

Sistem Pemberi Rekomendasi Anime Menggunakan Pendekatan Hybrid

Alif Ranadian Nadhifah¹, Agung Toto Wibowo²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹alifnadhifah@students.telkomuniversity.ac.id, ²agungtoto@telkomuniversity.ac.id,

Abstrak

Sistem pemberi rekomendasi adalah alat dan perangkat software yang dapat memberikan solusi berupa saran suatu item yang bisa digunakan oleh user untuk mengambil sebuah keputusan untuk memilih item yang dikehendaki contohnya seperti memilih item anime. Sistem pemberi rekomendasi terbagi dalam beberapa kelas pendekatan rekomendasi yang berbeda diantaranya ada: collaborative filtering, content-based, dan hybrid recommender system. Namun perlu diketahui bahwa masing-masing kelas pendekatan tersebut memiliki kelemahan masing-masing. Sehingga penelitian pembangunan sistem pemberi rekomendasi anime ini memilih pendekatan hybrid filtering dengan tujuan mengatasi kelemahan dari masing-masing pendekatan yang berdiri sendiri. Penelitian ini terbatas hanya menggunakan 3 filtering untuk dikombinasikan, yaitu k-means clustering, TF-IDF dengan cosine similarity, dan item-based filtering dengan algoritma SVD. Dataset yang digunakan dalam penelitian ini merupakan dataset anime yang diambil melalui situs web kaggle yang sebelumnya telah dilakukan proses crawling dari situs myanimelist.net yang dibagikan secara open source. Akhir dari penelitian ini menemukan bahwa meskipun di beberapa penelitian lain pendekatan hybrid dapat meningkatkan kualitas rekomendasi ternyata sebaliknya. Masih ditemukan masalah-masalah seperti sparsity dan scalability dalam penelitian ini yang menyebabkan rekomendasi yang dihasilkan sistem masih kurang baik.

Kata kunci : Sistem Pemberi Rekomendasi, K-Means Clustering, TF-IDF, Item-based Filtering, SVD

Abstract

Recommendation systems are software and devices that can provide solutions in the form of item suggestions that can be used by users to make decisions about choosing the desired item, for example choosing an anime item. Comprehensive recommendation systems in several different recommendation approach classes include: Collaborative Filtering, Content-Based, and Hybrid Recommendation System. However, it should be noted that each of these approach classes has its own weaknesses. So this anime recommendation system development research chooses a hybrid filtering approach with the aim of overcoming the weaknesses of each stand-alone approach. This study is limited to using only 3 filters that will be combined, namely k-means clustering, TF-IDF with cosine similarity, and item-based filtering with the SVD algorithm. The dataset used in this study is an anime dataset taken from the kaggle website which has previously been crawled from the myanimelist.net site which is shared as open source. The end of this study found that although in several other studies the hybrid approach can improve the quality of recommendations, the opposite actually happened. Problems such as sparsity and scalability were still found in this study, which caused the recommendations produced by the system to be less than optimal.

Keywords: Recommender System, K-Means Clustering, TF-IDF, Item-based Filtering, SVD

1. Pendahuluan

Latar Belakang

Dunia hiburan saat ini berkembang pesat seimbang dengan berkembangnya teknologi. Bahkan dari waktu ke waktu juga bermunculan banyak aplikasi hiburan seperti aplikasi *streaming online*. Ada banyak hal yang ditawarkan dalam berbagai aplikasi *streaming online*, salah satunya adalah anime. Anime adalah salah satu hiburan yang berasal dari Jepang yang dapat dinikmati oleh seorang peminat anime yang bisa dikatakan sebagai seorang pengguna aplikasi tersebut. Dari tahun ke tahun, ada banyak anime yang diproduksi tiap tahunnya. Bagi pengguna yang menyukai menonton anime saat melihat fenomena ini pasti akan membuat mereka kebingungan dalam menentukan anime apa yang akan dipilih untuk ditontonnya. Maka dari itu, sistem pemberi rekomendasi dapat dijadikan sebagai pilihannya.

Sistem pemberi rekomendasi adalah alat dan perangkat *software* yang dapat memberikan solusi berupa saran suatu *item* yang bisa digunakan oleh *user* untuk mengambil sebuah keputusan untuk memilih *item* yang dikehendaki. Sistem pemberi rekomendasi beberapa kelas pendekatan rekomendasi yang berbeda diantaranya ada: *collaborative filtering*, *content-based*, dan *hybrid recommender system* [1][10].

