

Optimasi Kinerja Metode Hybrid Content-Based Filtering dan Matrix Factorization pada Sistem Rekomendasi Film

Andre Januarta¹, Bagus Darmajati Perdana Putra², Deka Dwi Nugraha³, Sofyan Hidayat⁴, Mahendra Akmal Abiyasa⁵, Ade Saputra⁶, Zaidan Umar⁷, Faradi Deo Herlambang⁸

^{1,2,3,4,5,6,7} Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Purwokerto, Indonesia

⁸ Program Studi Teknik Informatika, Fakultas Teknik dan Sains, Universitas Muhammadiyah Purwokerto, Purwokerto, Indonesia
surel: ¹januarta.andre@gmail.com, ²bagusfernandez64@gmail.com, ³dekahryn03@gmail.com, ⁴sofyanking291201@gmail.com, ⁵mahendraabiyasa@gmail.com, ⁶deesapp256@gmail.com, ⁷zaedanomar2308@gmail.com, ⁸faradideoherlambang4@gmail.com

Info Artikel

Sejarah artikel:

Diterima 20-01-2026

Revisi 19-02-2026

Diterima 05-03-2026

Kata kunci:

Sistem Rekomendasi
Content-Based Filtering
Matrix Factorization
TF-IDF
Cosine Similarity

ABSTRAK

Perkembangan platform digital dan layanan streaming film menyebabkan peningkatan jumlah konten yang tersedia bagi pengguna. Kondisi ini menimbulkan permasalahan information overload, di mana pengguna mengalami kesulitan dalam menentukan film yang sesuai dengan preferensi mereka. Sistem rekomendasi hadir sebagai solusi untuk membantu pengguna menemukan film yang relevan secara cepat dan personal. Penelitian ini bertujuan untuk menganalisis kinerja sistem rekomendasi film menggunakan metode hybrid yang menggabungkan Content-Based Filtering (CBF) dan Matrix Factorization (MF). Pendekatan CBF memanfaatkan representasi genre film menggunakan Term Frequency–Inverse Document Frequency (TF-IDF) dan Cosine Similarity untuk mengukur kemiripan antar film, sedangkan MF diterapkan menggunakan teknik Singular Value Decomposition (SVD) untuk mempelajari pola laten interaksi pengguna dan film berdasarkan data rating. Dataset yang digunakan berasal dari The Movie Database (TMDB) dan MovieLens dengan jumlah data yang besar dan beragam. Hasil penelitian menunjukkan bahwa metode hybrid mampu meningkatkan kualitas rekomendasi dengan menggabungkan keunggulan kedua pendekatan, di mana CBF berperan sebagai penyaring awal berdasarkan kemiripan konten dan MF berfungsi untuk memprediksi serta mengurutkan film berdasarkan preferensi pengguna. Evaluasi kinerja sistem menggunakan confusion matrix dan metrik precision, recall, serta F1-score menunjukkan tingkat akurasi sebesar 73%, dengan performa yang baik pada kelas like. Hasil ini membuktikan bahwa pendekatan hybrid efektif dalam menghasilkan rekomendasi film yang lebih relevan dan personal.

Penulis yang sesuai:

Andre Januarta

Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Purwokerto

Email: januarta.andre@gmail.com

1. PENDAHULUAN

Pertumbuhan industri perfilman dan perkembangan platform streaming digital menyebabkan meningkatnya jumlah konten film yang dapat diakses oleh pengguna, kondisi ini memberikan kemudahan dalam memperoleh



hiburan. Berdasarkan laporan WIPO dan Omdia yang dikutip dalam [1], industri perfilman global pada tahun 2023 mencatat produksi sebanyak 9.511 judul film. Peningkatan tersebut sejalan dengan pesatnya perkembangan platform layanan streaming. Lebih lanjut, masih merujuk pada [1], laporan FlixPatrol menunjukkan bahwa jumlah pengguna salah satu platform streaming terbesar, yaitu Netflix, mengalami pertumbuhan yang sangat signifikan dalam kurun waktu enam tahun terakhir, dari 139 juta pengguna pada tahun 2018 menjadi 301 juta pengguna pada tahun 2024.

