

1. Pendahuluan

Latar Belakang

Konsep penjualan dengan menggunakan *platform e-commerce* di Indonesia semakin meningkat dari tahun ke tahun. Hal ini didukung dengan meningkatnya pengguna internet di Indonesia yang melihat peluang bisnis yang sangat menjanjikan bagi pelaku bisnis, baik dalam segi ekonomi maupun kemudahan yang ditawarkan. Namun, keberhasilan *e-commerce* dalam penjualan produk sangatlah bergantung pada manajemen pengelolaan dan pengiklanan produknya. Salah satu cara dalam meningkatkan penjualan produk tersebut adalah dengan menggunakan mekanisme sistem rekomendasi.

Metode *collaborative filtering* dan *content-based filtering* merupakan metode yang paling sering digunakan dalam membangun sistem rekomendasi [1,2,3,4]. Kedua metode tersebut memiliki kelebihan dan kekurangannya masing-masing terutama dalam menangani permasalahan *data sparsity* dan *cold start problem* yang sering dijumpai dalam sistem rekomendasi [3]. Kelebihan utama dari *collaborative filtering* adalah metode tersebut dapat memprediksi produk dengan menggunakan pola pengguna yang serupa, sehingga metode ini akan mudah untuk diimplementasikan. Sedangkan kelemahannya akan menimbulkan *cold start problem* bagi pengguna yang baru dan akan terjadi *sparse rating* pada produk yang sama. Metode lainnya yaitu *content-based filtering* juga memiliki kelebihan seperti kualitas rekomendasi yang selalu meningkat dikarenakan konten data dari produk yang selalu mengalami pembaruan. Tetapi, prediksi tersebut tidak dapat dilakukan terhadap jenis barang yang benar-benar berbeda yang tidak pernah diinginkan oleh pengguna tertentu, sehingga hal ini menjadi kelemahan utama dalam menggunakan metode ini, terutama dalam menangani produk baru [5].

Salah satu metode yang pernah dijumpai dalam penelitian sebelumnya adalah menggunakan metode *K-Nearest Neighbor* (KNN) yang menggunakan nilai k sebagai parameter utama dalam menghitung nilai *precision*, *recall*, dan *F-measure* dari metode tersebut [1]. Setelah dilakukan pengujian terhadap metode tersebut, metode KNN cukup bagus digunakan apabila data yang digunakan *sparse* karena nilai *false positive* tidak mempengaruhi nilai *recall*. Namun, metode KNN memiliki akurasi yang rendah yaitu dibawah 0,5 sehingga rekomendasi yang dihasilkan tidak akurat bagi para pengguna *e-commerce* tersebut [1].

Metode lainnya yang pernah digunakan untuk sistem rekomendasi adalah *Linear Regression Model*, yang menggunakan *baseline model* dan modifikasi model linear untuk melakukan pengujian terhadap dataset produk makanan di Amazon dalam menghitung nilai MSE [21]. Hasil MSE yang didapatkan dari kedua model tersebut yaitu 1,698 untuk *baseline model* dan 1,553 untuk model linear. Sehingga dapat dikatakan bahwa model tersebut tidaklah optimal, karena mendapatkan nilai yang cukup tinggi. Hal tersebut disebabkan oleh *rating* pada produk yang tidak lengkap (*sparse*). Hal itu menyebabkan prediksi yang dilakukan terhadap pengguna cenderung tidak akurat bahkan tidak dapat dilakukan.

Penyebab utama data mengalami *sparsity* yang tinggi disebabkan oleh suatu matriks rating yang merupakan hubungan antara *user* dengan produk yang tidak lengkap, hal ini dikarenakan user pada umumnya hanya memberikan rating pada sebagian produk saja. Salah satu cara untuk mengatasi masalah ini adalah menggunakan model variabel tersembunyi (*latent factor*) dari metode faktorisasi matriks [6] yang digunakan sebagai dimensi baru dari vektor pengguna dan vektor produk. Akan tetapi, semakin besar data yang *sparse*, maka akan menyebabkan tingginya kompleksitas komputasi, sehingga dibutuhkan teknik optimasi *gradient descent* dalam penanganan hal ini.

Sehingga untuk mengatasi masalah *sparsity* yang terdapat pada penelitian sebelumnya, penulis akan menggunakan metode faktorisasi matriks berbasis *gradient descent* yang menggunakan model variabel tersembunyi (*latent factor*) dalam penanganan masalah *data sparsity*.

Topik dan Batasannya

Tugas akhir ini difokuskan pada penelitian metode faktorisasi matriks berbasis *gradient descent* terhadap sistem rekomendasi produk pada *e-commerce*. Lalu berdasarkan data latih dan data uji, akan dihitung performansi dari rekomendasi produk menggunakan *root mean square error* (RMSE).

Untuk rumusan masalah yang akan dicakup dalam penelitian ini adalah bagaimana implementasi dan pengujian performansi sistem rekomendasi produk *e-commerce* pada dataset produk Amazon dengan menggunakan metode faktorisasi matriks berbasis *gradient descent* dalam menangani data yang memiliki tingkat *sparsity* yang tinggi. Diharapkan metode tersebut dapat menangani permasalahan *data sparsity*, sehingga dapat menghasilkan nilai RMSE yang optimal.

Terdapat beberapa batasan yang ada pada tugas akhir ini, yaitu tidak seluruh data pada domain produk *e-commerce* yang digunakan. Dataset yang digunakan diambil dari website *Stanford Network Analysis Project* (SNAP) dalam bentuk csv yang berisikan metadata mengenai produk pada *e-commerce* Amazon. Data yang digunakan berasal

dari kategori produk elektronik dan produk bahan makanan yang telah diberikan rating minimal oleh 50 *users*, dan *user* yang memberikan rating minimal terhadap 50 produk. Penggunaan kedua kategori dataset ini didasarkan oleh kategori dengan jumlah transaksi terbanyak pada *e-commerce* Amazon dengan tingkat data *sparsity* lebih dari 99%.

Tujuan

Tujuan dari tugas akhir ini adalah untuk pencarian nilai parameter optimal untuk mengimplementasikan metode faktorisasi matriks berbasis *gradient descent* pada sistem rekomendasi produk e-commerce dengan menggunakan dataset yang memiliki tingkat *sparsity* yang tinggi dan menguji performansi algoritma tersebut dalam penanganan masalah *data sparsity* dengan mengukur nilai RMSE pada data yang digunakan. Pengujian dilakukan terhadap dua dataset, yaitu data produk elektronik dan data produk bahan makanan dari website Amazon.

Organisasi Tulisan

Bagian-bagian selanjutnya pada tugas akhir ini akan memaparkan mengenai studi literatur terkait sistem rekomendasi yang akan di bangun pada bagian 2. Kemudian dilanjutkan dengan pembahasan mengenai perancangan dan pembangunan sistem rekomendasi produk pada bagian 3. Pada bagian 4 akan ditunjukkan hasil pengujian dan evaluasi sistem. Lalu pada bab 5 akan membahas kesimpulan dan saran dari penelitian tugas akhir ini.