

Evaluasi Komparatif Hybrid Filtering dan Model-Based SVD pada Sistem Rekomendasi Film Menggunakan Dataset MovieLens

Angelina Putri Ariani¹, Dita Ayu Handayani², Putri Muryanti Setyowati³, Deni Kurnianto Nugroho^{4*}, Marwan Noor Fauzy⁵

Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

¹angelinaputrii5@students.amikom.ac.id, ²ditahandayani@students.amikom.ac.id, ³putrimuryanti@students.amikom.ac.id,
⁴deni@amikom.ac.id*, ⁵marwannoorfauzy@amikom.ac.id

Abstract

The increasing number and variety of films have made it difficult for users to choose movies that suit their preferences. Therefore, recommendation systems have become an important solution to help users obtain relevant and personalized movie recommendations. This study aims to implement and evaluate the performance of recommendation systems using Item-Based Collaborative Filtering, Content-Based Filtering, Hybrid Filtering, and Model-Based Collaborative Filtering approaches on the MovieLens 100k dataset. Specifically, the Model-Based approach utilizes the Singular Value Decomposition (SVD) algorithm, optimized with a learning rate of 0.005, a regularization term of 0.02, and 50 latent factors. The main scientific contribution of this study is a comparative analysis between memory-based and model-based approaches in handling high-sparsity datasets. The research process includes data preprocessing, data separation using the random holdout method, and data exploration to analyze rating distribution. System performance evaluation was conducted using Root Mean Square Error (RMSE) and Precision@10 metrics. The results showed that the Model-Based SVD method produced the best performance with the lowest RMSE value of 0.877 and the highest Precision@10 of 67.45%. Meanwhile, the Hybrid Filtering method using a manual weighting scheme failed to improve performance, yielding a Precision@10 of only 0.40%; this low performance is attributed to the rigidity of static weights, which could not effectively balance the biases of the underlying models. These results indicate that machine learning-based approaches with latent models are more effective in handling datasets with high sparsity levels compared to content-based or hybrid methods with static weights.

Keywords: Recommendation System, Hybrid Filtering, Collaborative Filtering, Content-Based Filtering, Singular Value Decomposition.

Abstrak

Perkembangan jumlah dan variasi film yang semakin meningkat menyebabkan pengguna mengalami kesulitan dalam menentukan pilihan tontonan yang sesuai dengan preferensi mereka. Oleh karena itu, sistem rekomendasi menjadi solusi penting untuk membantu pengguna memperoleh rekomendasi film yang relevan dan personal. Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi kinerja sistem rekomendasi menggunakan pendekatan Item-Based Collaborative Filtering, Content-Based Filtering, Hybrid Filtering, serta Model-Based Collaborative Filtering pada dataset MovieLens 100k. Secara khusus, pendekatan Model-Based menerapkan algoritma Singular Value Decomposition (SVD) yang dioptimalkan dengan learning rate 0.005, regularisasi 0.02, dan 50 faktor laten. Penelitian ini memberikan kontribusi ilmiah berupa analisis komparatif antara pendekatan memory-based dan model-based pada dataset dengan tingkat sparsity tinggi. Proses penelitian meliputi tahapan preprocessing data, pemisahan data menggunakan metode random holdout, dan eksplorasi data untuk menganalisis distribusi rating. Evaluasi kinerja sistem dilakukan menggunakan metrik Root Mean Square Error (RMSE) dan Precision@10. Hasil penelitian menunjukkan bahwa metode Model-Based SVD menghasilkan performa terbaik

dengan nilai RMSE terendah sebesar 0.877 dan Precision@10 tertinggi sebesar 67.45%. Sementara itu, metode Hybrid Filtering yang menggunakan skema pembobotan manual menghasilkan kinerja rendah dengan Precision@10 sebesar 0.40%; hal ini disebabkan oleh penggunaan bobot statis yang tidak mampu mengakomodasi variasi bias antar-model secara efektif. Hasil ini menunjukkan bahwa pendekatan berbasis machine learning dengan model laten lebih efektif dalam menangani dataset sparse dibandingkan metode berbasis konten maupun hybrid konvensional dengan bobot statis.

Kata kunci: sistem rekomendasi, hybrid filtering, collaborative filtering, content-based filtering, singular value decomposition

© 2026 Author
Creative Commons Attribution 4.0 International License



1. Pendahuluan

Pada era sekarang konsumsi media hiburan, khususnya film mengalami peningkatan. Jumlah film tiap tahun mencapai ribuan dengan genre yang beragam. Menciptakan rekomendasi yang variasi [1]. Dengan adanya hal tersebut, penggemar film seringkali kesulitan dalam menentukan pilihan film yang akan ditonton sesuai dengan preferensi mereka. Salah satu solusi dari permasalahan ini adalah dengan membangun sistem rekomendasi untuk menciptakan rekomendasi konten yang relevan secara efisien. Sistem rekomendasi tidak hanya meningkatkan pengalaman pengguna, tetapi juga meningkatkan nilai bisnis pada platform penyedia layanan hiburan.

Umumnya sistem rekomendasi tradisional menggunakan dua metode, yaitu collaborative filtering dan content-based filtering. Collaborative filtering umumnya terbagi menjadi dua kategori yaitu user-based dan item-based [2]. Item-based collaborative filtering bekerja dengan memberikan rekomendasi item serupa dari hasil menganalisis kesamaan antar item berdasarkan pola rating pengguna. Sementara content-based filtering berfokus pada karakteristik item yang pernah disukai pengguna. Kedua metode ini memiliki keterbatasan. Item-based filtering menghadapi masalah cold start untuk pengguna atau item yang belum memiliki riwayat rating, serta masalah sparsity pada matriks user-item. Content-based filtering sendiri memiliki keterbatasan pada rekomendasi item yang serupa dengan yang pernah dilihat pengguna, sehingga kurang mampu memberikan rekomendasi yang beragam dan mengeksplorasi preferensi pengguna [3].

