

ABSTRACT

Recommendation system has benefitted users in finding product that match their character. But this kind of recommendation become less interesting when repetitively given. Users are expecting the recommendations to have *serendipity* criteria, i.e. novelty, unexpectedness, and relevance. Therefore, many researchers began to engage in the research of *serendipity* recommendation that usually uses collaborative filtering algorithms. Collaborative filtering algorithm is used because of its low level of complexity and its explainable decision. Collaborative filtering is mostly combined with kNN algorithm to produce accurate recommendation. The kNN algorithm support in selecting the most similar item using distance-based concept. In order to produce serendipitous recommendation, distance-based concept is applied using heirarchical clustering. However, heirarchical clustering cannot adapt with a large amount of data. Therefore, our proposed work uses K-Means Clustering algorithm to generate the clusters of items, which are obtained by collaborative filtering process. The process proceeds with finding a cluster that has potential *serendipity* among the generated clusters based on the distance between the clusters. This experiment using movielens *serendipity* dataset as an evaluation reference. The lowest value of MAE obtained from the proposed work with 13 variances of k is 0.239 at the k value of 9

Keywords: *Collaborative Filtering; serendipity; K-Means Clustering*

INTISARI

Rekomendasi yang sesuai dengan karakter konsumen telah banyak membantu dalam menemukan produk yang sesuai selera konsumen. Akan tetapi rekomendasi seperti ini menjadi tidak menarik apabila dilakukan secara berulang-ulang. Konsumen mengharapkan rekomendasi yang memiliki kriteria *serendipity*, antara lain *novelty*, *unexpectedness*, dan *relevance*. Oleh karena itu penelitian untuk mengembangkan sistem rekomendasi *serendipity* mulai banyak dilakukan. Pengembangan sistem rekomendasi *serendipity* pada umumnya menggunakan algoritma *collaborative filtering*. Algoritma ini dipilih karena memiliki kompleksitas yang rendah dan kemudahan dalam melakukan interpretasi hasil.

Algoritma *collaborative filtering* sering dikombinasikan dengan dengan algoritma kNN. Algoritma kNN merupakan algoritma yang menggunakan konsep jarak antar data untuk menghasilkan rekomendasi yang akurat dan memiliki tingkat kemiripan tinggi dengan data acuan. Akan tetapi, hal ini menjadikan algoritma kNN tidak sesuai untuk menghasilkan rekomendasi *serendipity*. *Hierarchical clustering* merupakan salah satu algoritma yang berdasarkan jarak antar data yang dapat menghasilkan rekomendasi *serendipity*. Akan tetapi algoritma ini memiliki keterbatasan karena hanya sesuai digunakan pada data yang memiliki hubungan hirarkis dan tidak sesuai untuk data yang berukuran besar. Oleh karena itu penelitian ini menggunakan algoritma pengklasteran *k-Means* untuk menghasilkan klaster dari data *collaborative filtering*. Klaster-klaster ini digunakan untuk menentukan rekomendasi *serendipity* dengan menggunakan jarak antar klaster. Rekomendasi yang dihasilkan diuji dengan menggunakan data *groundtruth* Movielens. Dari hasil variasi 13 jumlah klaster, diperoleh nilai MAE terendah yaitu 0,239 dari jumlah klaster 9.

Kata kunci -- *Collaborative Filtering; serendipity; K-Means Clustering*