

Recommender System Berbasis Hybrid Filtering di Twitter Menggunakan K-Means Clustering (Studi Kasus: Film di Disney+)

Farid Krida Mukti¹, Erwin Budi Setiawan²

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

¹faridkrida@students.telkomuniversity.ac.id, ²erwinbudisetiawan@telkomuniversity.ac.id,

Abstrak

Dengan perkembangan teknologi yang sudah sangat pesat, sehingga menonton film di rumah sudah menjadi sarana hiburan. Disney+ adalah platform menonton film yang menyediakan berbagai berbagai judul film. Namun, karena ada terlalu banyak film terlalu banyak judul film, hal ini menyebabkan kebingungan di antara para pengguna. Untuk menentukan yang mana film untuk ditonton. Solusi untuk masalah ini adalah dengan menyediakan sebuah sistem rekomendasi yang memberikan rekomendasi untuk film yang akan ditonton. Twitter adalah sebuah media sosial yang sosial yang digunakan untuk menulis posting yang disebut *tweet*. Dalam sistem ini, *tweet* merupakan data yang akan diolah menjadi rating. Penelitian ini dilakukan menggunakan *K-Means Clustering* dan *Hybrid Filtering*. Dengan menggunakan dataset yang diperoleh dari Kaggle dataset berisi IMDb, Disney Rotten Tomatoes, dan Metacritic. Kemudian melakukan preprocessing data dengan *text processing*, *polarity*, dan *labeling*. Dan mendapatkan dataset yang akan digunakan untuk eksperimen. Hasil pengujian dari penelitian ini menunjukkan bahwa *K-Means Clustering* dengan *Hybrid Filtering* mendapatkan hasil yang paling baik dengan mendapatkan nilai MAE sebesar 0.4236, dan nilai RMSE sebesar 0.5246.

Kata kunci : sistem rekomendasi, *hybrid filtering*, *collaborative filtering*, *content-based filtering*, *k-means clustering*.

Abstract

With the rapid development of technology, watching movies at home has become a means of entertainment. Disney+ is a movie-watching platform that provides a wide range of movie titles. However, since there are too many movies too many movie titles, it causes confusion among users. To decide which movie to watch. The solution to this problem is to provide a recommendation system that gives recommendations for movies to watch. Twitter is a social social media that is used to write posts called tweets. In this system, tweets are the data that will be processed into ratings. This research is conducted using K-Means Clustering and Hybrid Filtering. By using a dataset obtained from Kaggle dataset containing IMDb, Disney Rotten Tomatoes, and Metacritic. Then preprocessing the data with text processing, polarity, and labeling. And get a dataset that will be used for experiments. The test results of this study show that K-Means Clustering with Hybrid Filtering gets the best results by getting an MAE value of 0.4236, and an RMSE value of 0.5246.

Keywords: Recommendation System, *hybrid filtering*, *collaborative filtering*, *content-based filtering*, *k-means clustering*.

1. Pendahuluan

Latar Belakang

Dengan berkembangnya revolusi industri 4.0, jejaring sosial selama beberapa tahun mengalami perkembangan. Media sosial sendiri adalah salah satu alat yang selalu digunakan untuk mengekspresikan berbagai kegiatan dan menyampaikan laporan [1]. Twitter adalah salah satu *platform* media sosial paling populer yang memberikan pengguna untuk membuat cuitan singkat "*Tweet*", mencari informasi populer "*Trending*" dan bisa berkomunikasi secara singkat "*Direct Message*" [2]. Karena fungsinya, banyak pengguna Twitter yang hobi menonton film menggunakan media sosial untuk mencari informasi dan memberikan ulasan terkait film tersebut.

Film merupakan salah satu media hiburan yang menawarkan berbagai macam judul dan *genre* film. Berkat revolusi industri 4.0, kemudian akses terhadap berbagai macam film berkembang pesat. Banyak Perusahaan di industri film telah melakukan pasar *direct-to-customer* untuk menikmati kenyamanan hiburan internet [3]. Salah satu layanan yang sering digunakan adalah Disney+. Disney+ memungkinkan pengguna untuk menonton berbagai judul dan *genre*, bisa menambahkan film ke dalam *watchlist*, dan mengunduh film [2]. Dengan banyaknya judul dan *genre* yang tersedia di platform Disney+, diperlukan sistem rekomendasi untuk memudahkan pengguna dalam memilih film berdasarkan judul dan *genre* yang disukai. Di samping itu dapat diterapkan pada platform streaming film, sistem rekomendasi juga dapat diterapkan pada *platform* lain seperti musik, TV, *e-commerce* dan lainnya [4].

Sistem rekomendasi adalah sistem yang dapat membantu mengatasi banjir informasi dengan memberikan rekomendasi khusus kepada pengguna, dan rekomendasi tersebut diharapkan dapat memenuhi keinginan dan

kebutuhan pengguna [5]. Rekomendasi sistem memiliki beberapa metode yaitu *content-based*, *collaborative filtering* dan *hybrid based* [6]. Salah satu metode sistem rekomendasi yang akan digunakan adalah *collaborative filtering*. *Collaborative filtering* adalah sistem rekomendasi yang menggabungkan semua pengguna untuk memilih produk yang sama, seperti film, berdasarkan rating pengguna [7]. *Collaborative filtering* dibagi menjadi dua bagian, yaitu *user based* dan *item based*. *Collaborative filtering* adalah algoritma sistem rekomendasi yang paling sukses dan populer, tetapi memiliki akurasi yang buruk dan waktu berjalan yang lama, sehingga diperlukan clustering untuk mengatasi masalah ini [8]. Dengan adanya masalah pada *collaborative filtering* yang ditemukan, kami menggunakan salah satu dari cluster untuk menyelesaikannya menggunakan *K-Means clustering*.