Untuk mengatasi keterbatasan dari dua metode tersebut, maka digunakanlah pendekatan hybrid filtering yang menggabungkan kelebihan berbagai metode. Sistem hybrid dapat menggabungkan item-based collaborative filtering dan content-based filtering dengan berbagai strategi seperti weighted sehingga mampu meminimalkan kelemahan masing-masing metode dan memaksimalkan keakuratan rekomendasi [4]. Kombinasi item-based collaborative

filtering yang memanfaatkan pola preferensi pengguna dan content-based yang menganalisis karakteristik konten dapat menghasilkan rekomendasi yang lebih personal. Seiring kemajuan machine learning, model-based collaborative filtering menggunakan algoritma matrix factorization seperti Singular Value Decomposition (SVD) telah menunjukkan kinerja sistem yang lebih baik dalam memprediksi rating pengguna. SVD bekerja dengan membagi matriks user-item menjadi matriks laten berdimensi rendah yang menangkap faktor-faktor tersembunyi dalam preferensi pengguna dan karakteristik item sehingga mampu mengatasi sparsity dan menghasilkan prediksi yang lebih akurat[6]. Penerapan SVD sebagai model-based membuka peluang untuk menemukan pola kompleks dari data interaksi user-item yang sebelumnya tidak teridentifikasi oleh metode memory-based.

Dataset MovieLens telah menjadi patokan umum yang digunakan dalam penelitian sistem rekomendasi karena menyediakan data rating film yang luas dan real-world dari ribuan pengguna. Dataset ini memiliki data sparsity yang tinggi, distribusi rating yang tidak seimbang, dan variasi preferensi pengguna yang tinggi. Oleh karena itu, MovieLens cocok digunakan untuk menguji seberapa akurat suatu algoritma rekomendasi bekerja. Penelitian-penelitian sebelumnya telah menggunakan dataset ini telah melakukan perbandingan antara metode rekomendasi. Meskipun metode collaborative filtering dan hybrid telah banyak diteliti, terdapat kesenjangan penelitian yang perlu ditangani. Mayoritas penelitian sebelumnya cenderung memfokuskan evaluasi pada optimalisasi satu pendekatan spesifik secara terisolasi baik itu hanya item-based, hanya content-based, atau hanya model SVD tanpa melakukan analisis komparatif yang menyeluruh perbandingan metode tersebut pada dataset yang sama. Belum banyak studi yang secara eksplisit membandingkan efektivitas strategi hybrid sederhana (pembobotan manual) melawan pendekatan model-based (SVD) khususnya pada kondisi data dengan tingkat sparsity yang ekstrem (di atas 90%). Ketidaaan perbandingan langsung ini

menyulitkan pemahaman mengenai strategi mana yang lebih robust: apakah penggabungan metode konvensional (hybrid) atau penggunaan model laten (model-based) dalam mengatasi kelangkaan data rating.

Studi oleh Yudha dkk, menunjukkan bahwa metode hybrid mampu meningkatkan akurasi dibandingkan metode tunggal [4]. Selanjutnya, seiring kemajuan machine learning, pendekatan model-based collaborative filtering menggunakan teknik faktorisasi matriks seperti SVD mulai banyak diterapkan. Penelitian Sitanggang dkk menunjukkan bahwa SVD unggul dalam menangani sparsity dengan membagi matriks user-item menjadi matriks laten berdimensi rendah, yang mampu menangkap pola tersembunyi (latent factors) dari preferensi pengguna yang tidak terdeteksi oleh metode memory-based tradisional [6].

Berdasarkan masalah tersebut, penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi sistem rekomendasi hybrid filtering yang mengkombinasikan item-based collaborative filtering dengan content-based filtering, serta model-based collaborative filtering menggunakan algoritma Singular Value Decomposition (SVD) pada dataset MovieLens. Sebelum proses pengembangan model, dilakukan analisis eksploratif data (EDA) untuk memahami karakteristik dataset, khususnya tingkat sparsity dan distribusi rating pengguna. Evaluasi kinerja sistem dilakukan menggunakan metrik Root Mean Square Error (RMSE) untuk mengukur akurasi prediksi serta Precision@10 untuk menilai relevansi daftar rekomendasi yang dihasilkan. Selain itu, penelitian ini juga membandingkan performa pendekatan memory-based dan model-based serta menganalisis efektivitas strategi hybrid dalam meningkatkan kualitas rekomendasi.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimen kuantitatif dengan memanfaatkan dataset publik MovieLens untuk mengevaluasi kinerja sistem rekomendasi film. Metode penelitian meliputi tahapan pengumpulan data, preprocessing, pemisahan data, analisis eksploratif data, serta penerapan beberapa pendekatan sistem rekomendasi, yaitu item-based collaborative filtering, content-based filtering, hybrid filtering, dan model-based collaborative filtering menggunakan algoritma Singular Value Decomposition (SVD). evaluasi kinerja dilakukan secara komparatif menggunakan metrik Root Mean Square Error (RMSE) dan Precision@10. Kontribusi ilmiah utama dari penelitian ini adalah analisis komparatif yang membuktikan bahwa pada dataset dengan sparsity tinggi, pendekatan model-based (SVD) jauh lebih efektif dibandingkan pendekatan hybrid berbasis bobot statis.

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian hybrid filtering ini berasal dari dataset MovieLens yang diperoleh melalui situs resmi GroupLens[12]. Kumpulan data mencakup 100.000 rating dengan 9.000 film yang diberikan rating oleh 600 user.

2.2. Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan kualitas data sebelum proses pemodelan. Matriks user-item dibentuk dari data rating dan dilakukan identifikasi terhadap nilai yang hilang (missing value). Penanganan missing value dilakukan melalui metode imputasi agar semua sel dalam matriks terisi dengan nilai yang tepat. Setelah tahap imputasi selesai, dilakukan pemeriksaan ulang untuk memastikan tidak ada missing value yang tersisa pada matriks hasil pemrosesan.