Pertumbuhan jumlah pengguna yang diiringi dengan meningkatnya produksi dan ketersediaan konten film dari waktu ke waktu menimbulkan permasalahan tersendiri, yaitu kondisi information overload, di mana pengguna harus dihadapkan pada ribuan judul film yang terus bertambah. Situasi ini berpotensi menimbulkan kesulitan pada pengguna dalam menentukan film yang sesuai dengan preferensi mereka. Oleh karena itu, sistem rekomendasi banyak dikembangkan dan dimanfaatkan sebagai solusi untuk membantu pengguna menemukan konten yang relevan berdasarkan minat dan pola perilaku yang terekam sebelumnya.

Sejumlah penelitian telah dilakukan untuk mengatasi permasalahan tersebut. Salah satunya adalah penelitian pada [2] yang mengusulkan penerapan metode collaborative filtering. Namun, pendekatan ini memiliki keterbatasan karena sangat bergantung pada pola penilaian (rating) historis dari pengguna, sehingga kurang mampu melakukan prediksi terhadap film yang baru dirilis akibat ketiadaan data rating sebelumnya. Selain itu, penelitian tersebut tidak menyajikan hasil evaluasi kinerja sistem menggunakan metrik kuantitatif seperti Root Mean Square Error (RMSE), Mean Absolute Error (MAE), maupun confusion matrix, sehingga tingkat akurasi dan keandalan sistem rekomendasi yang dibangun belum dapat diukur secara objektif.

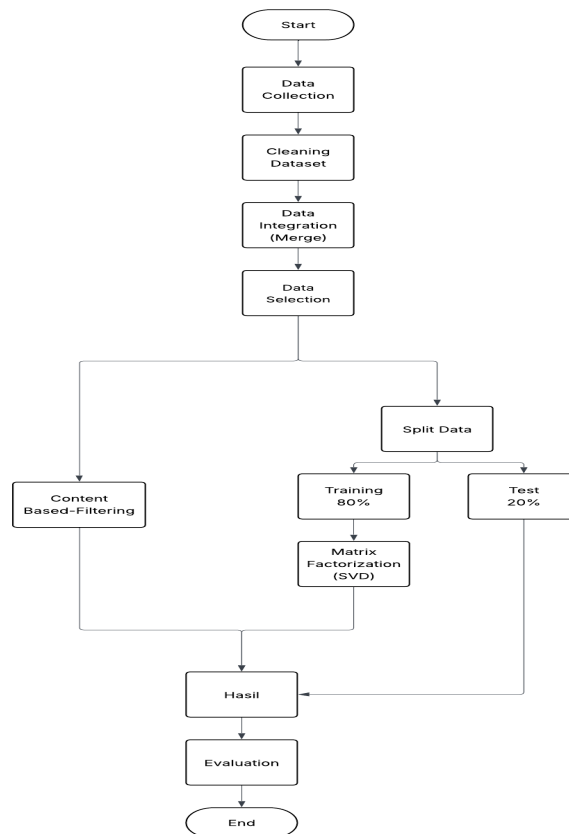
Selanjutnya, penelitian pada [3] juga menggunakan pendekatan yang sama, yaitu collaborative filtering. Keterbatasan utama dari penelitian ini terletak pada ketergantungan model terhadap data pengguna lain untuk mempelajari pola preferensi, sehingga sistem tidak mampu melakukan prediksi secara langsung berdasarkan karakteristik film semata (item-based) tanpa adanya riwayat rating sebelumnya. Penelitian sebelumnya telah mengembangkan sistem rekomendasi film dengan mengombinasikan pendekatan Content-Based Filtering dan Collaborative Filtering. Namun, pendekatan hibrida tersebut masih memiliki sejumlah keterbatasan, di antaranya permasalahan cold-start dan sparsity, yang berpotensi memengaruhi kualitas serta tingkat akurasi rekomendasi yang dihasilkan [4].

Untuk mengatasi berbagai keterbatasan tersebut, pendekatan hybrid dapat dijadikan sebagai alternatif dengan mengombinasikan beberapa metode rekomendasi. Pendekatan hybrid dipilih karena berdasarkan Penelitian [1] menunjukkan bahwa sistem rekomendasi dengan pendekatan hybrid yang mengintegrasikan Content-Based Filtering mampu meningkatkan tingkat akurasi serta keberagaman (diversity) rekomendasi dibandingkan dengan penggunaan satu metode saja dan mampu meningkatkan akurasi serta diversifikasi rekomendasi dibandingkan metode tunggal. Penelitian ini mengusulkan pengembangan sistem rekomendasi film menggunakan pendekatan hybrid yang mengombinasikan teknik Content-Based Filtering dan Matrix Factorization. Tujuan utama penelitian ini adalah menganalisis kinerja pendekatan hybrid dalam meningkatkan kualitas rekomendasi film serta mengevaluasi performa sistem menggunakan metrik evaluasi yang relevan.

