

Sistem Rekomendasi Film dengan Collaborative Filtering Berbasis SVD, KNN dan Explainable AI pada Dataset MovieLens

Movie Recommendation System with Collaborative Filtering, SVD, KNN, and Explainable AI on MovieLens Dataset

Rizky Adhita Saputra*, Nugroho Dwi Saputro, Khoiriya Latifah

Fakultas Teknik dan Informatika, Program Studi Informatika, Universitas PGRI Semarang, Semarang, Indonesia

Email: ¹*rizkyadhisaputra25@gmail.com, ²nugputra@upgris.ac.id, ³khairiyatlatifah@upgris.ac.id

Email Penulis Korespondensi: rizkyadhisaputra25@gmail.com

Received: 31/01/2026, Revised: 15/09/2026, Accepted: 25/09/2026, Published: 26/09/2026

Abstrak—Pertumbuhan pesat platform *streaming* memicu fenomena *information overload*, yang membutuhkan sistem rekomendasi efektif. Namun, metode *Collaborative Filtering* konvensional sering menghadapi hambatan *data sparsity* ekstrem, mencapai 98,47% pada dataset MovieLens, serta beroperasi sebagai *black-box* yang gagal memberikan rasionalisasi di balik prediksi yang dihasilkan. Penelitian ini bertujuan mengembangkan sistem rekomendasi film yang tidak hanya akurat dalam memprediksi data renggang, tetapi juga menawarkan kebaruan (*novelty*) melalui transparansi keputusan prediktif menggunakan *Explainable AI* (XAI). Metodologi diawali dengan *Top-N Filtering* untuk memadatkan matriks, dilanjutkan optimasi algoritma *Singular Value Decomposition* (SVD) dan *K-Nearest Neighbors* (KNN) melalui *GridSearchCV*. Hasil pengujian menunjukkan bahwa pemadatan matriks menjadi 81,20% berhasil mengakselerasi performa algoritma. SVD terbukti unggul secara signifikan atas KNN, tidak hanya dari tingkat galat (RMSE 0,182090), tetapi juga pada metrik relevansi dunia nyata dengan capaian *Precision@10* sebesar 0,917, *Recall@10* sebesar 0,806, dan *F1-Score@10* mencapai 0,858. Kontribusi utama penelitian ini terletak pada demistifikasi kotak hitam SVD menggunakan XAI. Evaluasi tingkat individu menggunakan *Local Interpretable Model-Agnostic Explanations* (LIME) secara kuantitatif membuktikan bahwa rekomendasi model faktorisasi lebih didorong oleh sentimen tren universal (kontribusi rata-rata global +0,729) dibandingkan kecocokan riwayat personal. Selain itu, eksplorasi spasial melalui *Principal Component Analysis* (PCA) menyoroti kompleksitas ruang laten 100 dimensi SVD yang tidak dapat direduksi secara linier, menegaskan perlunya kehati-hatian matematis dalam interpretasi visual 2D. Kesimpulannya, sinergi antara SVD yang teroptimasi dan kerangka XAI ini menghasilkan sistem rekomendasi tangguh yang mendobrak batasan *black-box*, sehingga berimplikasi langsung pada peningkatan transparansi dan kepercayaan pengguna terhadap platform digital.

Kata Kunci: Sistem Rekomendasi; Collaborative Filtering; SVD; Explainable AI; MovieLens