2.3. Pemisahan Data

Proses pemisahan data dilakukan dengan menggunakan pendekatan random holdout[9]. Seluruh data rating terlebih dahulu diacak guna memastikan reproducibilitas hasil. Sebanyak 80% dari data tersebut digunakan sebagai training set dalam proses pembelajaran model, sementara 20% sisanya digunakan sebagai test set untuk mengevaluasi kinerja model terhadap data yang tidak digunakan.

2.4. Exploratory Data Analysis (EDA)

Eksplorasi data (EDA) dilakukan untuk mengidentifikasi pola kecenderungan penilaian dari pengguna terhadap film. Analisis meliputi visualisasi distribusi rating pengguna terhadap film untuk mengidentifikasi kecenderungan pemberian rating, serta analisis tingkat sparsity untuk mengukur kepadatan interaksi antara pengguna dan film dalam bentuk rating. Hasil EDA digunakan sebagai dasar dalam pemilihan metode rekomendasi dan interpretasi hasil evaluasi.

2.5. Analisis Proses

2.5.1 Item-based Collaborative Filtering

Item-based Collaborative Filtering digunakan untuk mengidentifikasi item yang memiliki karakteristik serupa, sehingga sistem dapat merekomendasikan produk lain yang relevan. Selanjutnya, nilai kesamaan antar item dikombinasikan dengan preferensi pengguna untuk menghasilkan prediksi tingkat kegunaan atau relevansi suatu produk bagi pengguna. Produk yang memiliki nilai kegunaan tertinggi lah yang kemudian dijadikan rekomendasi [5].

2.5.1.1 Menghitung Similarity

Untuk menghitung nilai *similarity* antar item dan user digunakan persamaan (1).

$$S(i,j) = \frac{\sum_{v \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{v \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{v \in U} (R_{u,j} - \bar{R}_u)^2}} \quad (1)$$

Keterangan :

$S(i,j)$ = Nilai kemiripan antara item i dan item j
 $u \in U$ = Himpunan user yang memberi rating item i dan item j

R_u, i = Rating user u pada item i

R_u, j = Rating user u pada item j

\bar{R}_u = Nilai rating rata-rata user u

2.5.1.2 Menghitung Nilai Prediksi

Langkah selanjutnya adalah menghitung prediksi rating dari item-item tersebut. Cara menghitung nilai prediksi untuk item baru menggunakan persamaan Weighted Sum [1]. Persamaan Weighted Sum ditunjukkan pada persamaan (2).

$$P(a, j) = \frac{\sum_{i \in I} (R_a, i * R_S i, j)}{\sum_{i \in I} |S_i, j|}$$

(2)

Keterangan :

$P(a, j)$ = Prediksi rating item j oleh user a
 $i \in I$ = Himpunan item i yang mirip dengan item j

R_a, i = Rating user a pada item i

S_i, j = Nilai similarity antara item i dan j

2.5.2 Content-based Filtering

Pendekatan content-based filtering digunakan untuk merekomendasikan item yang memiliki kemiripan tinggi dengan item yang pernah disukai atau diberi penilaian oleh pengguna, sehingga bersifat personal[8]. Pendekatan ini membantu mengatasi keterbatasan analisis konten berbasis fitur eksplisit berupa genre [10].

2.5.2.1 Menghitung TF-IDF

Perhitungan TF-IDF untuk menghitung representasi genre film dalam bentuk vektor numerik. Persamaan TF-IDF ditunjukkan pada persamaan (3).

$$S(i, j) = TF(t, d) \times IDF(t)$$

$$IDF(t) = \log \left(\frac{N}{df_t} \right)$$

(3)

Keterangan :

$TF(t, d)$ = Frekuensi term t pada dokumen d (genre film)

$IDF(t)$ = Nilai Inverse Document Frequency dari term t

t = Term atau kata yang muncul pada genre film

N = Jumlah total dokumen (film) pada dataset

dt_t = Jumlah dokumen (film) yang mengandung term t

\log = Fungsi algoritma menstabilkan nilai pembobotan

2.5.2.2 Menghitung Cosine Similarity

Langkah selanjutnya, perhitungan cosine similarity untuk mengukur kemiripan genre antar film berdasarkan vektor TF-IDF. Perhitungan cosine similarity ditunjukkan pada persamaan (4).

$$CS(A, B) = \frac{A \cdot B}{|A| \times |B|}$$

(4)

Keterangan :

$CS(A, B)$ = Nilai kemiripan antara dua film

A = Vektor TF-IDF genre film pertama

B = Vektor TF-IDF genre film kedua

$A \cdot B$ = Hasil perkalian dot product vektor A dan B

$|A| \times |B|$ = Panjang hasil vektor A dan vektor B

2.5.2.3 Skor Rekomendasi

Kemudian, menghitung skor film yang dihitung dari rata-rata kemiripan film tersebut dengan film yang disukai user. Persamaan skor ditunjukkan pada persamaan (5).

$$Score(i) = \frac{1}{|L_u|} \sum_{j \in L_u} sim(i, j)$$

(5)

Keterangan :

$Score(i)$ = Nilai skor rekomendasi film i

L_u = Himpunan film yang disukai oleh pengguna u

$|L_u|$ = Jumlah film yang disukai oleh pengguna u

$sim(i, j)$ = Nilai cosine similarity antara film i dan film j

2.5.2.4 Prediksi Rating

Langkah terakhir menghitung pada prediksi rating film berdasarkan kemiripan yang pernah ditonton user dengan persamaan weighted average ditunjukkan pada persamaan (6).

$$\widehat{r}_{u,i} = \frac{\sum_{j \in L_u} sim(i, j) \times r_{u,j}}{\sum_{j \in L_u} |sim(i, j)|}$$

(6)

Keterangan :

