

ABSTRAK

Pada era digital ini, layanan streaming film menyediakan jumlah konten yang sangat besar sehingga menimbulkan fenomena kelebihan informasi (information overload), di mana pengguna mengalami kesulitan dalam menemukan film yang sesuai dengan preferensi mereka. Masalah tersebut dapat diatasi dengan menggunakan sistem rekomendasi. Terdapat dua metode populer yang digunakan, yaitu collaborative filtering yang rekomendasinya berdasarkan kemiripan antar perilaku pengguna dan content-based filtering yang berdasarkan kemiripan antar item. Collaborative filtering memiliki masalah cold start dan sparsity data, di mana pemberian rekomendasi tidak akurat pada data yang memiliki pengguna baru dan film baru. Sedangkan content-based filtering memiliki kelemahan overspecialization karena rekomendasinya bergantung pada kemiripan fitur item.

Kedua masalah tersebut dapat ditangani dengan metode hybrid yang menggabungkan beberapa metode sekaligus. Metode hybrid yang digunakan pada penelitian ini menggabungkan collaborative filtering berbasis Singular Value Decomposition (SVD) dengan Gaussian Mixture Models (GMM) untuk content-based clustering serta dilengkapi mekanisme adaptive weighting. GMM melakukan soft clustering pada film berdasarkan genre menggunakan smoothed one-hot encoding untuk merepresentasikan multi-genre film secara akurat. Adaptive weighting memberikan bobot lebih besar pada GMM ($\alpha=0,20$) untuk pengguna baru dan meningkatkan bobot SVD ($\alpha=0,50$) seiring bertambahnya riwayat rating pengguna.

Pengujian menggunakan Root Mean Squared Error (RMSE) dan Mean Absolute Error (MAE) untuk mengukur tingkat error, Top-K overlap untuk mengukur kesamaan daftar rekomendasi, serta uji paired t-test untuk signifikansi statistik. Hasil pengujian menunjukkan model hybrid berhasil menurunkan RMSE sebesar 1,04% (dari 0,968 menjadi 0,957) dan MAE sebesar 2,82% (dari 0,769 menjadi 0,747). Pada validasi silang 5-fold, model hybrid mencapai rata-rata RMSE 0,945 (peningkatan 1,01%) dan MAE 0,737 (peningkatan 2,9%). Uji paired t-test menghasilkan p-value 0,0009 (RMSE) dan 0,00001 (MAE) yang $< 0,05$, mengonfirmasi peningkatan akurasi signifikan secara statistik. Untuk pengguna baru, model hybrid menghasilkan 100% rekomendasi sesuai genre favorit dengan nilai Top-K overlap 0% pada $K=5$.

Kata kunci: Sistem rekomendasi film, collaborative filtering, Gaussian Mixture Models, Singular Value Decomposition, hybrid, adaptive weighting, cold start

ABSTRACT

In this digital era, movie streaming services provide a massive amount of content, leading to an information overload phenomenon where users face difficulties in finding movies that match their preferences. This issue can be addressed by implementing a recommendation system. Two popular methods are widely used: collaborative filtering, which provides recommendations based on similarities in user behavior, and content-based filtering, which is based on item feature similarities. Collaborative filtering suffers from cold start and data sparsity problems, where recommendations become inaccurate when dealing with new users and new movies. On the other hand, content-based filtering has an overspecialization drawback because its recommendations strictly depend on the similarity of item features.

Both limitations can be mitigated using a hybrid approach that combines multiple methods simultaneously. The hybrid method employed in this study integrates Singular Value Decomposition (SVD)-based collaborative filtering with Gaussian Mixture Models (GMM) for content-based clustering, enhanced by an adaptive weighting mechanism. GMM performs soft clustering on movies based on genres using smoothed one-hot encoding to accurately represent multi-genre movies. The adaptive weighting mechanism assigns a higher weight to GMM ($\alpha=0.20$) for new users and increases the SVD weight ($\alpha=0.50$) as the user's rating history grows.

Evaluation was conducted using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) to measure error rates, Top-K overlap to measure the similarity of recommendation lists, and a paired t-test for statistical significance. The experimental results show that the hybrid model successfully reduced the RMSE by 1.04% (from 0.968 to 0.957) and the MAE by 2.82% (from 0.769 to 0.747). In the 5-fold cross-validation, the hybrid model achieved an average RMSE of 0.945 (a 1.01% improvement) and an MAE of 0.737 (a 2.9% improvement). The paired t-test yielded p – values of 0.0009 for RMSE and 0.00001 for MAE (both < 0.05), confirming that the accuracy improvement is statistically significant. For new users, the hybrid model generated 100% recommendations matching their favorite genres with a Top-K overlap value of 0% at $K=5$.

Keywords: *Movie recommendation system, collaborative filtering, Gaussian Mixture Models, Singular Value Decomposition, hybrid, adaptive weighting, cold start*