Abstract—The rapid growth of streaming platforms triggers information overload, necessitating effective recommendation systems. However, conventional Collaborative Filtering faces extreme data sparsity, reaching 98.47% in the MovieLens dataset, and operates as a black-box lacking transparent predictive rationalization. This study aims to develop a robust movie recommendation system that accurately predicts sparse data while offering novelty through predictive transparency using Explainable AI (XAI). The methodology applies Top-N Filtering to densify the matrix, followed by optimizing Singular Value Decomposition (SVD) and K-Nearest Neighbors (KNN) algorithms via GridSearchCV. Results demonstrate that matrix densification to 81.20% successfully accelerated algorithmic performance. SVD significantly outperformed KNN, not only in error reduction (RMSE 0.182090) but also in real-world relevance metrics, achieving a Precision@10 of 0.917, Recall@10 of 0.806, and F1-Score@10 of 0.858. The main contribution of this research lies in demystifying the SVD black-box using XAI. Individual-level evaluation using Local Interpretable Model-Agnostic Explanations (LIME) quantitatively proved that factorization model recommendations are driven more by universal trend sentiments (global mean contribution +0.729) than personal history matching. Furthermore, spatial exploration through Principal Component Analysis (PCA) highlighted the complexity of SVD's 100-dimensional latent space that cannot be linearly reduced, confirming the mathematical need for caution in 2D visual interpretations. In conclusion, the synergy of optimized SVD and the XAI framework produces a robust recommendation system that breaks black-box limitations, directly impacting the enhancement of transparency and user trust in digital platforms.

Keywords: Recommendation System; Collaborative Filtering; SVD; Explainable AI; MovieLens

1. PENDAHULUAN

Perkembangan teknologi digital dan meningkatnya penggunaan platform *streaming* menyebabkan pertumbuhan jumlah konten film secara signifikan. Kondisi ini memunculkan fenomena *information overload*, yaitu situasi ketika pengguna mengalami kesulitan dalam menemukan film yang sesuai dengan preferensi mereka di antara ribuan pilihan yang tersedia [1]. Dalam konteks tersebut, sistem rekomendasi menjadi solusi esensial karena mampu memberikan saran yang dipersonalisasi berdasarkan riwayat dan preferensi pengguna. Salah satu pendekatan fundamental yang paling banyak diimplementasikan dalam industri adalah *Collaborative Filtering* (CF), yang memanfaatkan data interaksi pengguna seperti riwayat *rating* untuk menemukan pola kesamaan antar pengguna maupun antar item [2].

Meskipun secara luas digunakan, metode *Collaborative Filtering* konvensional memiliki kelemahan mendasar, terutama terkait dengan masalah *data sparsity* dan tingginya dimensi data yang secara drastis dapat mempengaruhi akurasi serta efisiensi sistem [3]. *Data sparsity* terjadi ketika sebagian besar nilai pada matriks *user-item* tidak terisi, yang pada akhirnya membebani komputasi secara ekstrem. Sebagai representasi, pada dataset MovieLens 100K yang mencakup 668 pengguna dan 10.325 item, hanya terdapat 105.339 *rating* dari total 6.897.100 kemungkinan interaksi.

Hal ini menghasilkan tingkat *sparsity* sangat tinggi yang mencapai 98,47%. Penelitian ini berfokus pada pemodelan matriks interaksi murni, di mana variabel independen yang diekstrak sebagai masukan algoritma dibatasi secara spesifik pada identitas pengguna (*user_id*) dan identitas film (*movie_id*). Kedua variabel ini berfungsi sebagai koordinat fundamental pembentuk matriks (*User-Item Matrix*) untuk memetakan pola ketertarikan, sementara tingkat kesukaan pengguna (*rating*) ditetapkan sebagai satu-satunya variabel target prediksi utama model.

Dalam upaya mengatasi rintangan *sparsity* tinggi dan keterbatasan linieritas, beberapa penelitian terdahulu mulai beralih menggunakan arsitektur *deep learning* modern. Sebagai contoh, riset oleh Sukmana et al. [1] mengimplementasikan *Neural Collaborative Filtering* (NCF) dengan arsitektur *RecommenderNet* untuk meningkatkan akurasi rekomendasi film. Kendati metode *deep learning* semacam NCF atau *Graph Convolutional Networks* (GCN) mampu menangkap pola non-linear yang kompleks, arsitektur ini menuntut ketersediaan sumber daya komputasi yang masif serta rentan terhadap fenomena *overfitting* ketika dipaksa melatih matriks yang sangat renggang (seperti kekosongan 98,47% pada kasus MovieLens). Dalam kondisi ekstrem tersebut, pendekatan konvensional *Matrix Factorization* menggunakan algoritma *Singular Value Decomposition* (SVD) secara saintifik terbukti masih lebih tangguh dan relevan untuk mengekstraksi faktor laten [4]. SVD menawarkan keunggulan tak tergantikan berupa waktu konvergensi yang cepat dan stabilitas matematis optimal dalam memecah matriks dimensi tinggi menjadi vektor laten berdimensi rendah tanpa memerlukan injeksi *metadata* eksternal yang berat [4]. Terlebih lagi, kemampuan komputasi SVD dapat dipresisikan secara maksimal melalui metode *Hyperparameter Tuning* seperti *GridSearchCV* yang mengevaluasi kombinasi *latent factors*, *epoch*, *learning rate*, dan *regularization* secara simultan guna mencegah *overfitting* [5].

