Jurnal Ilmu Komputer dan Sistem Informasi (JIRSI) **e-ISSN 2830-3954**
Volume: 5, Nomor: 3, September 2026: 736-748 **p-ISSN 2830-6031**
https://jurnal.unity-academy.sch.id/index.php/jirsi/index

Implementasi Content-Based Filtering dan Singular Value Decomposition untuk Rekomendasi Film

Implementation of Content-Based Filtering and Singular Value Decomposition for Movie Recommendation

Muhammad Iqbal Pradipta*[1] Mhd. Basri[2]
[1,2] Teknologi Informasi/Universitas Muhammadiyah Sumatera Utara
E-mail: [1]muhammad.iqbalpradipta@umsu.ac.id [2]mhd.basri@umsu.ac.id

Abstrak

Pertumbuhan katalog layanan video daring menimbulkan kelebihan informasi yang menyulitkan pengguna menemukan film sesuai preferensi. Penelitian ini bertujuan membangun dan mengevaluasi dua pendekatan sistem rekomendasi film, yaitu content-based filtering menggunakan TF-IDF dan cosine similarity serta collaborative filtering menggunakan Singular Value Decomposition (SVD). Data penelitian mencakup 4.803 film beserta kredit produksi dan 100.004 interaksi rating pengguna. Tahapan penelitian meliputi pembersihan dan integrasi data, pembentukan representasi konten, pemodelan faktor laten, pembagian data latih dan uji sebesar 75:25, serta evaluasi menggunakan Precision dan Root Mean Squared Error. Hasil pengujian menunjukkan bahwa rekomendasi berbasis konten memperoleh Precision 0,70 berdasarkan kesesuaian genre, sedangkan rekomendasi berbasis sutradara dan aktor mencapai presisi 1,00 karena fitur pemeringkatan identik dengan kriteria relevansi. Model SVD menghasilkan RMSE 0,9007 dan mampu memberikan rekomendasi personal berdasarkan pola rating pengguna. Kontribusi penelitian terletak pada evaluasi dua baseline dengan tujuan berbeda, analisis potensi bias pada pengukuran presisi, serta penegasan pentingnya konsistensi identitas film lintas dataset. Content-based filtering efektif untuk menemukan film serupa berdasarkan metadata, sedangkan SVD mampu memodelkan preferensi laten pengguna, tetapi hasil kedua pendekatan harus ditafsirkan menggunakan metrik yang sesuai.

Kata kunci: *content-based filtering, cosine similarity, rekomendasi film, SVD, TF-IDF*

Abstract

The growth of online video catalogs creates information overload that makes it difficult for users to discover movies matching their preferences. This study develops and evaluates two movie recommender approaches: content-based filtering using TF-IDF and cosine similarity, and collaborative filtering using Singular Value Decomposition (SVD). The dataset includes 4,803 movies with production-credit records and 100,004 user-rating interactions. The research stages comprise data cleaning and integration, content representation, latent-factor modeling, a 75:25 train-test split, and evaluation using Precision and Root Mean Squared Error. The content-based experiment achieved a Precision of 0.70 based on genre overlap, while director- and actor-based recommendations obtained a precision of 1.00 because the ranking features were identical to the relevance criteria. The SVD model achieved an RMSE of 0.9007 and generated personalized recommendations from user-rating patterns. The study contributes an evaluation of two baselines with different objectives, an analysis of potential bias in precision measurement, and an emphasis on consistent movie identifiers across datasets. Content-based filtering is effective for retrieving movies with similar metadata,

whereas SVD captures latent user preferences. However, the results of the two approaches must be interpreted using metrics appropriate to their respective objectives.

Keywords: *content-based filtering, cosine similarity, movie recommendation, SVD, TF-IDF*

1. PENDAHULUAN

Pertumbuhan katalog pada platform video daring memberikan semakin banyak pilihan kepada pengguna, tetapi tidak selalu mempermudah pengambilan keputusan. Pengguna harus menyeleksi ribuan judul berdasarkan informasi seperti sinopsis, genre, pemeran, sutradara, dan rating. Kondisi tersebut memunculkan *information overload*, yaitu keadaan ketika banyaknya informasi justru menyulitkan pengguna menemukan konten yang sesuai dengan preferensinya. Sistem rekomendasi dikembangkan sebagai mekanisme penyaringan untuk memprioritaskan item berdasarkan karakteristik konten, pola interaksi pengguna, atau kombinasi keduanya. Kajian sistematis menunjukkan bahwa rekomendasi film menjadi salah satu bidang penerapan sistem rekomendasi yang banyak diteliti karena tersedianya dataset publik serta beragamnya metode dan metrik evaluasi [11].