$\widehat{r}_{u,i}$ = Prediksi rating pengguna u terhadap film i

L_u = Himpunan film yang pernah diberi rating oleh pengguna u

$r_{u,j}$ = Rating diberikan pengguna u terhadap film j

$sim(i, j)$ = Nilai cosine similarity antara film i dan film j

$|sim(i, j)|$ = Nilai absolut kemiripan antara film i dan film j

2.5.3 Hybrid Filtering

Hybrid filtering merupakan model yang menggabungkan antara 2 pendekatan atau lebih yang memanfaatkan peningkatan akurasi saat memberikan suatu rekomendasi item [6]. Pendekatan ini menggabungkan pendekatan item-based collaborative filtering dan content-based filtering dimana hasilnya akhir nilainya dihitung dengan persamaan (7).

$$\text{Hybrid} = (0.4 \times \text{content-based}) + (0.6 \times \text{collaborative filtering}) \quad (7)$$

Bobot lebih besar diberikan pada collaborative filtering karena pendekatan ini dinilai lebih efektif dalam menangkap pola preferensi pengguna berdasarkan interaksi rating, sementara content-based filtering berperan sebagai pendukung untuk menyesuaikan rekomendasi berdasarkan karakteristik konten.

2.5.4 Model-based SVD

Singular Value Decomposition (SVD) merupakan pendekatan dekomposisi matriks yang banyak diterapkan dalam bidang machine learning [7]. Teknik ini digunakan untuk mengekstraksi pola laten, melakukan reduksi dimensi, serta merepresentasikan kembali data ke dalam bentuk yang lebih ringkas. Melalui SVD, sebuah matriks dapat diuraikan menjadi tiga matriks terpisah yang lebih sederhana, sehingga memudahkan proses analisis dan pemodelan data.

Untuk memastikan kinerja model yang optimal dan reproducibility, parameter model ditentukan secara eksplisit. Learning rate ditetapkan sebesar 0.005 dan parameter regularization sebesar 0.02 untuk mencegah overfitting pada data latih. Eksperimen dilakukan dengan skema grid search untuk mencari kombinasi terbaik dari jumlah latent factors pada rentang [20, 30, 40, 50] dan jumlah iterasi (epochs) pada rentang [20, 30, 40]. Seluruh proses pelatihan model menggunakan random state bernilai 42 untuk menjaga konsistensi hasil eksperimen.

2.5.5 RMSE

Root Mean Square Error (RMSE) merupakan matrik evaluasi yang banyak digunakan dalam menilai performa algoritma machine learning, termasuk model yang lebih kompleks dibandingkan regresi linier. Nilai RMSE dimanfaatkan untuk mengukur perbedaan kinerja model pada tahap kalibrasi dan validasi, serta untuk membandingkan akurasi prediksi antar model yang berbeda [8]. Persamaan RMSE ditunjukkan pada persamaan (8).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (8)$$

Keterangan :

n = Jumlah data

E = Error

y' = Nilai output (prediksi)

Y = Nilai aktual

2.5.6 Precision@10

Precision merupakan pendekatan dalam mengukur persentase film yang direkomendasikan benar-benar relevan [10]. Pengukuran ini dikenal dengan istilah precision at k, dimana k adalah jumlah hasil rekomendasi teratas yang diberikan [11]. Oleh karena itu, digunakan precision@10 yang mengambil 10 jumlah hasil rekomendasi teratas. Rumus dari precision@10 ditunjukkan pada persamaan (9).

$$\text{Precision}@10 = \frac{\text{Jumlah item relevan 10 rekomendasi teratas}}{10} \quad (9)$$

Keterangan :

Item relevan = Daftar film yang disukai atau rating tinggi

10 rekomendasi teratas = Film dengan skor prediksi tertinggi hasil content-based filtering

Precision@10 = Nilai jumlah item rekomendasi teratas

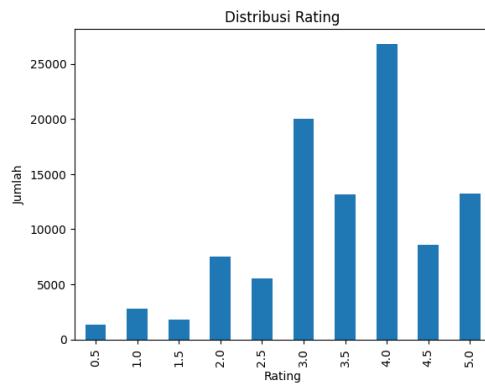
3. Hasil dan Pembahasan

Pada bab ini disajikan hasil dari proses implementasi dan evaluasi kinerja sistem rekomendasi hybrid filtering dan machine learning pada dataset MovieLens. Pembahasan dimulai dari tahap penerapan sistem. Setiap tahapan dijelaskan sesuai dengan hasil yang didapatkan selama penelitian.

Dataset yang digunakan dalam penelitian ini diperoleh dari MovieLens, yaitu dataset Movies yang memuat informasi movieId dan genre, serta dataset Ratings yang berisi data userId, movieId, dan rating.

Pada tahap preprocessing, data rating ditransformasikan ke dalam bentuk matriks user-item, di mana baris merepresentasikan pengguna dan kolom merepresentasikan film. Hasil pembuatan matriks menunjukkan adanya 5.365.412 nilai kosong (missing value) yang merepresentasikan film yang belum dirating oleh pengguna. Hal ini menunjukkan tingkat sparsity data yang tinggi. Untuk menangani missing value tersebut, nilai kosong diisi dengan nilai 0, yang merepresentasikan tidak adanya interaksi antara pengguna dan film. Proses ini bertujuan untuk membentuk matriks lengkap agar dapat diproses oleh algoritma Collaborative Filtering dan model berbasis matriks.

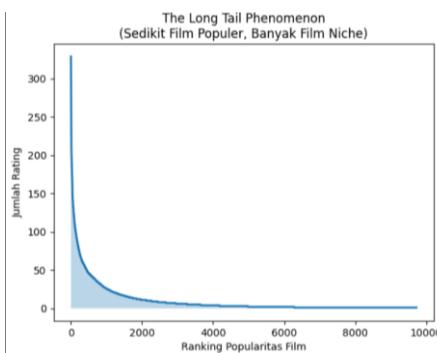
Distribusi rating dapat dilihat pada gambar 1 berikut.



