan tesis ini:

BAB I PENDAHULUAN

Bab I mendeskripsikan tentang latar belakang masalah, identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada bagian ini berisi penelitian yang pernah dilakukan terkait topik penelitian; dan teori-teori pendukung yang digunakan yang berhubungan erat dengan pokok-pokok landasan berpikir, yang relevan dengan topik penelitian.

BAB III METODE PENELITIAN

Pada bagian ini berisi metode penelitian menggunakan metode imputation mean, max, min dan KNN Imputer serta tahapan yang dilakukan dalam penelitian antara lain: langkah-langkah penelitian dan pengumpulan data.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Pada bab ini berisi pembahasan yang menjelaskan tentang data yang digunakan, uji validitas, proses imputation yang menggunakan metode mean, max, min, dan KNN Imputer.

BAB V SIMPULAN DAN SARAN

Pada bab ini berisi tentang rangkuman dari pembahasan, yang terdiri dari jawaban atas perumusan masalah, tujuan penelitian, dan hipotesis. Selain itu berisi saran untuk penelitian selanjutnya, sebagai hasil pemikiran penelitian atas keterbatasan penelitian yang dilakukan.

DAFTAR PUSTAKA

Daftar pustaka berisi tentang literatur yang dipakai atau dikutip dalam penelitian yang dapat berupa buku dan jurnal fisik ataupun elektronik.

LAMPIRAN

Bagian ini berisi data atau dokumen yang dapat mendukung penelitian seperti berbentuk gambar, tabel, formulir, flowchart, dan lain sebagainya.