

BAB I PENDAHULUAN

1.1 Latar Belakang

Semakin banyak layanan yang tersedia di internet atau meningkatnya pengembangan layanan berbasis internet memberikan sejumlah data yang besar atau informasi yang berlebihan, sehingga menimbulkan *user* akan kesusahan memilih *item*/barang terbaik yang sesuai dengan kebutuhan masing – masing [1]. Para *user* tersebut membutuhkan suatu metode sistem rekomendasi untuk memilah atau menentukan informasi yang sesuai dengan preferensi. Sistem rekomendasi dapat menentukan atau memberikan preferensi *user* tentang *item*, dan tujuan dari sistem tersebut adalah sebagai rekomendasi *item*/barang yang kemungkinan *user* menyukai atau sesuai dengan preferensi [2].

Dalam sistem rekomendasi terdapat tiga tipe rekomendasi yaitu *Collaborative Filtering*(CF), *Content Based* (CB) dan *Hybrid* [3]. Algoritma CF merupakan metode sistem rekomendasi yang paling populer dan banyak digunakan, metode ini melakukan pendekatan dengan mengidentifikasi sekumpulan pelanggan serupa yang berdasarkan transaksi sebelumnya atau memberikan umpan balik pelanggan tentang *item* atau produk, yang kemudian memprediksi peringkat *item* berdasarkan perilaku yang diamati [4]. Algoritma CF memiliki dua kategori yaitu algoritma *Memori Based* dan *Model Based*. Sistem rekomendasi memiliki beberapa masalah atau tantangan yang berhubungan dengan *sparsity data*, *scalability* dan *synonym*.

Data *sparsity* merupakan suatu kejadian yang ketika sejumlah *user* dan *item* sebagian besar belum dinilai atau data memiliki banyak data kosong yang belum diberi nilai atau *rating* [3]. Banyaknya data kosong berarti semakin jarang data yang dapat digunakan untuk acuan rasio prediksi jumlah *rating* atau nilai. Hasil prediksi akan kurang optimal atau kurang sesuai *preferences user*, maka mengatasi masalah *sparsity data* diperlukan untuk keakuratan rekomendasi.

Sistem rekomendasi memiliki berbagai pendekatan dalam menyelesaikan permasalahan sistem rekomendasi, namun dalam masalah data *sparsity* tidak semua pendekatan dapat menyelesaikan seperti CF *memori based*. Pendekatan *memori based* CF dinilai tidak cocok untuk menyelesaikan masalah ini, karena

algoritma ini didasarkan pada matriks kesamaan antara *user* dan *item* yang nilai *rating* data jarang atau sebagian besar kosong. Di sisi lain, pendekatan model based CF memberikan asumsi tertentu tentang data yang bisa disesuaikan dengan model dan penawaran yang diberikan prediksi untuk *item* yang tidak diberi nilai atau *rating*. Pendekatan yang pernah dilakukan para peneliti dalam penyelesaian masalah data *sparsity* yaitu *Singular Value Decomposition* (SVD) [5], *Non-negative Matrix Factorization* (NMF) [3, 7], *Regularization Single-Element-Based Non-negative Matrix Factorization* (RSNMF) [6], *Sparse Probabilistic Matrix Factorization* [8], *Hierarchical Itemspace Rank* (HIR) [9], *Nearly Completely Decomposable* (NCD) *architecture* [2], *Social Behaviour* dan *hybrid* [10]. Dari berbagai pendekatan yang digunakan oleh peneliti sebelumnya, menunjukkan bahwa kebanyakan menggunakan pendekatan berbasis *Matrix Factorization* (MF), dimana MF merupakan pendekatan model based CF yang mempelajari tentang masalah data *sparsity*.