Pendekatan Matrix Factorization dengan teknik Singular Value Decomposition (SVD) dipilih karena terbukti efektif dalam menangkap pola laten pada data penilaian berskala besar, sehingga berpotensi meningkatkan ketepatan dalam memodelkan dan memprediksi preferensi pengguna.

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental. Dataset yang digunakan diperoleh melalui proses scraping dari situs The Movie Database (TMDB) serta memanfaatkan dataset yang disediakan secara resmi oleh platform MovieLens. Sistem rekomendasi kemudian dikembangkan dan diuji menggunakan data film dan penilaian pengguna untuk mengevaluasi kinerja metode hibrida yang mengombinasikan Content-Based Filtering dan Matrix Factorization.



Gambar 1. Flowchart metode

2.1. Data Collection

Penelitian ini menggunakan dataset yang berasal dari dua sumber, yaitu The Movie Database (TMDB) dan MovieLens, yang terdiri atas empat file utama, yaitu ratings.csv, movies.csv, links.csv, dan dataset_tmdb_movies.csv. file ratings.csv berisi data penilaian (rating) yang diberikan oleh pengguna, movies.csv memuat daftar film pada platform MovieLens, sedangkan links.csv berfungsi sebagai penghubung antara identitas film pada MovieLens dan TMDB. Ketiga berkas tersebut diperoleh dari situs resmi MovieLens. Adapun file dataset_tmdb_movies.csv yang berisi informasi detail film yang bersumber dari situs TMDB, yang diperoleh melalui proses scraping. Jumlah data film yang dikumpulkan dalam dataset ini sebanyak 6.000 judul.

Tabel 1. Jumlah Dataset yang diperoleh

No.	Dataset	Jumlah
1	ratings.csv	32000204
2	movies.csv	87585
3	links.csv	87461
4	dataset_tmdb_movies.csv	6000

2.2. Cleaning Data

Cleaning Data merupakan tahapan yang bertujuan untuk mengidentifikasi serta memperbaiki kesalahan, inkonsistensi, dan ketidakakuratan dalam dataset guna meningkatkan kualitas data. Proses ini mencakup berbagai aktivitas, seperti penghapusan data duplikat, penanganan nilai hilang, penyeragaman format, perbaikan kesalahan penulisan, serta penyelarasan data yang berasal dari berbagai sumber [5]. Pada penelitian ini, proses cleaning data dilakukan untuk menyesuaikan tipe data pada setiap atribut yang digunakan dalam pembangunan sistem, sehingga memastikan kesesuaian format data dengan kebutuhan model dan kelancaran proses komputasi. Selain itu, data cleaning juga bertujuan untuk menangani keberadaan data duplikat pada dataset, sehingga informasi yang digunakan

bersifat unik, konsisten, dan tidak menimbulkan bias. Dengan demikian, kualitas data yang lebih baik diharapkan dapat mendukung kinerja sistem rekomendasi yang dirancang secara lebih optimal.

2.3. Integration Data

Data integration atau penggabungan dataset merupakan tahapan penyatuan data yang berasal dari berbagai sumber ke dalam satu struktur yang terorganisasi, sehingga informasi yang semula terpisah dapat saling terhubung dan digunakan secara terpadu. Proses ini bertujuan untuk menyelaraskan format, atribut, serta relasi antar data agar membentuk basis data yang konsisten, lengkap, dan siap digunakan dalam tahap analisis maupun pemodelan sistem, khususnya dalam membangun hubungan antara objek, pengguna, dan karakteristik yang relevan di dalam sistem rekomendasi [6].

2.4. Content Based-Filtering

Content-Based Filtering merupakan metode rekomendasi yang didasarkan pada karakteristik atau deskripsi dari suatu item. Metode ini bekerja dengan membandingkan profil preferensi pengguna terhadap representasi fitur item, kemudian merekomendasikan item-item yang memiliki tingkat kemiripan tertinggi dengan item yang sebelumnya diminati atau digunakan oleh pengguna [7].

