

BAB I PENDAHULUAN

1.1 Latar Belakang

Sistem rekomendasi sering dimanfaatkan untuk mengatasi persoalan dengan mencari informasi relevan dari kumpulan informasi yang tersedia. Sistem rekomendasi umumnya diterapkan pada bidang-bidang yang memiliki jumlah data besar dan terus bertambah seiring berjalannya waktu. Sistem ini akan mengolah informasi pengguna yang selanjutnya memberikan rekomendasi sesuai dengan karakteristik pengguna tersebut yaitu sesuai dengan peminatannya. Salah satu metode yang digunakan dalam memberikan rekomendasi sesuai peminatan pengguna adalah *Collaborative Filtering*.

Collaborative Filtering (CF) merupakan metode yang populer. Namun CF menghadapi 2 masalah utama yaitu *sparsity* dan *scalability*. *Sparsity* merupakan kondisi di mana data yang belum terisi penuh atau datanya tidak lengkap. Jika data *rating* kondisinya jarang maka sulit untuk menentukan kesamaan antara pengguna dan dengan demikian akan mempengaruhi kualitas rekomendasi. *Scalability* merupakan kondisi di mana sistem rekomendasi perlu meningkatkan kekuatan komputasi mereka untuk menawarkan rekomendasi yang tepat waktu. Hal tersebut dilakukan dengan kondisi data dalam skala besar dan membutuhkan banyak sumber daya dan komputasi yang handal.

Beberapa penelitian telah dilakukan untuk mengetasi masalah tersebut seperti yang dilakukan oleh P. A. Riyaz dan S. M. Varghese mengusulkan Algoritma *Collaborative Filtering* (CF) diimplementasikan di Apache Hadoop dengan memanfaatkan *MapReduce* untuk *Bigdata* (*scalability*) [1]. Wu, et al. [2] mengusulkan rekomendasi *hybrid*. Pendekatan tersebut mampu meningkatkan kinerja rekomendasi dan mengatasi masalah *scalability*. Das, et al [3] mengkombinasi algoritma DBSCAN (*Density-Based Spatial Klustering of Applications with Noise*) dan Borda akan menghasilkan rekomendasi berupa film yang paling disukai sesuai dengan genrenya. Tang dan Tong [4],

mengusulkan BordaRank, dimulai dari mengisi *matriks* yang *sparsity* dengan memprediksi *rating* dan dilanjutkan proses agregasi sehingga menghasilkan *ranking* produk untuk direkomendasikan ke *user target*. Ardiansyah, MI et al [5], mengusulkan *Matrix Factorization* untuk mengisi nilai *rating* yang kosong untuk mengatasi masalah *sparse data rating*.

Selain itu Lestari, S., et al sudah melakukan penelitian untuk mengatasi *sparsity* dengan pendekatan *ranking based*. Lestari, S., et al [6] mengusulkan metode NRF (Normalized Rating Frequency) for *Collaborative Filtering* untuk mengatasi masalah *sparsity*. Lestari, S., et al [7] mengusulkan WP-Rank (*Weight Point Rank*) yang memaksimalkan penggunaan data peringkat untuk menghasilkan bobot produk. Hasil eksperimen menunjukkan bahwa metode *WP-Rank* lebih unggul daripada metode *Borda*.

Pada penelitian ini, *dataset* diklaster terlebih dahulu, sehingga data yang jumlahnya besar akan semakin cepat diproses sehingga waktu proses menjadi semakin singkat. Selain itu, penelitian ini akan melakukan pengisian nilai yang kosong (imputasi) untuk mengatasi *sparsity* pada *dataset* yang ada. Selanjutnya, penelitian ini akan menyelidiki pengaruh klasterisasi khususnya K-Means dan DBSCAN terhadap peningkatan kualitas *ranking* (WP-Rank) dan dievaluasi dengan metode NDCG (*Normalized Discounted Cumulative Gain*) Selain itu, penelitian ini akan menyelidiki pengaruh imputasi terhadap peningkatan kualitas *ranking* dan dievaluasi menggunakan NDCG.

1.2 Batasan Masalah

Penelitian ini menggunakan *dataset* publik *rating* film dari website *movielens.org* khususnya *dataset* 100k. Kumpulan data MovieLens ini dikumpulkan oleh Proyek Penelitian GroupLens di Universitas Minnesota. Kumpulan data ini terdiri dari 100.000 *rating* dari 943 pengguna terhadap 1682 buah film. Nilai *rating* yang tersedia yaitu 1 hingga 5. *Dataset* tersebut diklaster menggunakan metode K-Means dan DBSCAN menggunakan RapidMiner 9.9. Selanjutnya hasil masing – masing klaster diatasi *sparsity* nya menggunakan

teknik imputasi dengan fungsi rata – rata. Kemudian hasil imputasinya diranking menggunakan metode WP-Rank serta dievaluasi menggunakan metode NDCG.

1.3 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah:

Apakah *clustering* dan *ranking based* dapat mengatasi masalah *scalability* dan *sparsity* pada *Collaborative Filtering*?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk :

- a. Menganalisis dampak penerapan K-Means dan DBSCAN terhadap permasalahan *scalability* pada *Collaborative Filtering*; dan
- b. Menganalisis dampak penerapan WP-Rank terhadap permasalahan *sparsity* pada *Collaborative Filtering*.
- c. Menganalisis dampak imputasi terhadap kualitas rekomendasi yang ditunjukkan dengan nilai NDCG

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

- a. Ditemukan dampak penerapan K-Means dan DBSCAN terhadap permasalahan *scalability* pada *Collaborative Filtering*;
- b. Ditemukan dampak penerapan WP-Rank terhadap permasalahan *sparsity* pada *Collaborative Filtering*;
- c. Ditemukan dampak imputasi terhadap kualitas rekomendasi yang ditunjukkan dengan nilai NDCG