Gambar 1. Distribusi Rating

Visualisasi tersebut menunjukkan bahwa sebagian besar pengguna memberikan rating tinggi pada film. Rating 4.0 menjadi nilai paling sering diberikan oleh pengguna, disusul oleh rating 3.0 dan 5.0. Hal ini menggambarkan adanya positive bias, yaitu kecenderungan pengguna memberi penilaian baik pada film yang mereka tonton. Sementara itu, rating rendah seperti 0.5, 1.0, dan 1.5 jarang diberikan, menunjukkan bahwa pengguna cenderung menghindari film yang diperkirakan buruk atau jarang memberikan penilaian negatif.

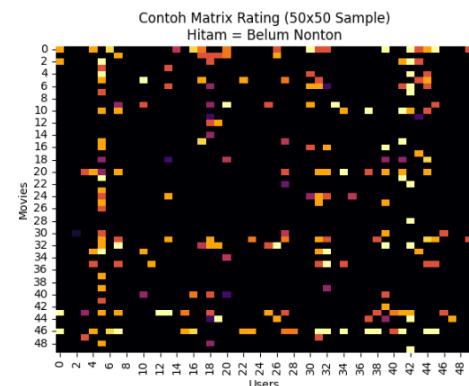
Fenomena Long Tail Pada Item-Based dapat dilihat pada gambar 2 berikut.



Gambar 2. Long Tail Item-Based Collaborative Filtering

Distribusi popularitas film menunjukkan adanya fenomena Long Tail, dimana hanya sedikit film yang memiliki jumlah rating yang sangat tinggi, sementara sebagian besar film lainnya memiliki jumlah rating yang jauh lebih rendah. Pola ini mengidentifikasi bahwa data bersifat tidak merata (sparse), sehingga sistem rekomendasi tidak hanya berfokus pada film populer, tetapi juga mampu mengeksplorasi film-film niche agar rekomendasi yang dihasilkan lebih beragam dan tetap relevan bagi pengguna.

Visualisasi matrix rating dapat dilihat pada gambar 3 berikut.



Gambar 3. Matrix Rating

Visualisasi matrix rating pengguna film (sample 50x50) menunjukkan tingkat sparsity data, di mana area berwarna hitam merepresentasikan film yang belum pernah diberi rating oleh pengguna, sedangkan area berwarna menunjukkan film yang pernah di rating. Visualisasi ini menggambarkan karakteristik data long tail dan ketidakmerataan distribusi rating pada dataset.

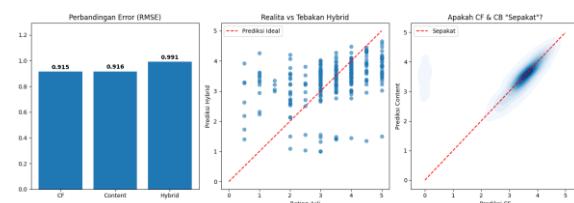
Visualisasi kemiripan film content-based dapat dilihat pada gambar 4 berikut.



Gambar 4. Kemiripan Film (CB - TF-IDF)

Heatmap kemiripan antar film yang dihasilkan oleh model Content-Based Filtering berbasis genre menggunakan pembobotan TF-IDF. Nilai kemiripan berada di rentang 0-1, di mana nilai 1 menunjukkan film dengan konten yang identik. Visualisasi ini memperlihatkan tingkat kemiripan antar film berdasarkan genre.

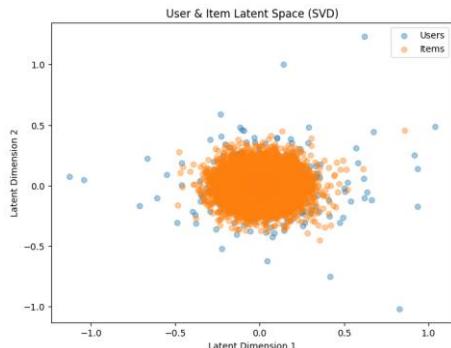
Perbandingan Error dan Performa Model dapat dilihat pada gambar 5 berikut.



Gambar 5. Perbandingan Error dan Performa Model

Visualisasi perbandingan RMSE menunjukkan bahwa model Collaborative dan Content-Based memiliki tingkat error yang hampir sama, sedangkan model Hybrid menghasilkan nilai RMSE yang sedikit lebih tinggi. Artinya penggabungan kedua pendekatan belum mampu meningkatkan akurasi dengan signifikan. Scatter plot yang menggambarkan hubungan antar rating asli dan prediksi hybrid menunjukkan bahwa model cenderung memberikan nilai yang lebih rendah dibandingkan dengan rating asli, terutama untuk film dengan penilaian tinggi. Di sisi lain, analisis kesesuaian antara prediksi Collaborative Filtering dan Content-Based menunjukkan bahwa distribusi titik mengikuti garis diagonal, yang menunjukkan adanya konsistensi dalam prediksi di banyak kasus. Namun, variasi kepadatan di sekitar garis tersebut menunjukkan bahwa masih terdapat perbedaan estimasi pada beberapa item.

Visualisasi ruang laten user dan item dapat dilihat pada gambar 6 berikut.



Gambar 6. Ruang Laten User dan Item

Titik-titik user dan item tersebar dalam dua dimensi laten, memperlihatkan struktur interaksi yang dipelajari model. Walaupun sebagian besar titik terlihat berkelompok rapat di sekitar pusat, terdapat sejumlah pengguna dan item yang terletak jauh dari pusat, menunjukkan perilaku atau karakteristik yang unik. Visualisasi ini memperlihatkan bahwa SVD mampu menangkap variasi preferensi pengguna dan hubungan antar item dengan lebih baik dibandingkan pendekatan berbasis kesamaan sederhana.