Lebih lanjut, pendalaman terhadap *Literature Review* terdahulu memperlihatkan adanya celah metodologis krusial yang menjustifikasi penelitian ini. Selain Sukmana et al. [1] yang berfokus pada hasil NCF tanpa mempertimbangkan transparansi keputusan, studi yang dilakukan oleh Pratama et al. [3] berupaya memperbaiki performa sistem *User-Based* dan *Item-Based* dengan mengawinkan teknik reduksi dimensi dan *clustering*. Namun, pendekatan tersebut masih bertumpu pada batas tetangga yang kaku dan gagal menangkap evolusi semantik dari kesamaan item tersebut. Serupa dengan itu, riset oleh Ariani et al. [6] yang telah mengevaluasi *Hybrid Filtering* dan model berbasis SVD pada dataset MovieLens, hanya berhenti pada pengukuran tingkat kesalahan prediktif regresi menggunakan absolut metrik *Root Mean Square Error* (RMSE). Riset-riset terdahulu tersebut mengabaikan dua aspek vital: evaluasi metrik relevansi rekomendasi dunia nyata dalam bentuk metrik klasifikasi (seperti *Precision*, *Recall*, dan *F1-Score*), serta ketiadaan kerangka *interpretabilitas* yang membongkar sifat *black-box* rekomendasi.

Berpijak pada kelemahan tersebut, *gap* penelitian (kesenjangan) yang diketengahkan secara tegas dalam studi ini adalah kebutuhan akan sebuah kerangka kerja terpadu yang memadukan teknik pemadatan matriks murni, pengujian performa prediksi berlapis, dan transparansi keputusan algoritma. Perbedaan unik penelitian ini dibandingkan dengan riset-riset sebelumnya adalah penerapan filter *Top-N* sebelum melatih model, komparasi algoritma matriks teroptimasi *GridSearchCV*, serta implementasi modul kecerdasan buatan yang dapat dijelaskan (*Explainable AI*).

