

INTISARI

Pengaruh K-Means Clustering Pada Collaborative Filtering Dalam Metode Rekomendasi Anime

Oleh

Intan Hervianda Putri

19/439153/PA/18976

Anime merujuk kepada semua produk animasi yang berasal dari Jepang. Terdata pada dataset MyAnimeList, ada sebanyak 14.000 anime yang telah diproduksi hingga saat ini. Sistem rekomendasi dapat memberikan rekomendasi anime berdasarkan data profil pengguna. Salah satu metode sistem rekomendasi yang sering digunakan adalah *collaborative filtering* (CF). Ada beberapa pendekatan dalam CF, yaitu *user-based* CF dan *item-based* CF. Akan tetapi, metode ini memiliki kekurangan di bagian komputasi yang cukup memakan waktu, terutama saat mengolah dataset besar seperti MyAnimeList, sehingga dibutuhkan metode yang dapat mengurangi waktu komputasi model.

Penelitian membagi *user* ke dalam beberapa klaster menggunakan *k-means clustering* dan mencari prediksi *rating* menggunakan CF. Hasil dari beberapa percobaan membuktikan *k-means clustering* dapat mengurangi waktu inferensi model CF, dan semakin banyak klaster yang digunakan, semakin cepat model bekerja. *Clustering* juga dapat sedikit meningkatkan performa *user-based* CF tetapi cenderung mengurangi akurasi model *item-based* CF.

ABSTRACT

The Effect of K-Means Clustering on Collaborative Filtering in an Anime Recommendation

by

Intan Hervianda Putri

19/439153/PA/18976

The word anime refers to all animation originating from Japan. As recorded on the MyAnimeList dataset, there are as many as 14,000 anime that have been produced up to date. The recommendation system can provide anime recommendations based on user profile data. One recommendation system method that is often used is collaborative filtering (CF). There are several approaches to CF, namely user-based CF and item-based CF. However, this approach has a weakness in that the computation is quite time consuming, especially when processing large datasets such as MyAnimeList, so a method that can reduce the computing time is needed.

This research will split users into several clusters with k-means clustering and use collaborative filtering to compute for rating predictions. The results of several experiments shows that k-means clustering can decrease the inference time of both CF models, and the more clusters used, the faster the model works. Clustering can also slightly improve the performance of user-based CF, but instead tend to lower the accuracy for item-based CF model.