3.1 Cara Kerja Model dan Algoritma

Pada pendekatan Item-Based Collaborative Filtering, sistem mengevaluasi seberapa mirip film satu dengan yang lain berdasarkan pola rating yang diberikan oleh pengguna lainnya. Film-film yang memiliki kemiripan yang tinggi digunakan sebagai dasar untuk memprediksi rating pengguna terhadap film yang belum mereka tonton. Prediksi nilai rating ini didapatkan melalui penggabungan nilai rating dari film-film serupa dengan memperhatikan bobot kemiripan. Model ini dapat menghasilkan rekomendasi yang sesuai, tetapi kinerjanya dipengaruhi oleh tingginya tingkat sparsity pada matriks rating.

Pendekatan Content-Based Filtering memanfaatkan data genre film. Genre film direpresentasikan dalam bentuk vektor menggunakan metode TF-IDF, yang kemudian digunakan untuk menghitung tingkat kesamaan antara film. Profil preferensi pengguna dibentuk berdasarkan film-film yang sebelumnya mereka suka, sehingga rekomendasi yang diberikan lebih fokus pada film dengan ciri konten yang mirip. Keterbatasan variasi dan informasi mendalam dari fitur konten dalam dataset MovieLens mengakibatkan pendekatan ini memiliki kinerja yang cukup rendah.

Pendekatan Hybrid Filtering dalam penelitian ini mengintegrasikan hasil prediksi dari Item-Based Collaborative Filtering dan Content-Based Filtering. Penggabungan ini dilakukan melalui skema jumlah berbobot dengan bobot yang ditentukan secara manual dan bersifat statis. Tujuan dari model ini adalah untuk mengkombinasikan keunggulan dari kedua pendekatan, tetapi karena pembobotan belum disesuaikan secara adaptif, peningkatan kinerja yang didapat masih belum signifikan jika dibandingkan dengan pendekatan tunggal.

Pendekatan Model-Based menggunakan algoritma Singular Value Decomposition (SVD) bekerja dengan memfaktorkan matriks user-item menjadi representasi laten pengguna dan item. Model ini mampu menangkap pola preferensi tersembunyi serta hubungan kompleks antara pengguna dan film, sehingga lebih efektif dalam menangani data dengan tingkat sparsity yang tinggi. Hasil prediksi rating yang dihasilkan oleh model SVD kemudian digunakan untuk menyusun daftar rekomendasi film bagi pengguna.

3.2 Hasil Analisis Rekomendasi

Hasil analisis rekomendasi menunjukkan bahwa setiap model menghasilkan karakteristik rekomendasi yang berbeda. Item-Based Collaborative Filtering cenderung merekomendasikan film-film populer dengan genre yang sesuai dengan riwayat pengguna, namun memiliki keterbatasan dalam memberikan variasi rekomendasi akibat sparsity data. Content-Based Filtering menghasilkan rekomendasi dengan tingkat kemiripan konten yang tinggi, terutama berdasarkan genre, namun kurang mampu menangkap preferensi laten pengguna. Model Hybrid Filtering menggabungkan kedua pendekatan tersebut, tetapi efektivitasnya sangat bergantung pada skema pembobotan yang digunakan. Sementara itu, model Model-Based SVD mampu menghasilkan rekomendasi yang lebih relevan dan beragam karena mempelajari pola laten dari interaksi pengguna dan item, sehingga lebih fleksibel terhadap karakteristik dataset MovieLens.

3.3 Hasil Rekomendasi

Hasil rekomendasi dan prediksi rating pendekatan Item-Based Collaborative Filtering dapat dilihat pada tabel 1 berikut.

Tabel 1. Hasil rekomendasi dan prediksi rating Item-Based

	Judul	Genre	Prediksi Rating
1.	Groundhog Day (1993)	Comedy, Fantasy, Romance	4.09
2.	Indiana Jones and the Temple of Doom (1984)	Action, Adventure, Fantasy	4.09
3.	The Sixth Sense (1999)	Drama, Horror, Mystery	3.98
4.	Ferris Bueller's Day Off (1986)	Comedy	3.97
5.	Who Framed Roger Rabbit? (1988)	Adventure, Animation, Fantasy	3.95

Tabel menunjukkan hasil rekomendasi film untuk User 140 berdasarkan pendekatan Item-Based Collaborative Filtering. Rekomendasi diurutkan berdasarkan tingkat kemiripan antar item yang pernah diberi rating oleh pengguna, sehingga film yang direkomendasikan memiliki pola preferensi yang serupa dengan film yang telah ditonton sebelumnya. Nilai prediksi rating yang dihasilkan berada pada rentang 3.95 hingga 4.09 yang menunjukkan bahwa film-film tersebut diperkirakan disukai oleh pengguna. Hasil mengindikasikan bahwa pendekatan Item-Based Collaborative Filtering mampu memberikan rekomendasi yang relevan berdasarkan hubungan antar item dalam data.

3.3.1 Hasil rekomendasi dan score rekomendasi pendekatan Content-Based Filtering dapat dilihat pada tabel 2 berikut.

Tabel 2. Hasil rekomendasi CBF dan score rekomendasi

	Judul	Genre	Score Rekomendasi
1.	Little Big Soldier (Da bing xiao jiang) (2010)	Action, Adventure, Comedy, Drama, War	0.207
2.	Three Kings (1999)	Action, Adventure, Comedy, Drama, War	0.207
3.	Forbidden Kingdom, The (2008)	Action, Adventure, Comedy, Fantasy	0.204
4.	Dungeons & Dragons (2000)	Action, Adventure, Comedy, Fantasy	0.204
5.	Gamers, Dorkness The: Rising (2008)	Action, Adventure, Comedy, Fantasy	0.204

Tabel tersebut menunjukkan hasil rekomendasi untuk user 1 berdasarkan pendekatan Content-Based Filtering. Rekomendasi diurutkan berdasarkan tingkat kemiripan konten tertinggi yang diperoleh dari perhitungan Cosine Similarity antar genre film. Sistem merekomendasikan film yang memiliki kesamaan genre dengan film yang disukai oleh pengguna, seperti Action, Adventure Comedy, Drama, dan Fantasy. Nilai skor rekomendasi yang dihasilkan berada pada rentang 0.204 hingga 0.207, yang menunjukkan tingkat kemiripan konten yang relatif tinggi antar film rekomendasi. Hasil ini mengindikasikan bahwa pendekatan Content-Based Filtering mampu memberikan rekomendasi yang relevan dengan memanfaatkan kesesuaian konten item terhadap preferensi pengguna.