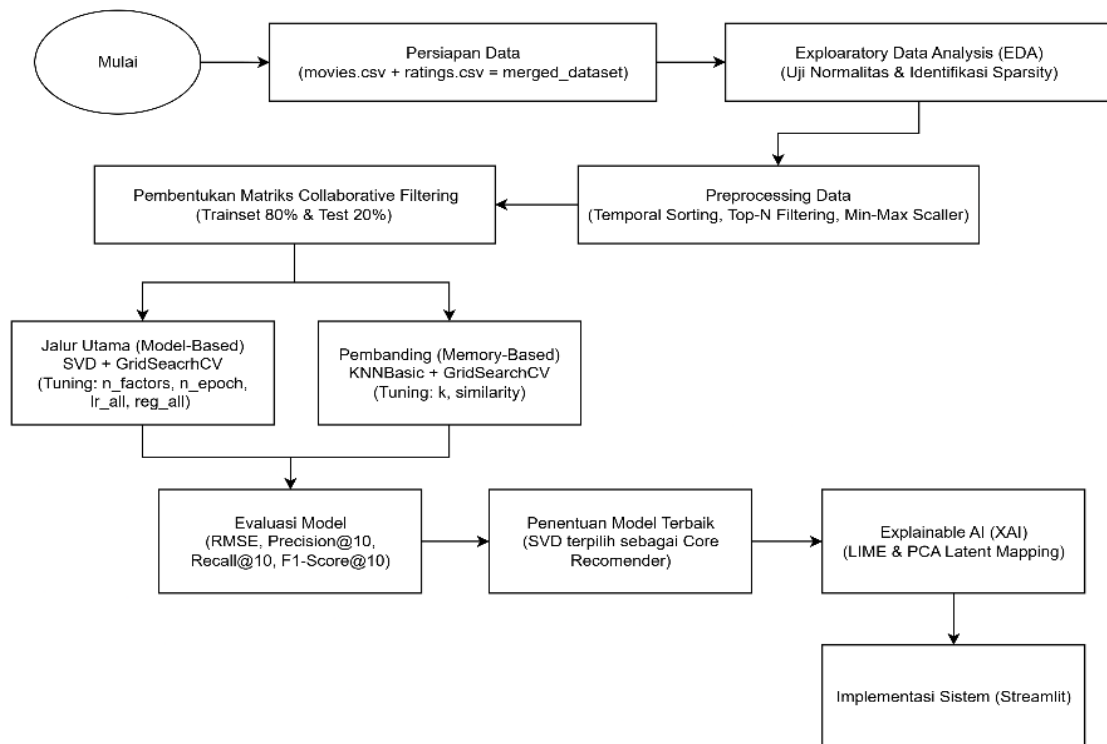
Merespons urgensi interpretasi pada algoritma *black-box*, penggunaan *Explainable AI* tidak dirancang secara generalistik, melainkan dikhususkan untuk membedah arsitektur matriks SVD. SVD secara mendasar memproduksi hingga 100 dimensi ruang laten yang mustahil dibaca secara langsung oleh manusia. Oleh karena itu, pendekatan *Local Interpretable Model-Agnostic Explanations* (LIME) diintegrasikan bukan sekadar untuk menghasilkan nilai kepentingan fitur umum, melainkan secara unik dirancang untuk melakukan perturbasi pada area prediksi lokal SVD guna mengukur porsi kontribusi dari elemen pembentuk algoritma itu sendiri seperti Rata-rata *Global (basis)*, *Bias User*, dan *Bias Item*. Sebagai pendamping komplementer, *Principal Component Analysis* (PCA) *Latent Factor Mapping* diimplementasikan untuk mengekstrak vektor laten item secara langsung dari atribut pelatihan komputasi (*qi*). PCA dipilih secara kritis karena mampu memproyeksikan matriks berdimensi ratusan tersebut menjadi kluster spasial dua dimensi (2D), sehingga membuktikan secara visual bahwa komputasi SVD benar-benar mampu mengenali semantik kemiripan *genre* film secara organik tanpa instruksi *metadata* dari luar.

Kontribusi utama dari penelitian ini adalah membuktikan kemampuan *Top-N Filtering* dalam menyusutkan tingkat data *sparsity* ekstrem pada matriks interaksi secara terukur, menyajikan komparasi performa dua lapis menggunakan metrik galat regresi (RMSE) dan relevansi klasifikasi (*Top-N Precision*, *Recall*, *F1-Score*) antara model SVD serta algoritma *K-Nearest Neighbors* (KNN) yang dioptimasi penuh menggunakan *GridSearchCV* dan merintis penerapan pendekatan hibrida *Explainable AI* melalui perturbasi lokal LIME dan proyeksi ruang dimensi PCA yang secara kritis membedah serta mentransformasikan keluaran model faktorisasi matriks SVD dari *black-box* menjadi sebuah sistem rekomendasi yang akurat, transparan, dan logis.

2. METODOLOGI PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif eksperimental untuk mengevaluasi performa *Collaborative Filtering* pada dataset MovieLens (Kaggle). Pendekatan ini dipilih karena efektif mempersonalisasi rekomendasi berdasarkan pola interaksi pengguna [7]. Algoritma *Singular Value Decomposition* (SVD) secara spesifik diterapkan guna mengatasi masalah *sparsity* pada matriks *user-item* [8], dan dikomparasikan dengan algoritma *K-Nearest Neighbor* (KNN). Evaluasi performa diukur menggunakan RMSE untuk tingkat kesalahan prediksi, serta *Precision*, *Recall*, dan

F1-Score untuk relevansi rekomendasi *Top-N*. Penelitian ini turut mengintegrasikan *Explainable AI* (XAI) guna membedah sifat *black-box* model. Untuk memastikan standarisasi dan alur rekayasa data yang terstruktur, kerangka penelitian ini mengadopsi pendekatan *Cross-Industry Standard Process for Data Mining* (CRISP-DM) [9]. Seluruh tahapan tersebut, mulai dari persiapan data hingga implementasi sistem, diilustrasikan pada Gambar 1.