Salah satu pendekatan yang umum digunakan adalah *content-based filtering*, yaitu rekomendasi berdasarkan kemiripan atribut antara suatu film dan film yang sebelumnya diminati pengguna. Atribut tersebut dapat berupa sinopsis, genre, kata kunci, sutradara, pemeran, bahasa, dan informasi visual. Pendekatan ini tetap dapat menghasilkan rekomendasi ketika data interaksi pengguna terbatas, tetapi cenderung memberikan item yang terlalu serupa dengan riwayat pengguna. Bendouch, Frasincar, dan Robal [4] menunjukkan bahwa pengayaan fitur visual dan semantik dapat memperluas representasi konten. Dalam implementasi yang lebih sederhana dan transparan, TF-IDF dapat digunakan untuk membentuk representasi numerik sinopsis, sedangkan *cosine similarity* digunakan untuk mengukur kemiripan antarfilm.

Pendekatan lain adalah *collaborative filtering*, yang memanfaatkan pola rating atau interaksi kolektif untuk memperkirakan preferensi pengguna terhadap film yang belum dinilai. Pendekatan ini tidak memerlukan analisis langsung terhadap isi film, tetapi kinerjanya bergantung pada ketersediaan interaksi dalam matriks pengguna-item. Metode *matrix factorization*, termasuk *Singular Value Decomposition* (SVD), memproyeksikan pengguna dan film ke ruang faktor laten berdimensi lebih rendah. Metode tersebut masih digunakan sebagai *baseline* karena memiliki akurasi yang kompetitif dan implementasi yang relatif efisien [2], [3]. Selain itu, pendekatan berbasis graf dan fitur temporal telah dikembangkan untuk menangkap hubungan pengguna-item yang lebih kompleks serta perubahan preferensi dari waktu ke waktu [2], [6].

Penelitian terdahulu menunjukkan bahwa pendekatan berbasis konten dan kolaboratif memiliki karakteristik yang berbeda. *Content-based filtering* cenderung lebih stabil ketika interaksi pengguna terbatas, sedangkan *collaborative filtering*

memperoleh manfaat ketika tersedia data rating yang memadai [5]. Airen dan Agrawal [1] menggunakan *co-clustering* dan penyetelan parameter untuk meningkatkan rekomendasi. Yang, Woradit, dan Cosh [8] mengembangkan sistem hibrida berbasis partisi pengguna dan perbandingan konten, sedangkan Sarhan et al. [9] mengintegrasikan pembelajaran mesin dan analisis sentimen. Perkembangan tersebut menunjukkan kecenderungan penggunaan model yang semakin kompleks untuk meningkatkan personalisasi.

Namun, peningkatan kompleksitas model tidak selalu menjamin evaluasi yang lebih valid. Nilai presisi yang tinggi dapat muncul apabila fitur yang digunakan untuk menentukan peringkat sama dengan kriteria yang digunakan untuk menetapkan relevansi. Sebagai contoh, sistem yang mengurutkan film berdasarkan kesamaan sutradara akan memperoleh presisi tinggi apabila relevansi juga ditentukan hanya berdasarkan kesamaan sutradara. Selain itu, integrasi dataset metadata dan rating dapat menghasilkan hubungan film yang tidak tepat apabila keduanya menggunakan skema identitas yang berbeda dan tidak diselaraskan melalui *movie identifier* yang konsisten.

Berdasarkan kesenjangan tersebut, penggunaan TF-IDF, *cosine similarity*, dan SVD dalam penelitian ini tidak diposisikan sebagai kebaruan algoritmik karena telah banyak digunakan sebagai metode *baseline*. Inovasi penelitian terletak pada kerangka evaluasi transparan yang membedakan tujuan rekomendasi berbasis kemiripan konten dan rekomendasi personal berbasis faktor laten. Kerangka tersebut juga mengkaji potensi bias pada pengukuran presisi dan menempatkan konsistensi identitas film lintas dataset sebagai bagian dari validitas hasil rekomendasi.

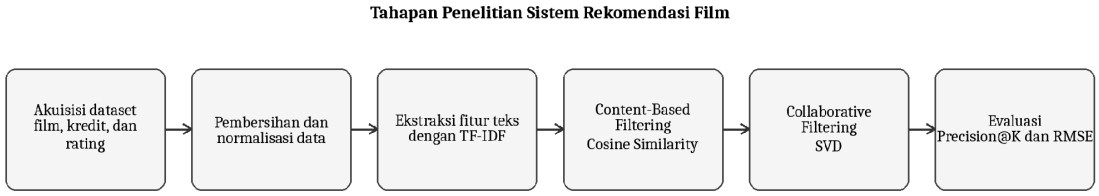
Penelitian ini memberikan tiga kontribusi. Pertama, penelitian menyediakan alur pengolahan data yang dapat direplikasi, meliputi pembersihan data, ekstraksi metadata, pembentukan representasi konten, dan pemodelan faktor laten. Kedua, penelitian memberikan bukti empiris mengenai karakteristik hasil *content-based filtering* dan SVD ketika diterapkan pada data film dan rating pengguna. Ketiga, penelitian memberikan kontribusi evaluatif-metodologis dengan menunjukkan bahwa Precision dan RMSE mengukur keluaran yang berbeda sehingga tidak dapat dibandingkan secara langsung.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membangun rekomendasi film berbasis konten menggunakan sinopsis dan metadata film. Membangun rekomendasi personal menggunakan SVD berdasarkan pola rating pengguna, dan mengevaluasi kedua pendekatan secara proporsional berdasarkan tujuan model, definisi relevansi, pemilihan metrik, dan konsistensi identitas film lintas dataset.

