

INTISARI

SOLUSI MASALAH *SCALABILITY* DAN *SPARSITY* PADA *COLLABORATIVE FILTERING* DENGAN PENDEKATAN *CLUSTERING* DAN *RANKING BASED*

Oleh

MOHAMAD FAHMI HAFIDZ

fahmidzic2015@gmail.com

Seiring waktu berjalan, maka jumlah data semakin banyak. Karena itu, dibutuhkan suatu system rekomendasi yang cepat dan akurat. Salah satu system rekomendasi yang terkenal adalah *Collaborative Filtering*. Tetapi, *Collaborative Filtering* memiliki kelemahan di antaranya *Scalability* dan *Sparsity*. Penelitian ini menggunakan *dataset* Movielens 100k yang diklaster menggunakan metode K-Means dan DBSCAN kemudian diimputasi menggunakan nilai rata – rata. Selanjutnya hasilnya diranking menggunakan WP-Rank dan dievaluasi menggunakan metode NDCG. Hasil riset menunjukkan bahwa, *dataset* ini sebaiknya diklaster dengan K-Means dan DBSCAN menjadi 2 buah klaster. Pada metode K-Means, *Davies Bouldin Index* paling kecil yaitu 0,503 didapatkan pada klaster berjumlah 2 buah. Pada metode DBSCAN, *epsilon* terbaik bernilai 2. Hasil riset juga menunjukkan bahwa metode K-Means dan DBSCAN mampu mempersingkat waktu proses. Waktu proses Metode K-Means yaitu 1,5355 kali lipat lebih cepat daripada *dataset* tidak diklaster sedangkan Metode DBSCAN mampu mempersingkat waktu proses 1,0407 kali lipat daripada *dataset* tidak diklaster. Keduanya terbukti dapat mempercepat waktu proses sehingga mengatasi *scalability*. Selain itu, NDCG pada K-Means mengalami peningkatan, sedangkan NDCG pada DBSCAN mengalami penurunan daripada NDCG *dataset* aslinya. Hal ini menunjukkan bahwa K-Means lebih baik daripada DBSCAN dalam mengklaster *dataset* ini. Selanjutnya, Hasil klasterisasi K-Means menghasilkan persentase null sebesar 3,570% dan 13,032%. Hasil klasterisasi DBSCAN menghasilkan persentase null sebesar 0% dan 100%. Ini menunjukkan bahwa imputasi hasil klaster K-Means lebih baik daripada imputasi pada hasil klaster DBSCAN. Hal ini menunjukkan bahwa hasil klaster K-Means menghasilkan imputasi yang lebih baik daripada hasil klaster DBSCAN.

Kata Kunci : K-Means, DBSCAN, *Scalability*, WP-Rank, NDCG

ABSTRACT**SOLUTION OF SCALABILITY AND SPARSITY PROBLEMS
ON COLLABORATIVE FILTERING
WITH CLUSTERING AND RANKING BASED APPROACH**

By

MOHAMAD FAHMI HAFIDZ

fahmidzic2015@gmail.com

As time goes by, the amount of data increases. Therefore, we need a recommendation system that is fast and accurate. One of the well-known recommendation systems is Collaborative Filtering. However, Collaborative Filtering has weaknesses including Scalability and Sparsity. This study uses the 100k Movielens dataset which is clustered using the K-Means and DBSCAN methods then imputed using the average value. Then the results are ranked using WP-Rank and evaluated using the NDCG method. The research results show that this dataset should be clustered with K-Means and DBSCAN into 2 clusters. In the K-Means method, the smallest Davies Bouldin Index, which is 0.503, is obtained in 2 clusters. In the DBSCAN method, the best epsilon is worth 2. The research results also show that the K-Means and DBSCAN methods can shorten processing time. The processing time for the K-Means method is 1.5355 times faster than unclustered datasets, while the DBSCAN method can shorten the processing time by 1.0407 times than unclustered datasets. Both are proven to be able to speed up processing time so as to overcome scalability. In addition, NDCG on K-Means has increased, while NDCG on DBSCAN has decreased compared to the NDCG of the original dataset. This shows that K-Means is better than DBSCAN in clustering this dataset. Furthermore, the results of K-Means clustering produce null percentages of 3.570% and 13.032%. The DBSCAN clustering results produce a null percentage of 0% and 100%. This shows that the imputation of the K-Means cluster results is better than that of the DBSCAN cluster results. This shows that the K-Means cluster results produce better imputation than the DBSCAN cluster results.

Keyword : K-Means, DBSCAN, Scalability, WP-Rank, NDCG