Hasil rekomendasi dan prediksi rating pendekatan Hybrid Filtering dapat dilihat pada tabel 3 berikut.

Tabel 3. Hasil rekomendasi Hybrid dan prediksi rating

	Judul	Prediksi Rating
1.	Room at the Top (1959)	3.90
2.	Long, Hot Summer, The (1958)	3.90
3.	Hamlet (1964)	3.84
4.	My Sister's Keeper (2009)	3.81
5.	Iron Man (1931)	3.81

Tabel tersebut menunjukkan hasil rekomendasi film untuk User 140 berdasarkan pendekatan Hybrid Filtering, yaitu penggabungan Collaborative Filtering dan Content-Based Filtering. Rekomendasi disusun berdasarkan nilai prediksi rating yang dihasilkan dari kombinasi kedua pendekatan, sehingga film yang muncul memiliki tingkat relevansi yang konsisten baik dari sisi kesamaan perilaku pengguna maupun kemiripan karakteristik konten. Nilai prediksi berada pada rentang 3.81 hingga 3.90, yang menunjukkan bahwa film-film tersebut diperkirakan sesuai dengan preferensi pengguna. Hasil dari mengindikasikan bahwa pendekatan hybrid mampu memberikan rekomendasi yang stabil dan relevan dengan memanfaatkan keunggulan dari kedua pendekatan, meskipun data rating pengguna relatif terbatas.

Hasil rekomendasi dan prediksi rating pendekatan Model-Based SVD dapat dilihat pada tabel 4 berikut.

Tabel 4. Hasil rekomendasi dan prediksi rating *Model-Based SVD*

	Judul	Prediksi Rating
1.	Amelie (Fabuleux destin d'Amélie Poulain, Le) (2001)	4.24
2.	Streetcar Named Desire, A (1951)	4.23
3.	North by Northwest (1959)	4.20
4.	Dark Knight, The (2008)	4.18
5.	Cinema Paradiso (Nuovo cinema Paradiso) (1989)	4.17

Tabel tersebut menunjukkan hasil rekomendasi film untuk User 140 yang dihasilkan menggunakan pendekatan Model-Based SVD. rekomendasi diurutkan berdasarkan hasil dekomposisi matriks rating user-item sehingga mampu memprediksi preferensi pengguna terhadap film yang belum pernah ditonton. Nilai prediksi rating yang dihasilkan berada pada rentang 4.17 - 4.24 yang menunjukkan tingkat ketertarikan pengguna yang tinggi terhadap film-film yang direkomendasikan.

3.3.2 Perbandingan Kinerja Antar Model

Perbandingan evaluasi RMSE antar model dapat dilihat pada tabel 5 berikut.

Table 5. Perbandingan RMSE antar model

Model	RMSE
Item-Based	0.915
Content-Based	0.916
Hybrid Filtering	0.990
Model-Based SVD	0.877

Tabel tersebut menunjukkan hasil evaluasi kinerja sistem rekomendasi menggunakan metrik Root Mean Square Error (RMSE). Berdasarkan hasil tersebut, pendekatan Model-Based SVD menghasilkan nilai RMSE yang paling rendah yaitu sebesar 0.877, yang menunjukkan tingkat kesalahan prediksi paling kecil dibandingkan model lainnya. Pendekatan Item-based Collaborative Filtering dan Content-Based Filtering memiliki nilai RMSE yang relatif mirip, masing-masing sebesar 0.915 dan 0.916. Sementara itu, model Hybrid Filtering menunjukkan nilai RMSE paling tinggi yaitu sebesar 0.990, yang disebabkan oleh proses penggabungan prediksi yang masih menggunakan manual weighted sum dengan bobot tetap, sehingga kontribusi masing-masing model belum optimal terhadap karakteristik data.

Perbandingan evaluasi precision@10 antar model dapat dilihat pada tabel 6 berikut.

Tabel 6. Perbandingan Precision@10 antar model

Model	Precision
Item-Based	16.88%
Content-Based	0.69%
Hybrid Filtering	0.40%
Model-Based SVD	67.45%

Tabel tersebut menunjukkan hasil evaluasi kinerja sistem menggunakan metrik Precision@10. Berdasarkan hasil tersebut, model SVD menghasilkan performa terbaik dengan nilai precision mencapai 67 persen, jauh lebih tinggi dibandingkan model lainnya. Model Item-Based Collaborative Filtering berada pada posisi kedua dengan precision 17 persen, sementara Content-Based Filtering dan Hybrid Filtering berada pada tingkat yang sangat rendah, masing-masing mendekati nol. Perbedaan yang mencolok ini mengindikasikan bahwa SVD lebih efektif dalam mengidentifikasi item yang relevan bagi pengguna pada skenario top-N recommendation, sedangkan model Content-Based Filtering dan Hybrid Filtering kurang mampu menangkap pola yang diperlukan untuk menghasilkan rekomendasi yang tepat.

4. Kesimpulan

Berdasarkan hasil implementasi dan evaluasi kinerja sistem rekomendasi pada dataset MovieLens, dapat disimpulkan bahwa pendekatan Model-Based Collaborative Filtering menggunakan algoritma Singular Value Decomposition (SVD) menunjukkan performa terbaik dibandingkan pendekatan lainnya. Hal ini ditunjukkan oleh nilai RMSE terendah sebesar 0.877 serta nilai Precision@10 tertinggi sebesar 67.45%, yang menunjukkan akurasi prediksi dan relevansi rekomendasi yang lebih baik.