2. METODOLOGI PENELITIAN

Penelitian menggunakan desain eksperimen komputasional. Dataset diolah dengan Python melalui pustaka pandas untuk manipulasi data, scikit-learn untuk representasi TF-IDF dan perhitungan cosine similarity, serta Surprise untuk

pemodelan SVD. Alur penelitian ditunjukkan pada Gambar 1. Setiap cabang model diperlakukan sebagai sistem terpisah karena sumber sinyal dan metrik evaluasinya berbeda.



Gambar 1. Tahapan penelitian sistem rekomendasi film

2.1. Dataset dan Unit Analisis

Data film diperoleh dari himpunan data berbasis The Movie Database (TMDB) yang disediakan melalui repositori Hugging Face [13]. Data tersebut mencakup tabel movies dan credits. Interaksi pengguna berasal dari ratings_small yang mengikuti struktur MovieLens. MovieLens merupakan dataset yang luas digunakan untuk pengembangan dan evaluasi sistem rekomendasi [12]. Ringkasan ukuran dataset disajikan pada Tabel 1.

Tabel 1. Ringkasan dataset penelitian

Dataset	Jumlah baris	Jumlah kolom	Fungsi utama
movies	4.803	20	Metadata film: judul, sinopsis, genre, popularitas, runtime, dan atribut produksi
credits	4.803	4	Pemeran dan kru dalam struktur JSON
ratings_small	100.004	4	Interaksi userId, movieId, rating, dan timestamp

Unit analisis pada cabang content-based adalah film, sedangkan unit analisis pada cabang collaborative filtering adalah pasangan pengguna-film. Rating berada pada skala 0,5 sampai 5,0. Atribut metadata bertipe JSON terlebih dahulu diparsing menjadi daftar nilai yang dapat dianalisis. Kolom overview menjadi sumber teks utama karena memuat ringkasan cerita, sedangkan director dan cast_names digunakan sebagai fitur identitas tambahan.

2.2. Pra-pemrosesan Data

Pra-pemrosesan bertujuan mencegah kegagalan transformasi dan memastikan setiap fitur memiliki representasi konsisten. Pemeriksaan awal menunjukkan nilai hilang pada homepage, overview, release_date, runtime, dan tagline. Nilai hilang pada kolom yang tidak digunakan untuk pemodelan utama tidak perlu dipaksakan menjadi data semu. Untuk overview, nilai kosong diubah menjadi string kosong sebelum normalisasi teks, runtime dapat diimputasi dengan nilai rata-rata apabila digunakan dalam analisis deskriptif. Ringkasan perlakuan data ditunjukkan pada Tabel 2.

Tabel 2. Masalah kualitas data dan perlakuan

Atribut	Nilai hilang	Perlakuan	Alasan
homepage	3.091	Dibiarkan kosong/tidak digunakan	Tidak berkontribusi langsung pada model
overview	3	Diisi string kosong lalu disaring	TF-IDF membutuhkan masukan teks valid
release_date	1	Dibiarkan kosong atau dihapus sesuai analisis	Tidak digunakan sebagai fitur model utama
runtime	2	Imputasi mean untuk analisis deskriptif	Menjaga tipe numerik dan jumlah observasi
tagline	844	Diisi string kosong bila digunakan	Tagline tidak tersedia untuk seluruh film

Normalisasi overview dilakukan dengan mengubah huruf menjadi lowercase dan menghapus angka serta karakter nonalfabet. Baris dengan teks kosong setelah normalisasi dikeluarkan dari pemodelan konten. Data cast, crew, genres, production_companies, dan production_countries yang tersimpan sebagai string JSON dikonversi menjadi daftar. Nama sutradara diekstraksi dari anggota crew dengan job Director. Duplikasi judul ditangani dengan mempertahankan satu observasi pertama agar indeks judul tidak ambigu saat pencarian rekomendasi.