Oleh karena itu pada penelitian ini kami akan menguji sistem rekomendasi yang dibangun dengan menggunakan *Hybrid Filtering* dengan metode *K-Means Clustering*. Diharapkan dengan adanya sistem ini rekomendasi film Disney+, penonton akan menemukan film yang sesuai dengan minat mereka, dan akan mendapatkan perangkat performa yang lebih akurat dan tepat.

Topik dan Batasannya

Topik pada penelitian ini yang diteliti oleh penulis adalah mengembangkan metode *hybrid filtering* yang digabungkan dengan *k-means clustering*. Keterbatasan dalam penelitian ini adalah pada Dataset bersumber dari website Kaggle dan menggunakan judul film sebagai crawling data di Twitter. Serta tweet yang di crawling pada Twitter adalah tweet yang menggunakan Bahasa Indonesia.

Tujuan

Tujuan dari penelitian ini adalah untuk menerapkan sistem penggabungan teknik *Hybrid Filtering* dengan metode *K-Means Clustering*. Dengan harapan penerapan klasifikasi menggunakan *K-Means Clustering* setelah diproses menggunakan *Hybrid Filtering* dapat menghasilkan model rekomendasi film yang baik dan dapat memberikan prediksi yang akurat untuk film yang direkomendasikan dan tidak direkomendasikan dari proses penambahan metode klasifikasi.

Organisasi Tulisan

Struktur penelitian pada jurnal ini setelah bagian pendahuluan, bagian yang akan dibahas selanjutnya adalah studi terkait pada bagian 2. Dilanjutkan bagian 3 menyajikan teori dan perancangan sistem yang dibangun. Kemudian bagian 4 menyajikan hasil yang didapatkan dan hasil analisis. Pada bagian akhir yaitu bagian 5 berisikan kesimpulan dari penelitian yang telah dilakukan.

2. Studi Terkait

Bagi Pada penelitian ini penulis menggunakan berbagai jenis referensi penelitian yang pernah dilakukan sebelumnya. Referensi penelitian yang dimaksud agar penulis mendapatkan pengetahuan dari penelitian sebelumnya dan dapat sebagai pembanding dan pedoman pada penulisan Tugas Akhir. Referensi penelitian yang dituju berupa segi penelitian, metode, dan kesimpulan yang diperoleh dari penelitian tersebut.

Dari banyaknya penelitian terkait dengan Sistem Rekomendasi menggunakan *Collaborative Filtering* dan ditambah dengan klasifikasi *K-Means*. Salah satunya penelitian yang ditulis oleh Phorasim, Phongsavanh, and Lasheng Yu. Sharma dengan judul penelitian "Movies Recommendation system using Collaborative Filtering and K-Means" [10]. Pada penelitian tersebut menggunakan *Collaborative Filtering* untuk rekomendasi sistem pada restoran dan membandingkan algoritma *Collaborative* dan *K-Means* pada klasifikasi. Untuk dataset yang digunakan bersumber dari Phorasim dengan data berjumlah 1.093.360 ulasan oleh 162.541 pengguna. Berdasarkan hasil uji algoritma KNN dalam uji MSE memperoleh hasil direntang 0.94 – 1.2 dengan memasukan nilai K dari 2 hingga 50 sementara itu pada algoritma *K-Means* dalam uji MSE memperoleh hasil direntang 0.70 – 0,80 yang menunjukkan bahwa *user-based K-Means* dan *item-based K-Means* mengungguli KNN.

Penelitian yang selanjutnya yang ditulis oleh M. Irawan, dll berjudul "User-Based dan Item-Based Collaborative Filtering pada Recommender System Buku dengan Metode Naïve Bayes Classification" [11]. Pada penelitian tersebut membandingkan *Collaborative Filtering* dengan *Collaborative Filtering* dikombinasikan dengan *Naïve Bayes Classification* kemudian diimplementasikan pada *user-based* dan *item-based*. Penelitian menggunakan dataset goodbooks-10k dari website kaggle dengan jumlah data rating 981.756, data users 53424 dan data buku 10000. Berdasarkan hasil yang diperoleh dalam menghitung nilai precision dan recall. Menghasilkan bahwa *Collaborative Filtering* dikombinasikan dengan *Naïve Bayes Classification* lebih baik dari *Collaborative Filtering* saja berdasarkan nilai performansi pada setiap metode [9].

Berdasarkan penelitian Yessica Putri Santoso dll. Dengan topik "Implementasi metode K-Means Clustering pada Sistem Rekomendasi Dosen Tetap Berdasarkan Penilaian Dosen". Menampilkan 1 dari 70 data yang diuji terdapat 39 data pembicara yang dapat direkomendasikan sebagian pembicara tetap dan 31 data presenter tidak memenuhi syarat menjadi dosen tetap karena hasil penghitungan akurasi sebesar 55.67% sehingga bisa disimpulkan bahwa algoritma *K-Means* tidak cocok untuk kasus ini [12].