Pendekatan berbasis MF memiliki berbagai macam metode seperti SVD, NMF, PMF dan lainnya. SVD melakukan faktorisasi dimana data asli matriks direduksi menjadi matriks yang lebih kecil dengan optimal *rank* 'k'. NMF merupakan teknik yang berdasarkan reduksi dimensi seperti SVD, namun NMF mempertahankan nilai positif. PMF merupakan teknik yang berdasarkan MF serta probabilitas, PMF melakukan reduksi dimensi atau faktorisasi yang kemudian menghitung probabilitas MF. NMF sering digunakan dalam menyelidiki data *sparsity* sistem rekomendasi karena dinilai unggul atau lebih baik, dibuktikan dengan hasil evaluasi yang memiliki nilai lebih baik dari metode lain dan metode NMF merupakan algoritma dengan komputasi dan kompleksitas penyimpanan yang cukup rendah [3, 6]. Algoritma NMF akan menentukan nilai yang sesuai untuk data kosong yang kemudian memperkirakan *rating* yang sesuai dari matriks asli [6]. Setelah pengurangan dimensi, algoritma akan merekonstruksi matriks data asli dan mengkonversi data yang jarang atau kosong menjadi data padat dan akhirnya akan dilakukan prediksi *rating* yang sesuai dan mengevaluasi data tersebut.

Matriks evaluasi kinerja sistem yang berdasarkan model peringkat terdapat *Discounted Cumulative Gain* (DCG), *Mean Average Precesion* (MAP) dan lain-

lainya. Model matriks evaluasi seperti DCG dapat diterapkan untuk mengoptimal suatu sistem seperti prediksi *rating* pada rekomendasi, dengan menggunakan optimasi matriks evaluasi akan dapat meningkatkan kinerja rekomendasi [11]. Metode berbasis NMF terkadang menerapkan optimasi dalam meningkatkan kinerja rekomendasi seperti menggunakan model *least square* seperti *Measure Square Error* (MSE) dan *Euclidean Distance* [6]. Model DCG lebih banyak digunakan terutama untuk model yang menerapkan data bertingkat relevansi, selain itu DCG mampu memberikan penyelesaian tugas rekomendasi dengan kualitas yang baik [11, 12].

Pada penelitian ini, melakukan proses optimasi data *sparsity* menggunakan metode berbasis Non-negative Matrix Factorization dengan Optimasi *Discounted Cumulative Gain* (DNMF), dimana DCG sebagai optimasi kriteria. DNMF akan mengoptimalkan data asli yang difaktorisasi sehingga menghasilkan nilai atau *rating* yang sesuai pada matriks kosong. Metode DNMF yang muncul merupakan kombinasi dari NMF [3, 6, 7] dan optimasi DCG [11, 12], dimana NMF memiliki nilai lebih dari metode lain seperti SVD yang memiliki hasil kurang optimal dengan baik. Sedangkan, optimasi DCG memiliki nilai optimasi yang baik dengan meminimalkan kerugian atau ketidakakuratan dan optimasi DCG dapat menangani masalah distribusi data miring atau kosong dalam peringkat.

Menggunakan metode dalam penelitian ini, diharapkan bisa memberikan rekomendasi yang sesuai dengan preferensi *user*. Selain itu, diharapkan dapat mengatasi berbagai macam masalah data *sparsity* dan *scalability* dengan tingkat akurasi yang tinggi. Dengan menggunakan metode optimasi akan meningkatkan akurasi prediksi *rating*, sehingga hasil penelitian ini bisa menjadi alternatif atau rujukan dalam membangun suatu sistem rekomendasi yang terkendala dengan masalah data *sparsity*.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Sparsity merupakan suatu masalah yang umum terjadi dalam sistem rekomendasi. Masalah ini berkaitan dengan sedikitnya nilai atau informasi yang telah terisi dari sebagian besar dataset yang ada, sehingga dapat mengakibatkan

menurunnya tingkat akurasi rekomendasi *item* terhadap *user* dikarenakan semakin sedikitnya data yang bisa dibandingkan. Metode NMF dan beberapa metode lain telah digunakan oleh peneliti sebelumnya dalam memberikan solusi dari permasalahan ini, terutama metode NMF yang dinilai baik atau bagus dalam mengatasi *sparsity* data.

1.2.2 Metode Usulan

Solusi permasalahan atau metode usulan untuk mengatasi masalah data *sparsity* yaitu dengan metode DNMF.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan dan usulan solusi atau metode maka didapatkan rumusan masalah :

1. Berapa hasil evaluasi rekomendasi yang dihasilkan oleh metode DNMF dalam mengatasi *sparsity* data?
2. Apakah hasil evaluasi metode DNMF lebih baik daripada hasil evaluasi metode NMF dalam mengatasi *sparsity* data?