2.5. TF-IDF

TF-IDF (Term Frequency–Inverse Document Frequency) merupakan metode pembobotan kata yang digunakan untuk merepresentasikan tingkat kepentingan suatu term dalam sebuah dokumen terhadap keseluruhan kumpulan dokumen. Komponen Term Frequency (TF) menyatakan jumlah kemunculan suatu kata dalam sebuah dokumen, di mana semakin sering kata tersebut muncul, semakin besar pula nilai frekuensinya [8]. Nilai TF ini kemudian digunakan sebagai dasar dalam perhitungan bobot kata, yang secara matematis dirumuskan melalui persamaan term frequency pada dokumen [9].

$$TF(d, t) = f(d, t) \quad (1)$$

Dimana :

f = Frequency Kemunculan Kata

t = Term (Kata/Frasa)

d = Dokumen

Semakin jarang suatu term muncul, maka nilai *Inverse Document Frequency* (IDF) akan semakin besar. Nilai tersebut dihitung menggunakan rumus sebagai berikut:

$$IDF = \text{LOG}\left(\frac{d}{df(t)}\right) \quad (2)$$

Dimana :

D = jumlah dokumen yang memuat kata kunci

df(t) = jumlah dokumen yang memuat term (t)

Selanjutnya, bobot suatu term dalam dokumen diperoleh dari hasil perkalian nilai Term Frequency (TF) dan Inverse Document Frequency (IDF), sebagaimana ditunjukkan pada rumus berikut:

$$W(d, t) = TF \times IDF \quad (3)$$

Dimana :

W = bobot setiap dokumen yang memuat kata kunci

TF = Term Frequency

IDF = Inverse Dokumen Frequency

Hasil dari proses perhitungan tersebut menghasilkan representasi kata kunci (keyword extraction) yang merefleksikan tingkat kepentingan setiap term dalam dokumen [10].

2.6. Cosine Similarity

Metode Cosine Similarity digunakan untuk menghitung tingkat kemiripan antara dua objek yang direpresentasikan dalam bentuk vektor, dengan memanfaatkan kata kunci yang diekstraksi dari dokumen sebagai dasar perbandingan. Nilai *Cosine Similarity* berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan tingkat kemiripan yang sangat tinggi atau identik antara dua objek, sedangkan nilai yang mendekati 0 menunjukkan kemiripan yang rendah. Perhitungan tingkat kesamaan tersebut dirumuskan dengan persamaan sebagai berikut:

$$x, y = x \text{ dan } y \text{ dihitung dengan } \sum_k^n = 1x_k y_k \quad (4)$$

Dimana :

$\|x\|$ = Panjang vektor x, dihitung dengan $\sum_k^n = 1x_k y_k$

$\|y\|$ = Panjang vektor y, dihitung dengan $\sum_k^n = 1x_k y_k$

2.7. Matrix Factorisasi

Matrix Factorization merupakan algoritma yang melakukan dekomposisi suatu matriks ke dalam beberapa matriks berukuran lebih kecil untuk memperoleh representasi fitur laten yang merefleksikan pola preferensi pengguna serta memungkinkan estimasi terhadap nilai peringkat yang belum teramati, sehingga dapat meningkatkan kualitas rekomendasi [11]. Proses faktorisasi tersebut kemudian dioptimalkan menggunakan teknik Singular Value Decomposition (SVD) guna merepresentasikan preferensi pengguna secara lebih akurat [12]. Hasil dari proses optimasi tersebut memungkinkan sistem untuk menghasilkan prediksi penilaian film, yaitu estimasi nilai rating yang berpotensi diberikan oleh pengguna terhadap film yang belum pernah mereka tonton [12].

2.8. Metrik Evaluasi

Penelitian ini menggunakan dua metrik evaluasi berbasis confusion matrix, yaitu Mean Absolute Error (MAE) dan Root Mean Square Error (RMSE), untuk menilai kinerja sistem rekomendasi. MAE digunakan untuk mengukur tingkat keakuratan prediksi dengan menghitung rata-rata selisih absolut antara nilai prediksi dan nilai aktual. Sementara itu, RMSE digunakan untuk mengukur besarnya kesalahan prediksi dengan menghitung akar dari rata-rata kuadrat selisih antara nilai prediksi dan nilai sebenarnya, sehingga memberikan penekanan yang lebih besar pada kesalahan dengan deviasi yang tinggi [12].