Aspek penting yang harus diperhatikan adalah identitas film. movieId pada MovieLens bukan secara otomatis sama dengan id TMDB. Pemetaan yang benar memerlukan tabel penghubung, misalnya links.csv yang mengaitkan movieId dengan tmdbId. Tanpa normalisasi identitas, pelatihan SVD terhadap rating tetap dapat berjalan, tetapi pemberian judul pada hasil rekomendasi tidak dapat dipastikan benar. Oleh karena itu, penelitian ini memisahkan evaluasi numerik SVD dari interpretasi judul dan menjadikan konsistensi ID sebagai batas validitas utama.

2.3. Content-Based Filtering

Content-based filtering dibangun dari tiga kelompok fitur, yaitu sinopsis yang telah dibersihkan, nama sutradara, dan nama aktor. Untuk sinopsis, TF-IDF menghitung bobot sebuah istilah berdasarkan frekuensinya pada satu dokumen dan kelangkaannya pada seluruh korpus. Jika tf(t,d) adalah frekuensi istilah t pada dokumen d dan df(t) adalah jumlah dokumen yang memuat istilah t, maka bentuk umum bobot TF-IDF dapat dinyatakan sebagai berikut:

$$TF\text{-}IDF(t,d) = tf(t,d) \times \log(N / df(t)) \quad (1)$$

Setelah setiap film direpresentasikan sebagai vektor, cosine similarity digunakan untuk menghitung kemiripan antara film i dan j. Nilai mendekati 1 menunjukkan arah vektor yang sangat serupa, sedangkan nilai mendekati 0 menunjukkan kesamaan yang rendah. Rumusnya adalah:

$$\cos(i,j) = (v_i \cdot v_j) / (||v_i|| \, ||v_j||) \quad (2)$$

Daftar rekomendasi dibentuk dengan mengambil K film berperingkat tertinggi setelah film acuan dikeluarkan. Untuk fitur sutradara dan aktor, TF-IDF pada token nama menghasilkan kemiripan maksimum ketika identitas yang sama muncul.

Karena itu, keluaran tersebut sebaiknya dipahami sebagai retrieval berbasis atribut, bukan bukti bahwa keseluruhan isi film identik.

Tabel 3. Fitur content-based dan maknanya

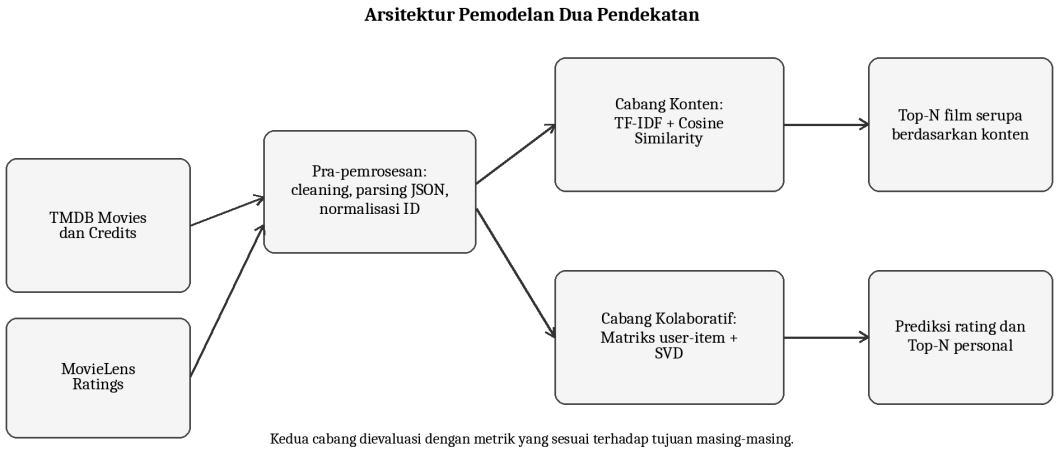
Fitur	Representasi	Sinyal kemiripan	Risiko
cleaned_overview	TF-IDF kata	Kesamaan tema dan narasi	Sinonim dan konteks semantik dapat terlewat
director	Token nama sutradara	Kesamaan pembuat film	Nilai 1,0 bersifat deterministik untuk nama sama
cast_names	Token nama pemeran	Kesamaan pemeran	Film berbeda dapat dianggap identik pada satu aktor

2.4. Collaborative Filtering dengan SVD

SVD digunakan untuk memodelkan hubungan laten antara pengguna dan film dari matriks rating. Data rating dikonversi ke format Surprise dengan skala 0,5-5,0, kemudian dibagi menjadi 75% data latih dan 25% data uji. Model mempelajari bias global, bias pengguna, bias item, serta vektor faktor laten. Prediksi rating dapat dituliskan secara ringkas sebagai:

$$r\_hat(u,i) = mu + b\_u + b\_i + q\_i\text{^}T\ p\_u \quad (3)$$

dengan mu sebagai rata-rata global, b_u sebagai bias pengguna, b_i sebagai bias film, serta p_u dan q_i sebagai vektor faktor laten pengguna dan film. Setelah model dilatih, rating diprediksi untuk film yang belum diberi rating oleh pengguna. Film dengan estimasi tertinggi dipilih menjadi Top-N recommendation. Arsitektur dua cabang model disajikan pada Gambar 2.



