

ABSTRAK

Hilmi, Fauzan Nafazi Zul. 2026. Sistem Rekomendasi Anime Menggunakan Pendekatan *Hybrid* dengan Algoritma *Random Forest* dan *K-Nearest Neighbor* (KNN). Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing: Dr. Ilham Saifudin, S.Pd., M.Si.; Dudi Irawan, S.T., M.Kom.

Meningkatnya jumlah judul anime dengan variasi genre, tipe, *studio*, dan segmentasi usia menyebabkan proses pencarian tontonan yang sesuai preferensi pengguna menjadi semakin kompleks, sehingga dibutuhkan sistem rekomendasi yang mampu menyaring pilihan konten secara relevan dan efisien. Penelitian-penelitian terdahulu umumnya menerapkan algoritma klasifikasi dan algoritma perhitungan kemiripan secara terpisah, sehingga integrasi antara algoritma *random forest* dan *K-Nearest Neighbor* (KNN) dalam sistem rekomendasi anime berbasis *content-based filtering* masih belum banyak dikaji. Berdasarkan celah tersebut, rumusan masalah dalam penelitian ini adalah seberapa besar kinerja pendekatan *weighted hybrid* yang mengintegrasikan skor klasifikasi *random forest* dan skor kemiripan KNN dalam menghasilkan rekomendasi anime, ditinjau dari nilai *Normalized Discounted Cumulative Gain* (NDCG), *precision*, dan waktu komputasi. Berdasarkan rumusan masalah tersebut, penelitian ini bertujuan menganalisis kinerja pendekatan *weighted hybrid* yang mengintegrasikan skor klasifikasi algoritma *random forest* dan skor kemiripan *K-Nearest Neighbor* (KNN) dalam menghasilkan rekomendasi anime, berdasarkan metrik NDCG, *precision*, dan waktu komputasi.

Dataset yang digunakan merupakan data sekunder dari situs Kaggle yang terdiri atas 24.905 *records*, meliputi atribut *name*, *genres*, *type*, *studios*, *rating*, dan *score*. Atribut *score* digunakan sebagai dasar pembentukan label kelayakan konten ("Layak" dan "Tidak Layak") melalui pendekatan *mean thresholding* pada tahap *pre-processing*, dilanjutkan dengan transformasi fitur menggunakan *Multi-Label Binarizer* dan *Label Encoder*. Algoritma *random forest* dilatih untuk mengklasifikasikan kelayakan anime berdasarkan atribut *genres*, *type*, *studios*, dan *rating*, sedangkan algoritma KNN digunakan untuk menghitung tingkat kemiripan antar anime berdasarkan fitur *genres* menggunakan *cosine similarity*. Kedua skor tersebut digabungkan melalui skema pembobotan (*weighted hybrid*) untuk menghasilkan urutan rekomendasi akhir yang ditampilkan kepada pengguna melalui antarmuka berbasis *website*.

Evaluasi dilakukan pada tiga skenario pembagian data (70:30, 80:20, dan 90:10) menggunakan skema pengujian Top-K (K=10, 20, dan 30) serta evaluasi global. Hasil pengujian menunjukkan bahwa sistem memperoleh nilai *precision* tertinggi sebesar 85% dan NDCG sebesar 94,9% pada pengujian Top-K, serta nilai *precision* sebesar 75,6% dan NDCG sebesar 97,7% pada evaluasi keseluruhan, yang mengindikasikan bahwa urutan rekomendasi yang dihasilkan sudah mendekati urutan ideal yang diharapkan. Dari sisi efisiensi, sistem menghasilkan rekomendasi dengan waktu pemrosesan kurang dari 2,5 detik pada pengujian Top-K dan waktu tercepat sebesar 4,464 detik pada skenario pembagian data 90:10.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa pendekatan *weighted hybrid* yang mengintegrasikan skor klasifikasi *random forest* dan skor kemiripan KNN terbukti memiliki kinerja yang baik dalam menghasilkan rekomendasi anime, baik dari segi kualitas rekomendasi (NDCG dan *precision*) maupun efisiensi waktu komputasi, sehingga pendekatan ini efektif digunakan sebagai metode rekomendasi anime.

Kata kunci: *Cosine Similarity*, *K-Nearest Neighbor*, NDCG, *Random Forest*, Sistem Rekomendasi Anime, *Weighted Hybrid*



ABSTRACT

Hilmi, Fauzan Nafazi Zul. 2026. *Anime Recommendation System Using a Hybrid Approach with Random Forest and K-Nearest Neighbor (KNN) Algorithms*. Undergraduate Thesis. Bachelor's Program in Informatics Engineering. University of Muhammadiyah Jember.

Advisors: Dr. Ilham Saifudin, S.Pd., M.Si.; Dudi Irawan, S.T., M.Kom.

The growing number of anime titles with diverse genres, types, studios, and age segmentations has made the process of finding content that matches user preferences increasingly complex, creating a need for a recommendation system capable of filtering relevant and efficient content choices. Previous studies have generally applied classification algorithms and similarity-based algorithms separately, leaving the integration of random forest and K-Nearest Neighbor (KNN) algorithms in a content-based anime recommendation system largely unexplored. Based on this gap, the research problem addressed in this study is how well the weighted hybrid approach, which integrates the classification score of random forest and the similarity score of KNN, performs in generating anime recommendations, as measured by Normalized Discounted Cumulative Gain (NDCG), precision, and computational time. Accordingly, this study aims to analyze the performance of the weighted hybrid approach that integrates the classification score of the random forest algorithm and the similarity score of the K-Nearest Neighbor (KNN) algorithm in generating anime recommendations, based on the NDCG, precision, and computational time metrics.

The dataset used is secondary data obtained from Kaggle, consisting of 24,905 records with attributes including name, genres, type, studios, rating, and score. The score attribute serves as the basis for constructing a content eligibility label ("Eligible" and "Not Eligible") using a mean thresholding approach during the preprocessing stage, followed by feature transformation using Multi-Label Binarizer and Label Encoder. The random forest algorithm is trained to classify anime eligibility based on the genres, type, studios, and rating attributes, while the KNN algorithm is used to calculate the similarity between anime based on the genres feature using cosine similarity. Both scores are then combined through a weighted hybrid scheme to generate a final ranking of recommendations presented to users through a website-based interface.

Evaluation was carried out using three data-splitting scenarios (70:30, 80:20, and 90:10) with Top-K testing schemes ($K = 10, 20$, and 30) as well as a global evaluation. The experimental results show that the system achieved a highest precision of 85% and an NDCG of 94.9% in the Top-K testing, as well as a precision of 75.6% and an NDCG of 97.7% in the overall evaluation, indicating that the resulting recommendation ranking closely approximates the ideal ranking. In terms of efficiency, the system generated recommendations with a processing time of less than 2.5 seconds in the Top-K testing, with the fastest time of 4.464 seconds recorded in the 90:10 data-splitting scenario.

Based on these results, it can be concluded that the weighted hybrid approach integrating the classification score of random forest and the similarity score of KNN demonstrates good performance in generating anime

recommendations, both in terms of recommendation quality (NDCG and precision) and computational time efficiency, making this approach effective for use as an anime recommendation method.

Keywords: *Cosine Similarity, K-Nearest Neighbor, NDCG, Random Forest, Anime Recommendation System, Weighted Hybrid*