2.9. Confusion Matrix

Confusion Matrix digunakan untuk mengevaluasi kinerja hasil penggabungan metode hibrida. Metode ini bekerja dengan membandingkan nilai prediksi dengan nilai aktual, sehingga menghasilkan empat kategori klasifikasi, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), yang merepresentasikan kondisi prediksi benar maupun salah pada masing-masing kelas [13].

3. HASIL DAN PEMBAHASAN

Data yang telah dikumpulkan kemudian melalui tahap pembersihan dan proses integrasi. Proses penggabungan data dilakukan dalam dua tahap. Tahap pertama menggabungkan file links.csv, movies.csv, dan dataset_tmdb_movies.csv menjadi satu kesatuan data.

Selanjutnya digunakan dalam proses pemodelan menggunakan metode Content-Based Filtering. Tahap kedua dilakukan dengan menggabungkan hasil integrasi data tersebut dengan file ratings.csv. Penggabungan ini menghasilkan data yang memuat informasi keterkaitan antara pengguna, film, dan nilai rating, sehingga dapat dimanfaatkan pada tahap pemodelan Matrix Factorization untuk mempelajari pola preferensi pengguna terhadap film.

3.1. Hasil Penerapan Content-Based Filtering

Rekomendasi Berdasarkan Genre untuk film: Harry Potter

	title_x	genres_x
276	The Equalizer	['Thriller', 'Action', 'Crime']
302	John Wick: Chapter 4	['Action', 'Thriller', 'Crime']
666	The Veteran	['Action', 'Crime', 'Thriller']
671	John Wick: Chapter 2	['Action', 'Thriller', 'Crime']
973	Kill Kane	['Thriller', 'Crime', 'Action']
1133	Bad Boys for Life	['Thriller', 'Action', 'Crime']
1219	The Equalizer 3	['Action', 'Thriller', 'Crime']
1279	Memory	['Action', 'Thriller', 'Crime']
1405	The Fate of the Furious	['Action', 'Crime', 'Thriller']
1446	War	['Action', 'Crime', 'Thriller']

Gambar 2. Hasil Penerapan Content-Based Filtering

Berdasarkan hasil pengujian, pendekatan Content-Based Filtering mampu menghasilkan rekomendasi film yang memiliki tingkat kemiripan konten Judul film digunakan sebagai kata kunci pencarian pada gambar 2. Selanjutnya, sistem mengekstraksi informasi genre dari film tersebut dan merepresentasikannya dalam bentuk vektor fitur. Vektor ini kemudian dibandingkan dengan vektor film lain yang terdapat dalam dataset berdasarkan atribut yang

telah ditetapkan, khususnya atribut *genres*, untuk mengukur tingkat kemiripan antarjudul. Kemiripan tersebut dihitung berdasarkan kesamaan genre film yang direpresentasikan dalam bentuk vektor TF-IDF, kemudian diukur menggunakan Cosine Similarity.

Pendekatan ini terbukti efektif sebagai tahap awal dalam proses rekomendasi, khususnya untuk memberikan saran film kepada pengguna baru yang belum memiliki riwayat rating (*cold-start*). Temuan ini sejalan dengan penelitian yang menyatakan bahwa Content-Based Filtering mampu memberikan rekomendasi relevan dengan memanfaatkan atribut item tanpa bergantung pada data pengguna lain [13].

3.2. Hasil Penerapan Matrix Factorization

Penerapan metode Matrix Factorization dengan teknik Singular Value Decomposition (SVD) menunjukkan bahwa model mampu mengekstraksi faktor laten yang merepresentasikan pola hubungan antara pengguna dan film berdasarkan data penilaian. Proses ini diawali dengan integrasi dataset utama dengan berkas *ratings.csv* sebagai sumber informasi rating yang diperlukan dalam pembentukan model prediktif. Selanjutnya, data dibagi menjadi himpunan pelatihan dan pengujian untuk keperluan evaluasi kinerja model. Pada tahap pelatihan, matriks interaksi pengguna film didekomposisi menjadi vektor-vektor laten melalui SVD, yang kemudian digunakan dalam proses pemodelan dan prediksi untuk mengukur tingkat akurasi sistem rekomendasi.