Gambar 2. Arsitektur pemodelan dua pendekatan

2.5. Metode Evaluasi

Evaluasi content-based menggunakan Precision dengan definisi relevansi eksplisit. Untuk pengujian sinopsis, sebuah rekomendasi dianggap relevan apabila memiliki sedikitnya satu genre yang sama dengan film acuan. Precision menghitung proporsi item relevan pada K rekomendasi teratas:

$$Precision = jumlah\ item\ relevan\ pada\ Top\text{-}K\ /\ K \quad (4)$$

Untuk hasil berbasis sutradara dan aktor, relevansi ditentukan dari kesamaan identitas fitur yang sama. Nilai ini bersifat sanity check karena kriteria evaluasi identik dengan fitur pembentuk ranking. Evaluasi SVD menggunakan RMSE untuk mengukur deviasi prediksi rating terhadap rating aktual:

$$RMSE = sqrt((1/n) sum((r\_ui - r\_hat\_ui)^2)) \quad (5)$$

RMSE yang lebih rendah menunjukkan prediksi lebih dekat terhadap rating aktual. Namun, RMSE tidak mengukur kualitas urutan Top-N secara langsung. Oleh karena itu, Precision dan RMSE tidak digunakan untuk menentukan model mana yang mutlak lebih baik, keduanya hanya menilai tujuan yang berbeda.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Persiapan dan Representasi Data

Setelah pembersihan, overview yang valid diubah menjadi matriks TF-IDF. Istilah dengan rata-rata bobot tinggi pada korpus antara lain love, life, story, time, world, dan man. Nilai rata-rata TF-IDF tidak dapat langsung ditafsirkan sebagai popularitas kata, nilai tersebut menunjukkan kontribusi relatif istilah setelah mempertimbangkan frekuensi dokumen. Tabel 4 menyajikan sepuluh istilah yang dilaporkan memiliki rata-rata bobot tertinggi pada eksperimen.

Tabel 4. Contoh istilah dengan rata-rata bobot TF-IDF tinggi

Peringkat	Istilah	Rata-rata TF-IDF
1	love	0,018286
2	life	0,014836
3	story	0,012618
4	time	0,011256
5	world	0,011158
6	man	0,010973
7	just	0,009113
8	hes	0,007696
9	way	0,007256
10	new	0,007132

Kehadiran token hes menunjukkan bahwa normalisasi yang hanya menghapus tanda baca dapat menggabungkan kontraksi bahasa Inggris, misalnya he's menjadi hes. Temuan ini mengindikasikan bahwa preprocessing dapat ditingkatkan menggunakan tokenisasi, stop-word removal, lemmatization, atau pemodelan berbasis embedding. Dengan demikian, TF-IDF memberikan baseline yang transparan, tetapi kualitas semantik sangat dipengaruhi aturan pembersihan teks.

3.2. Hasil Rekomendasi Berdasarkan Sinopsis

Pengujian content-based menggunakan sebuah film acuan dan mengambil sepuluh film dengan cosine similarity tertinggi. Hasil yang dilaporkan ditunjukkan pada Tabel 5. Taken 3 menempati urutan pertama dengan skor 0,747440, sedangkan film pada peringkat kesepuluh memperoleh skor 0,443218. Rentang ini

menunjukkan bahwa rekomendasi pertama memiliki kemiripan vektor lebih kuat daripada item lain, tetapi skor tersebut tidak dapat dibaca sebagai probabilitas kesukaan pengguna.

Tabel 5. Top-10 rekomendasi berdasarkan cleaned_overview

No.	Judul film	Cosine similarity
1	Taken 3	0,747440
2	Veer-Zaara	0,549093
3	Coyote Ugly	0,537654
4	Hercules	0,524978
5	Stoker	0,516975
6	Dark City	0,491365
7	The Black Hole	0,464159
8	Pan	0,463130
9	From Hell	0,443636
10	Coach Carter	0,443218

Validasi berbasis genre mengidentifikasi tujuh dari sepuluh film sebagai relevan terhadap genre film acuan. Rinciannya ditunjukkan pada Tabel 6. Definisi relevansi berupa minimal satu genre yang beririsan merupakan aturan yang sederhana dan dapat direplikasi, tetapi tidak menangkap tingkat kedekatan tema. Film dengan satu genre yang sama diperlakukan setara dengan film yang memiliki tiga genre yang sama. Selain itu, sinopsis dapat menghasilkan kesamaan semantik meskipun genre resminya berbeda.