Namun perlu diketahui bahwa masing-masing kelas pendekatan tersebut memiliki kelemahan masing-masing. Misalnya dalam pendekatan *collaborative filtering* yang paling populer dalam sistem pemberi rekomendasi memiliki masalah yang umum ditemukan, seperti *cold start*, *sparsity* dan *scalability* [5]. Untuk itu, pada penelitian ini dibangun sistem pemberi rekomendasi anime menggunakan pendekatan *hybrid filtering* yang mengkombinasikan beberapa pendekatan rekomendasi sebagai *filtering* rekomendasi anime. Apakah dengan dipilihnya pendekatan *hybrid filtering* dapat bisa meminimalisir ditemukannya masalah umum pada salah satu pendekatan *filtering* yang digunakan. Disamping itu, dataset yang digunakan dalam penelitian ini merupakan dataset anime yang didapat dari situs *web kaggle* yang sebelumnya telah dilakukan proses *crawling* dari situs *myanimelist.net* yang dibagikan secara *open source*.

Topik dan Batasannya

Penelitian ini dilakukan dengan mengimplementasikan pendekatan *hybrid filtering* pada sistem pemberi rekomendasi anime. Namun pendekatan *hybrid filtering* yang diterapkan terbatas dengan hanya menggunakan 3 *filtering* diantaranya *k-means clustering*, TF-IDF dengan *cosine similarity*, serta *item based collaborative filtering* dengan algoritma SVD.

Adapun dataset anime yang dimiliki terbagi dalam 3 data, diantara data *item* anime sebanyak 16216 *item*, data profil *user* sebanyak 47885 *user*, dan data ulasan *rating* sebanyak 130519 ulasan. Namun, dalam penelitian ini dataset anime yang digunakan hanya sebanyak 20% dari data *item* anime, 20% dari data profil *user*, dan data ulasan *rating* menyesuaikan dari data *item* anime dan data profil *user* yang tersisa. Pengurangan jumlah *dataset* ini adalah agar proses yang dibutuhkan untuk membangun sistem tidak memakan banyak waktu dan menghindari masalah *scalability*.

Tujuan

Penelitian ini bertujuan untuk mengimplementasikan pendekatan *hybrid filtering* pada sistem pemberi rekomendasi anime untuk bisa melihat apakah dengan melakukan kombinasi berbagai pendekatan *filtering* bisa meminimalisir ditemukannya masalah umum pada salah satu pendekatan *filtering* yang digunakan. Adapun hasil penelitian dilihat melalui performansi akurasi terhadap sistem yang dibangun menggunakan metrik evaluasi, seperti *recall*, *precision*, dan *f1-score*.

Organisasi Tulisan

Struktur penulisan jurnal penelitian setelah sub-bab pendahuluan akan dilanjutkan dengan pembahasan studi terkait pada sub-bab kedua. Kemudian pada sub-bab ketiga disajikan pembahasan mengenai sistem yang dibangun. Pada sub-bab kelima disajikan hasil penelitian dan pembahasan analisa dari hasil penelitian. Dan pada sub-bab kelima menyajikan kesimpulan dari penelitian yang telah dilakukan.

2. Studi Terkait

2.1 Sistem Pemberi Rekomendasi

Sistem pemberi rekomendasi adalah sebuah sistem berupa perangkat lunak yang dibangun untuk menghasilkan keluaran berupa sejumlah *item* teratas berdasarkan nilai yang telah diperhitungkan menggunakan sejumlah algoritma. Adapun sejumlah *item* yang dihasilkan tersebut merupakan keluaran atas permintaan pengguna yang membutuhkan dan sebelumnya pengguna telah memberikan kriteria tertentu dari sejumlah *item* yang diminta. Sistem pemberi rekomendasi yang digunakan dalam penelitian itu bersifat *personalized*, yaitu pengguna yang berbeda akan mendapatkan saran rekomendasi *item* yang beragam dan ditampilkan dalam bentuk urutan peringkat *item* [10].

Terdapat 3 jenis objek data yang digunakan dalam sistem pemberi rekomendasi, yaitu *users*, *items*, dan *transactions*. *Users* adalah pengguna sistem pemberi rekomendasi. Berbagai informasi mengenai *user* perlu dieksploitasi oleh sistem pemberi rekomendasi agar bisa mempersonalisasi rekomendasi dan interaksi manusia-komputer. Sedangkan *items* adalah objek yang direkomendasikan. Karakter *item* mungkin bisa dilihat dari kompleksitas dan nilai kegunaannya. Misalnya jika *item* bernilai positif maka *item* berguna bagi *user* atau sebaliknya jika *item* bernilai negatif maka *item* tidak cocok karena *user* mungkin membuat kesalahan dalam menentukan pilihannya. Sedangkan *transactions* adalah data yang digunakan oleh sistem dalam algoritma pembuatan rekomendasi dan berupa seperti *log* penyimpanan informasi penting yang dihasilkan dari interaksi manusia-komputer. Data *transactions* paling populer yang dikumpulkan sistem pemberi rekomendasi adalah data *rating item* yang diberikan oleh *user* dan dapat dikumpulkan secara *explicit* atau *implicit* [10].