RMSE : 0.9105
MAE : 0.6945

Gambar 3. Hasil Evaluasi

Prediksi rating yang dihasilkan memiliki tingkat kesalahan yang relatif rendah, berdasarkan hasil evaluasi pada gambar 3 yang di tunjukkan pada MAE dan RMSE mengindikasikan bahwa model dapat merepresentasikan preferensi pengguna secara cukup akurat. Nilai RMSE sebesar 0,9105 dan MAE sebesar 0,6945 menunjukkan bahwa rata-rata kesalahan prediksi berada di bawah satu satuan skala rating. Hal ini mengindikasikan bahwa model Matrix Factorization berbasis SVD mampu menghasilkan estimasi rating yang cukup akurat dan merepresentasikan preferensi pengguna secara memadai. Temuan ini sejalan dengan hasil penelitian sebelumnya [13] yang menyatakan bahwa Matrix Factorization efektif dalam menangani data rating yang bersifat sparse serta mampu meningkatkan kualitas prediksi dalam sistem rekomendasi. Selain itu, pendekatan SVD juga dinilai efisien dalam pengolahan dataset berskala besar, sebagaimana dilaporkan oleh [13].

3.2.1. Hasil Penerapan Metode Hybrid

Penggabungan metode Content-Based Filtering dan Matrix Factorization membentuk suatu sistem rekomendasi hybrid yang mampu mengombinasikan keunggulan masing-masing pendekatan [14]. Content-Based Filtering digunakan pada tahap awal untuk menyaring kandidat film berdasarkan tingkat kemiripan atribut, khususnya genre, sedangkan Matrix Factorization berperan dalam melakukan pemeringkatan kandidat tersebut berdasarkan nilai prediksi rating pengguna. Dalam penelitian ini, skema hybrid diterapkan pada kasus prediksi untuk pengguna ke-2 terhadap film *Zootopia 2* yang merupakan judul baru,

movieId	judul Film	Genre	rating
290 68954.0	Up	[Animation, 'Comedy', 'Family', 'Adventure']	4.053617
2846 44022.0	Ice Age: The Meltdown	[Animation, 'Family', 'Comedy', 'Adventure']	3.963100
683 97913.0	Wreck-It Ralph	[Family, 'Animation', 'Comedy', 'Adventure']	3.890759
249 5218.0	Ice Age	[Animation, 'Comedy', 'Family', 'Adventure']	3.890082
593 45517.0	Cars	[Animation, 'Adventure', 'Comedy', 'Family']	3.889538
545 291815.0	PAW Patrol: The Mighty Movie	[Family, 'Comedy', 'Adventure', 'Animation']	3.850581
2651 179443.0	Tad, the Lost Explorer, and the Secret of King...	[Animation, 'Adventure', 'Comedy', 'Family']	3.848122
1323 201588.0	Toy Story 4	[Family, 'Comedy', 'Animation', 'Adventure']	3.753898
1960 253128.0	The Boss Baby: Family Business	[Animation, 'Comedy', 'Adventure', 'Family']	3.700579
332 33615.0	Madagascar	[Family, 'Animation', 'Adventure', 'Comedy']	3.652493

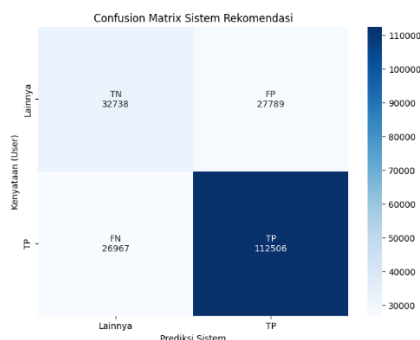
Gambar 4. Hasil Penerapan Metode Hybrid

Hasil evaluasi pada gambar 4 bahwa pendekatan hybrid mampu menghasilkan rekomendasi yang lebih relevan dan bersifat personal dibandingkan dengan penggunaan metode tunggal [15]. Dalam mekanisme sistem, tahap Content-Based Filtering menghasilkan daftar film yang relevan berdasarkan kemiripan terhadap judul yang digunakan sebagai kata kunci pencarian. Selanjutnya, film-film yang telah terpilih tersebut diproses menggunakan Matrix Factorization untuk memperkirakan nilai rating yang kemungkinan diberikan oleh pengguna, dalam hal ini pengguna