Tabel 6. Penilaian relevansi berbasis irisan genre

Judul	Genre	Relevan
Taken 3	Thriller, Action	Ya
Veer-Zaara	Drama, Romance	Ya
Coyote Ugly	Comedy	Tidak
Hercules	Action, Adventure	Ya
Stoker	Drama, Horror, Thriller	Ya
Dark City	Mystery, Science Fiction	Tidak
The Black Hole	Adventure, Family, Science Fiction, Action	Ya
Pan	Adventure, Family, Fantasy	Tidak
From Hell	Horror, Mystery, Thriller	Ya
Coach Carter	Drama	Ya

Berdasarkan tujuh item relevan dari sepuluh rekomendasi, Precision di 10 adalah 7/10 atau 0,70. Nilai ini menunjukkan bahwa mayoritas rekomendasi memiliki irisan genre, tetapi belum dapat digeneralisasi sebagai performa keseluruhan sistem karena pengujian hanya menggunakan satu film acuan dan

penilaian relevansi berbasis aturan. Evaluasi yang lebih kuat membutuhkan banyak query film, anotasi pengguna, serta metrik seperti Recall di K, NDCG di K, diversity, dan coverage.

3.3. Hasil Rekomendasi Berdasarkan Sutradara dan Aktor

Eksperimen berbasis nama sutradara menghasilkan delapan judul yang seluruhnya disutradarai Christopher Nolan. Karena vektor dibentuk dari token sutradara yang sama, cosine similarity bernilai 1,0. Hasil ini berguna untuk memeriksa bahwa pipeline ekstraksi crew bekerja, tetapi tidak menunjukkan kemiripan pada aspek lain seperti genre, periode produksi, atau gaya naratif. Tabel 7 menyajikan hasil tersebut.

Tabel 7. Rekomendasi berdasarkan sutradara Christopher Nolan

No.	Judul film	Cosine similarity
1	The Dark Knight	1,0
2	Interstellar	1,0
3	Inception	1,0
4	Batman Begins	1,0
5	Insomnia	1,0
6	The Prestige	1,0
7	Memento	1,0
8	The Dark Knight Rises	1,0

Eksperimen aktor menggunakan Christian Bale sebagai atribut acuan dan menghasilkan sepuluh film dengan skor 1,0. Secara operasional, seluruh item relevan karena memuat aktor yang sama, sehingga Precision di 10 adalah 1,00. Akan tetapi, nilai tersebut bersifat tautologis: model mengurutkan berdasarkan identitas aktor dan evaluasi juga menyatakan relevan berdasarkan identitas aktor. Oleh sebab itu, hasil tersebut tidak boleh dibandingkan langsung dengan Precision di 10 sinopsis.

Tabel 8. Rekomendasi berdasarkan aktor Christian Bale

No.	Judul film	Cosine similarity
1	The Big Short	1,0
2	Out of the Furnace	1,0
3	Public Enemies	1,0
4	The Fighter	1,0
5	Rescue Dawn	1,0
6	The Dark Knight Rises	1,0
7	The Dark Knight	1,0
8	Pocahontas	1,0
9	3:10 to Yuma	1,0
10	Equilibrium	1,0

Keuntungan fitur identitas adalah interpretabilitas: pengguna dapat memahami bahwa rekomendasi muncul karena aktor atau sutradara yang sama. Kekurangannya adalah kurangnya diskriminasi ketika banyak film berbagi atribut tersebut. Sistem produksi sebaiknya menggabungkan beberapa fitur dalam satu representasi dengan bobot terkontrol, misalnya sinopsis, genre, sutradara, aktor utama, dan popularitas. Penggabungan tersebut perlu divalidasi agar fitur dengan token langka tidak mendominasi seluruh skor.

3.4. Hasil Collaborative Filtering dengan SVD

Model SVD dilatih pada 75% data rating dan diuji pada 25% sisanya. Eksperimen menghasilkan RMSE 0,9007. Pada skala rating 0,5-5,0, nilai tersebut berarti akar rata-rata kuadrat selisih prediksi sekitar 0,90 poin rating. Nilai ini menunjukkan baseline yang berfungsi, tetapi belum cukup untuk menyatakan model optimal karena tidak tersedia pembanding dengan mean predictor, KNN, NMF, atau konfigurasi SVD lain. Literatur menunjukkan bahwa kinerja dekomposisi matriks sensitif terhadap kepadatan data, jumlah faktor laten, regularisasi, dan dinamika waktu [2], [3].