1.3 Tujuan & Manfaat

1.3.1 Tujuan Penelitian

Mengacu pada latar belakang serta rumusan masalah diatas maka berikut tujuan dari penelitian ini :

1. Untuk mengetahui hasil evaluasi DNMF dalam mengatasi *sparsity* data.
2. Untuk mengetahui perbandingan hasil evaluasi antara DNMF dan NMF.

1.3.2 Manfaat

Hasil penelitian ini bisa menjadi alternatif atau rujukan dalam membangun suatu sistem rekomendasi yang terkendala dengan masalah *sparsity* data.

1.4 Batasan-batasan

Batasan masalah pada penelitian ini adalah dataset yang digunakan yaitu dataset dari *MovieLens* 100K (<https://grouplens.org/datasets/movielens/>).

1.5 Metode Penelitian

Secara umum, penelitian dilakukan dengan beberapa tahapan yaitu :

1.5.1 Studi Pustaka

Pada tahap ini, semua informasi yang diperlukan dalam penelitian dikumpulkan serta dipelajari sehingga dapat membantu dalam menyelesaikan proses pengerjaan riset. Data yang dikumpulkan berupa artikel yang berisikan materi tentang NMF dan penelitian yang terkait dengan sistem rekomendasi dalam permasalahan data *Sparsity*. Selain itu, dataset yang digunakan pada penelitian ini adalah dataset *movie* dari MovieLens atau GroupLens.

1.5.2 Analisa dan Perancangan Sistem

Menganalisa permasalahan yang diperoleh setelah mempelajari berbagai data yang telah terkumpul. Dari data yang telah didapat, dikumpulkan dan dipelajari maka diperoleh suatu permasalahan yaitu, optimasi data *Sparsity* dalam sistem rekomendasi. Beberapa penelitian telah melakukan riset untuk mengatasi atau mengoptimalkan data *Sparsity*, sehingga didapat rekomendasi yang sesuai dengan kebutuhan meskipun data referensi hanyalah sedikit (data banyak kosong), namun dari beberapa penelitian terdapat kelemahan masing – masing dalam permasalahan tersebut.

1.5.3 Implementasi Sistem

Rancangan sistem yang telah dibuat dari analisa permasalahan, kemudian akan diimplementasikan. Tahap implementasi dimulai dari mempersiapkan data *rating movie*, yang kemudian melakukan tahapan prediksi *rating* dengan metode DNMF. Proses implementasi sistem ini akan dilakukan dengan menggunakan bahasa pemrograman python.

1.5.4 Uji Coba Sistem

Pada tahapan ini, dilakukan proses uji coba sistem dimana data *training* dilakukan perhitungan dengan metode DNMF sehingga diperoleh matriks prediksi *rating* serta bobot optimasi.

1.5.5 Analisa dan Evaluasi

Matriks prediksi *rating* dan bobot optimasi yang dihasilkan dari proses DNMF, selanjutnya akan dilakukan pengujian dengan menggunakan data uji. Tahapan ini dilakukan untuk mengetahui tingkat akurasi dari prediksi *rating*. Apabila tingkat akurasi pada proses pengujian tidak sesuai dengan harapan, maka akan dilakukan evaluasi terhadap rancangan arsitektur sistem yang dibuat.

1.5.6 Dokumentasi

Dokumentasi dilakukan mulai dari tahapan studi pustaka, analisa dan perancangan sistem hingga tahapan evaluasi atau pengujian.

1.6 Sistematika Proposal

Sistematika penulisan laporan Proposal ini terdiri dari beberapa bab, yaitu :

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, serta sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas tentang landasan teori dan penelitian sebelumnya.

BAB III METODE USULAN

Pada bab ini membahas tentang tahapan penelitian, rancangan solusi (Input, Proses, Ouput), serta rencana pengujian.

BAB IV IMPLEMENTASI METODE

Pada bab ini membahas tentang implementasi penelitian dari rancangan yang telah dibuat di bab sebelumnya, dimana bab ini berisi hasil implementasi serta hasil evaluasi dari beberapa percobaan.

BAB V PENUTUP

Pada bab ini membahas hasil kesimpulan dari hasil uji coba implementasi metode serta saran untuk mengembangkan sistem dengan penelitian terkait.

REFERENSI

Berisi tentang jurnal-jurnal atau referensi yang dijadikan sebagai acuan dalam penelitian ini.