ke-2, apabila menonton film yang direkomendasikan. Dengan demikian, sistem tidak hanya menyajikan film yang serupa secara konten, tetapi juga mampu memprediksi tingkat ketertarikan pengguna secara kuantitatif melalui estimasi rating. maka Temuan ini sejalan dengan hasil penelitian [13] yang menyatakan bahwa pendekatan hybrid dapat meningkatkan akurasi dan stabilitas rekomendasi.

3.2.2. Evaluasi Kinerja Sistem

Evaluasi kinerja sistem rekomendasi hybrid dilakukan menggunakan confusion matrix dan metrik evaluasi seperti precision, recall, dan F1-score.



Gambar 5. Hasil Evaluasi Confusion Matrix

Berdasarkan hasil evaluasi pada gambar 5 menggunakan confusion matrix, terlihat bahwa jumlah prediksi benar positif (True Positive) sebesar 112.506 lebih tinggi secara signifikan dibandingkan dengan kesalahan prediksi berupa False Positive (27.789) dan False Negative (26.967). Temuan ini menunjukkan bahwa sistem rekomendasi hybrid memiliki kapabilitas yang baik dalam mengenali film yang sesuai dengan preferensi pengguna. Walaupun masih ditemukan sejumlah kesalahan klasifikasi, dominasi prediksi yang tepat mengindikasikan bahwa model mampu membedakan item relevan dan tidak relevan secara efektif, sehingga performa sistem dapat dinilai cukup andal dalam mendukung proses personalisasi rekomendasi.

	precision	recall	f1-score	support
Dislike	0.55	0.54	0.54	60527
Like	0.80	0.81	0.80	139473
accuracy			0.73	200000
macro avg	0.68	0.67	0.67	200000
weighted avg	0.73	0.73	0.73	200000

Gambar 6. Hasil Matrix Evaluasi

Selanjutnya, hasil pengujian menunjukkan bahwa sistem mencapai tingkat akurasi sebesar 73%. Pada kelas *like*, diperoleh nilai *precision* sebesar 0,80, *recall* sebesar 0,81, dan *F1-score* sebesar 0,81, yang mengindikasikan bahwa sistem memiliki kinerja yang cukup baik dalam merekomendasikan film yang sesuai dengan preferensi pengguna. Namun, pada kelas *dislike*, nilai *precision* dan *F1-score* masih berada pada tingkat yang relatif rendah, sehingga kemampuan sistem dalam mengenali film yang kurang diminati pengguna belum optimal. Temuan ini sejalan dengan penelitian [16] yang menyatakan bahwa pendekatan sistem rekomendasi hybrid umumnya menunjukkan performa yang lebih tinggi pada prediksi item relevan dibandingkan item tidak relevan, sehingga diperlukan pengembangan lebih lanjut untuk meningkatkan keseimbangan kinerja antar kelas.

4. KESIMPULAN

Penelitian ini berhasil menganalisis kinerja sistem rekomendasi film menggunakan metode hybrid Content-Based Filtering dan Matrix Factorization. Hasil penelitian menunjukkan bahwa pendekatan hybrid mampu meningkatkan kualitas rekomendasi dengan mengombinasikan kemiripan konten dan pola preferensi pengguna. Evaluasi kinerja sistem menunjukkan tingkat akurasi yang cukup baik, khususnya pada rekomendasi film yang disukai pengguna. Ke depan, penelitian ini dapat dikembangkan dengan menambahkan fitur konten lain, seperti deskripsi film atau ulasan pengguna, serta menerapkan teknik pembelajaran mendalam untuk meningkatkan performa sistem rekomendasi.