Untuk userId 3, implementasi melaporkan tiga film dengan prediksi tertinggi, yaitu Beetlejuice, Galaxy Quest, dan The Good Thief. Daftar ini dapat dipandang sebagai contoh keluaran personal. Namun, kesesuaian judul hanya valid apabila movieId rating telah dipetakan dengan benar ke identitas film pada metadata. Jika proses menggabungkan movieId langsung dengan id TMDB tanpa tabel pemetaan, judul tersebut tidak dapat dikonfirmasi. Evaluasi RMSE tetap merefleksikan prediksi pada pasangan ID rating, sedangkan interpretasi nama film harus ditunda sampai pemetaan ID diverifikasi.

Tabel 9. Ringkasan hasil evaluasi

Pendekatan	Objek evaluasi	Metrik	Hasil	Interpretasi terbatas
TF-IDF + cosine	Top-10 sinopsis	Precision di 10	0,70	7 dari 10 film memiliki irisan genre
Token sutradara	8 film	Precision di 8	1,00	Sanity check karena aturan relevansi sama dengan fitur
Token aktor	Top-10 film	Precision di 10	1,00	Sanity check karena aturan relevansi sama dengan fitur
SVD	Rating data uji	RMSE	0,9007	Deviasi prediksi pada skala rating, bukan kualitas ranking

3.5. Analisis Perbandingan dan Implikasi

Kedua pendekatan menyelesaikan masalah yang berbeda. Content-based beroperasi pada atribut film dan dapat digunakan ketika pengguna belum memiliki banyak interaksi. SVD memerlukan riwayat rating, tetapi mampu menemukan faktor laten yang tidak eksplisit pada metadata. Temuan ini konsisten dengan

penelitian yang membedakan kebutuhan data content-based dan collaborative filtering [5]. Sistem produksi dapat menjalankan keduanya secara paralel: content-based untuk item similarity dan cold-start item, serta SVD untuk personalisasi pengguna yang telah aktif.

Hasil tidak mendukung kesimpulan bahwa satu model lebih akurat dari model lain karena metriknya berbeda. Precision di 10 menilai relevansi daftar berdasarkan aturan genre, sedangkan RMSE menilai galat prediksi rating. Perbandingan yang adil memerlukan sasaran yang sama, misalnya kedua model menghasilkan Top-10 lalu dinilai dengan Precision di 10 dan NDCG di 10 pada ground truth yang sama. Studi terbaru juga menunjukkan bahwa model kompleks seperti graph neural network, model temporal, dan sentiment-aware recommendation dapat meningkatkan representasi, tetapi harus dibandingkan terhadap baseline dengan protokol evaluasi konsisten [2], [6], [9].

Aspek skalabilitas dan privasi juga relevan. SVD pada dataset besar membutuhkan komputasi matriks yang lebih tinggi, penelitian lain mengeksplorasi dekomposisi alternatif dan mekanisme outsourcing aman [3], [7]. Walaupun dataset penelitian ini berukuran moderat, rancangan aplikasi nyata perlu mempertimbangkan pembaruan model, penyimpanan interaksi, anonimisasi pengguna, dan audit rekomendasi. Kualitas sistem tidak hanya ditentukan akurasi, tetapi juga diversity, novelty, fairness, explainability, dan kemampuan menangani cold-start.

Tabel 10. Keterbatasan dan arah perbaikan

Keterbatasan	Dampak	Perbaikan yang disarankan
Pengujian sinopsis hanya pada satu film acuan	Nilai Precision di 10 belum mewakili seluruh katalog	Gunakan banyak query dan agregasi macro/micro
Relevansi berbasis minimal satu genre	Penilaian dapat terlalu longgar	Gunakan anotasi pengguna atau graded relevance
Evaluasi aktor/sutradara tautologis	Presisi 1,00 tidak mengukur kualitas umum	Gabungkan fitur dan evaluasi pada preferensi nyata
Pemetaan movieId dan TMDB id belum dibuktikan	Judul rekomendasi SVD dapat salah	Gunakan links.csv atau kunci pemetaan resmi
Belum ada tuning dan baseline pembanding	RMSE belum menunjukkan model terbaik	Lakukan cross-validation dan grid search
Belum mengukur diversity dan novelty	Daftar dapat homogen	Tambahkan reranking multiobjektif

KESIMPULAN

Penelitian ini berhasil membangun dua *baseline* sistem rekomendasi film yang memiliki fungsi berbeda. Pendekatan *content-based filtering* menggunakan TF-IDF dan *cosine similarity* mampu menghasilkan daftar film yang serupa berdasarkan sinopsis, sutradara, dan aktor. Pada pengujian menggunakan satu film acuan, rekomendasi berbasis sinopsis memperoleh nilai Precision@10 sebesar 0,70 berdasarkan kesesuaian genre. Rekomendasi berdasarkan sutradara dan aktor menghasilkan presisi 1,00, tetapi hasil tersebut lebih tepat diposisikan sebagai verifikasi alur sistem karena fitur pencarian identik dengan kriteria relevansi.