Model Item-Based Collaborative Filtering dan Content-Based Filtering menghasilkan performa yang relatif lebih rendah, dengan nilai RMSE masing-masing sebesar 0.915 dan 0.916 serta Precision@10 sebesar 16.88% dan 0.69%. Kinerja kedua model tersebut dipengaruhi oleh karakteristik data MovieLens yang memiliki tingkat sparsity tinggi serta keterbatasan fitur konten yang digunakan dalam pendekatan Content-Based Filtering sehingga kurang mampu menangkap preferensi laten pengguna secara menyeluruhan.

Sementara itu, pendekatan Hybrid Filtering yang menggabungkan Item-Based Collaborative Filtering dan Content-Based Filtering dalam penelitian ini belum menunjukkan peningkatan kinerja yang signifikan, dengan nilai RMSE sebesar 0.990 dan Precision@10 sebesar 0.40%. Hal ini disebabkan oleh proses penggabungan prediksi yang masih menggunakan skema manual weighted sum dengan bobot statis, sehingga belum mampu

mengoptimalkan kontribusi masing-masing model.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pendekatan berbasis machine learning dengan model laten lebih efektif dalam menghasilkan rekomendasi yang akurat dan relevan pada dataset dengan tingkat sparsity tinggi dibandingkan pendekatan berbasis konten maupun hybrid dengan bobot manual.

Daftar Rujukan

- [1] Salsabilla, C. dan Utomo, D.W. (2025) “Sistemasi: Jurnal Sistem Informasi Sistem Rekomendasi Wisata Kabupaten Pekalongan dengan Content Based Filtering Pekalongan Regency Tourism Recommendation System with Content based Filtering,” SISTEMASI (Jurnal Sistem Informasi), 14(1), hlm. 256–270. Tersedia pada: <https://doi.org/10.32520/stmsi.v14i1.4839>.
- [2] Khusnah, M., Gernowo, R. dan Surarso, B. (2025) “Implementasi E-Commerce dengan Sistem Informasi Rekomendasi menggunakan Metode Collaborative Filtering untuk Pengembangan Penjualan pada UMKM,” Jurnal Sistem Informasi Bisnis, 15(1), hlm. 135–142. Tersedia pada: <https://doi.org/10.14710/vol15iss1pp135-142>.
- [3] Maulana, F. dan Setiawan, E.B. (2024) “Performance of Deep Feed-Forward Neural Network Algorithm Based on Content-Based Filtering Approach,” INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi, 8(2), hlm. 278–294. Tersedia pada: <https://doi.org/10.29407/intensif.v8i2.22904>.
- [4] Yudha, A.A.P., Munir dan Anisyah, A. (2025) “Perancangan Sistem Rekomendasi Akomodasi pada Event Konser dengan Metode Hybrid Filtering,” JUKTISI (Jurnal Komputer Teknologi Informasi Sistem Informasi), 4(2), hlm. 631–641. Tersedia pada: <https://doi.org/10.62712/juktisi.v4i2.493>.
- [5] Putri, R.P., Maulindar, J. dan Pradana, A.I. (2025) “Sistem Rekomendasi Kuliner Karanganyar Menggunakan Metode Hybrid Recommendation,” Progresif: Jurnal Ilmiah Komputer, 21(2), hlm. 910–920. Tersedia pada: <https://doi.org/10.35889/progresif.v21i2.3105>.
- [6] Sitanggang, A. dkk. (2023) “Sistem Rekomendasi Anime Menggunakan Metode Singular Value Decomposition (SVD) dan Cosine Similarity,” JTI (Jurnal Teknologi Informasi), 2(2), hlm. 90–94. Tersedia pada: <https://doi.org/10.35308/jti.v2i2.7787>
- [7] Kurniawan, A.A. dan Aribowo, W. (2024) “Perancangan dan Implementasi Sistem Pembangkit Listrik Tenaga Hybrid,” JURNAL TEKNIK ELEKTRO, 14(2), hlm. 102–107. Tersedia pada: <https://doi.org/10.26740/jte.v14n2.p102-107>.
- [8] Yuricha dan Phan, I.K. (2025) “REKOMENDASI DRAMA KOREA MENGGUNAKAN CONTENT-BASED FILTERING DENGAN WEIGHTED TREE SIMILARITY,” JITET (Jurnal Informatika dan Teknik Elektro Terapan), 13(1), hlm. 1716–1725. Tersedia pada: <https://doi.org/10.23960/jitet.v13i1.6050>.
- [9] Hendrawan, S. dan Yakub (2025) “OPTIMASI SISTEM REKOMENDASI FILM MENGGUNAKAN METODE HYBRID FILTERING,” POTERS (Proceedings of Technology, Engineering and Computers), 1(2), hlm. 316–323. Tersedia pada: <https://jurnal.ubd.ac.id/index.php/poters/index>.
- [10] Velamentosa, D. dan Zuliarsa, E. (2025) “SISTEM REKOMENDASI FILM MENGGUNAKAN METODE CONTENT-BASED FILTERING,” JATI (Jurnal Mahasiswa Teknik Informatika), 9(2), hlm. 2918–2922. Tersedia pada: <https://doi.org/https://doi.org/10.36040/jati.v9i2.13251>.
- [11] Roziqiin, N.M. (2024) SISTEM REKOMENDASI PEMILIHAN FILM ANIMASI MENGGUNAKAN METODE CONTENT BASED FILTERING. Tersedia pada: <http://etheses.uin-malang.ac.id/id/eprint/65972> (Diakses: 10 Februari 2026).
- [12] Pratama, R.A. dkk. (2025) “Perbandingan User-Based dan Item-Based pada Sistem Rekomendasi Film Kombinasi Teknik Reduksi Dimensi dan Clustering,” Jurnal Tekno Insentif, 19(1), hlm. 1–14. Tersedia pada: <https://doi.org/10.36787/jti.v19i1.1662>.