REFERENSI

- [1] D. Sukmana, R. G. Guntara, and M. R. Nugraha, "Peningkatan Akurasi Rekomendasi Film Menggunakan Neural Collaborative Filtering dengan Arsitektur RecommenderNet," *Jurnal Algoritma*, vol. 22, no. 2, Nov. 2025, doi: 10.33364/algoritma/v.22-2.3013.
- [2] E. Ryana Agustian and E. Prasetyo Nugroho, "Sistem Rekomendasi Film Menggunakan Metode Collaborative Filtering dan K-Nearest Neighbors Film Recommendation System Using Collaborative Filtering Method and K-Nearest Neighbors," 2020. [Online]. Available: <https://ejournal.upi.edu/index.php/JATIKOM>
- [3] F. Z. Munthe and F. A. Siregar, "Sistem Rekomendasi Film Berbasis Collaborative Filtering Menggunakan Algoritma Matrix Factorization (Studi Kasus Platform VIU)," *Hello World Jurnal Ilmu Komputer*, vol. 4, no. 3, pp. 151–162, Oct. 2025, doi: 10.56211/helloworld.v4i3.1164.
- [4] R. Aswilyarti *et al.*, "IMPLEMENTASI SISTEM REKOMENDASI PRODUK TREATMENT KECANTIKAN MENGGUNAKAN ALGORITMA SINGULAR VALUE DECOMPOSITION (STUDI KASUS : TYALASHES STUDIO)," 2025.
- [5] Y. LESTANTO and R. MUALIFA, "Development of a Data Cleaning System for Consumer Master Data using Sorted Neighborhood and N-Gram Methods," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 13, no. 1, p. 57, Feb. 2025, doi: 10.26760/elkomika.v13i1.57.
- [6] L. Maulidya, L. H. Suadaa, A. W. Wijayanto, and F. Ridho, "Multi-Source Data Fusion For Data Extraction and Integration of Scientific Publications in Academic Institution STIS," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 14, no. 2, pp. 348–363, Jul. 2025, doi: 10.23887/janapati.v14i2.87050.
- [7] Y. Christian and K. Kelvin, "RANCANG BANGUN APLIKASI KURSUS ONLINE BERBASIS WEB DENGAN SISTEM REKOMENDASI METODE CONTENT-BASED FILTERING," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 7, no. 1, pp. 23–36, Jan. 2022, doi: 10.36341/rabit.v7i1.2181.
- [8] M. Hafizh Mahendra, D. Triantoro Mardiansyah, and K. Muslim Lhaksana, "Dike : Jurnal Ilmu Multidisiplin Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer," 2023.
- [9] I. S. Wibowo, A. Witanti, and I. Susilawati, "Keyword Extraction Judul Berita Online Di Indonesia Menggunakan Metode TF-IDF", [Online]. Available: <http://jurnal.mdp.ac.id>
- [10] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *Jurnal Bumigora Information Technology (BITE)*, vol. 3, no. 1, pp. 19–27, Jul. 2021, doi: 10.30812/bite.v3i1.1110.
- [11] F. B. Arviyanto and A. A. Soebroto, "Sistem Rekomendasi Model Sandal Pada Home Industri Menggunakan Collaborative Filtering dan Algoritma Matrix Factorization," 2017. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [12] Roni Merdiansah, Khofifah Wulandari, Mentari Hasibuan, and Yuyun Umaidah, "Perbandingan Kinerja Model RNN, LSTM, dan BLSTM dalam Memprediksi Jumlah Gempa Bulanan di Indonesia," *Jurnal Penelitian Rumpun Ilmu Teknik*, vol. 3, no. 1, pp. 262–277, Feb. 2024, doi: 10.55606/juprit.v3i1.3466.
- [13] "Jurnal 13".
- [14] H. Darwis, F. A. Syahrir, and L. N. Hayati, "A Hybrid Movie Recommendation System to Address Data Sparsity Using Genre-Based K-Means and Neural Collaborative Filtering," *ILKOM Jurnal Ilmiah*, vol. 17, no. 2, pp. 203–212, Sep. 2025, doi: 10.33096/ilkom.v17i2.2868.203-212.
- [15] A. N. Hasoon, S. K. Abdulateef, R. S. Abdulameer, and M. L. Shuwandy, "An Intelligent Hybrid AI Course Recommendation Framework Integrating BERT Embeddings and Random Forest Classification," *Computers*, vol. 14, no. 9, Sep. 2025, doi: 10.3390/computers14090353.
- [16] Z. Z. Darban and M. H. Valipour, "GHRs: Graph-based Hybrid Recommendation System with Application to Movie Recommendation," Mar. 2022, doi: 10.1016/j.eswa.2022.116850.