Sementara itu, pendekatan *collaborative filtering* menggunakan SVD menghasilkan RMSE sebesar 0,9007 pada data uji sebesar 25% dan mampu menghasilkan rekomendasi personal berdasarkan pola rating pengguna.

Hasil penelitian menunjukkan bahwa *content-based filtering* dan SVD bersifat komplementer karena keduanya menyelesaikan tujuan rekomendasi yang berbeda dan tidak dapat dibandingkan secara langsung menggunakan metrik yang berbeda. Validitas rekomendasi SVD juga bergantung pada konsistensi pemetaan `movieId` antara dataset rating dan metadata film. Oleh karena itu, hasil penelitian ini dapat menjadi dasar pengembangan sistem rekomendasi hibrida yang menggabungkan informasi konten dan preferensi laten pengguna.

Pada penelitian selanjutnya, pemetaan identitas film perlu diperbaiki menggunakan links.csv atau tabel pemetaan resmi. Pengembangan berikutnya juga dapat mencakup *cross-validation*, penyetelan parameter, evaluasi Top-N dengan *ground truth* yang sama, serta pengukuran *diversity*, *novelty*, *coverage*, dan kepuasan pengguna. Sistem dapat dikembangkan lebih lanjut dengan menambahkan konteks waktu, popularitas, bahasa, genre, sentimen ulasan, dan mekanisme penjelasan rekomendasi. Dari sisi penerapan, model dapat diintegrasikan ke dalam aplikasi katalog film atau platform video daring dan diadaptasi untuk rekomendasi buku, musik, berita, serta produk digital lainnya.

DAFTAR PUSTAKA

[1] S. Airen and J. Agrawal, “Movie Recommender System Using Parameter Tuning of User and Movie Neighbourhood via Co-Clustering,” Procedia Computer Science, vol. 218, pp. 1176-1183, 2023, doi: 10.1016/j.procs.2023.01.096.
[2] G. Behera and N. Nain, “Collaborative Filtering with Temporal Features for Movie Recommendation System,” Procedia Computer Science, vol. 218, pp. 1366-1373, 2023, doi: 10.1016/j.procs.2023.01.115.
[3] F. Horasan, A. H. Yurttakal, and S. Gündüz, “A novel model based collaborative filtering recommender system via truncated ULV decomposition,” Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 8, art. 101724, 2023, doi: 10.1016/j.jksuci.2023.101724.
[4] M. M. Bendouch, F. Frasincar, and T. Robal, “A visual-semantic approach for building content-based recommender systems,” Information Systems, vol. 117, art. 102243, 2023, doi: 10.1016/j.is.2023.102243.
[5] H. H. Kurniawan, W. S. Lukman, R. Fredyan, and M. A. Ibrahim, “Movie Recommendation System: A Comparison of Content-Based and Collaborative Filtering,” Procedia Computer Science, vol. 245, pp. 860-868, 2024, doi: 10.1016/j.procs.2024.10.313.
[6] R. Nesmaoui, M. Louhichi, and M. Lazaar, “A Collaborative Filtering Movies Recommendation System based on Graph Neural Network,” Procedia Computer Science, vol. 220, pp. 456-461, 2023, doi: 10.1016/j.procs.2023.03.058.
[7] Y. Tao, F. Kong, Y. Shi, J. Yu, and X. Wang, “Efficient, secure and verifiable outsourcing scheme for SVD-based collaborative filtering recommender system,” Future Generation Computer Systems, vol. 149, pp. 445-454, 2023, doi: 10.1016/j.future.2023.07.042.

[8] Y. Yang, K. Woradit, and K. Cosh, "Hybrid Movie Recommendation System With User Partitioning and Log Likelihood Content Comparison," IEEE Access, vol. 13, pp. 11609-11622, 2025, doi: 10.1109/ACCESS.2025.3529515.
[9] A. M. Sarhan, H. Ayman, M. Wagdi, B. Ali, A. Adel, and R. Osama, "Integrating machine learning and sentiment analysis in movie recommendation systems," Journal of Electrical Systems and Information Technology, vol. 11, art. 53, 2024, doi: 10.1186/s43067-024-00177-7.
[10] F. Fkih, "Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 9, pp. 7645-7669, 2022, doi: 10.1016/j.jksuci.2021.09.014.
[11] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," Journal of Big Data, vol. 9, art. 59, 2022, doi: 10.1186/s40537-022-00592-5.
[12] F. M. Harper and J. A. Konstan, "The MovieLens Datasets: History and Context," ACM Transactions on Interactive Intelligent Systems, vol. 5, no. 4, art. 19, 2015, doi: 10.1145/2827872.
[13] M. I. Pradipta, "Movie Dataset," Hugging Face Datasets, 2024. [Online]. Available: https://huggingface.co/datasets/mhdiqbalpradipta/movie. [Accessed: Jul. 24, 2026].