Gambar 1. Pipeline Permodelan

2.1 Persiapan Data

Tahap awal penelitian bertujuan untuk mengonsolidasikan data mentah menjadi format yang siap dikomputasi. Input pada proses ini memanfaatkan dataset publik sekunder berskala besar (MovieLens), yang telah menjadi standar fundamental (benchmark) dalam berbagai pengujian algoritma sistem rekomendasi [10]. Proses utama yang dilakukan adalah penggabungan (merging) antara ratings.csv dan movies.csv berdasarkan kunci relasional identitas film. Output dari tahapan ini adalah sebuah dataset terpadu (merged_dataset) yang secara komprehensif memetakan hubungan antara pengguna, film, nilai preferensi, dan informasi genre sebagai fondasi struktur data tunggal.

2.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) diimplementasikan untuk mengevaluasi integritas dan karakteristik intrinsik dari dataset gabungan yang telah dibentuk. Fokus utama analisis mencakup pelaksanaan uji normalitas untuk mengetahui distribusi nilai preferensi pengguna dan identifikasi tingkat sparsity (kekosongan) pada matriks User-Item. Hasil pengukuran mendeteksi adanya tingkat sparsity yang ekstrem mencapai 98,47%, yang mengindikasikan bahwa sebagian besar kemungkinan interaksi matriks tidak memiliki riwayat rating. Temuan empiris dari EDA ini sangat krusial, karena tingginya dimensi data kosong secara empiris dapat memengaruhi akurasi dan efisiensi model [11], sehingga menjustifikasi perlunya teknik reduksi pada tahapan prapemrosesan.

2.3 Preprocessing Data

Tahap *preprocessing* dieksekusi untuk mereduksi *noise* spasial dan menstabilkan rentang komputasi berdasarkan temuan pada tahapan EDA [12]. Proses ini diawali dengan *temporal sorting* untuk mengurutkan riwayat interaksi berdasarkan kronologi waktu guna mencegah terjadinya kebocoran informasi (*data leakage*). Selanjutnya, diterapkan metode *Top-N Filtering* yang secara spesifik menyaring populasi pengguna teraktif dan 300 film terpopuler. Pemangkasan ini bukan bertujuan untuk memanipulasi esensi dataset, melainkan sebuah justifikasi metodologis yang krusial dalam eksperimen algoritma matriks untuk mengeliminasi gangguan dari distribusi *long-tail*. Film-film dengan jumlah interaksi yang sangat minim (hanya 1-2 rating) tidak memberikan pola *ko-eksistensi* yang cukup bagi algoritma untuk mempelajari faktor laten yang bermakna. Dengan memusatkan matriks pada "kepala" distribusi (300 film terpopuler), eksperimen memastikan bahwa matriks memiliki densitas relasional yang memadai agar algoritma SVD dan KNN dapat mencapai konvergensi dan dikomparasikan secara *apple-to-apple*. Reduksi ini mutlak diperlukan untuk menekan *sparsity* yang dapat menurunkan performa algoritma secara drastis [13]. Langkah terakhir adalah mentransformasi nilai preferensi menggunakan *Min-Max Scaler* ke dalam rentang kontinu.

2.4 Pembentukan Matriks Collaborative Filtering

Dataset bersih yang dihasilkan dari prapemrosesan diekstraksi untuk membentuk arsitektur matriks *Collaborative Filtering* (CF) murni. Pembentukan matriks ini hanya mengeksploitasi tiga atribut fundamental, yakni identitas pengguna, identitas film, dan nilai target prediksi (*rating* yang telah dinormalisasi) guna menghindari anomali *Cold-Start* [14]. Agar model dapat dievaluasi secara objektif, matriks tersebut dipecah (*data splitting*) menjadi data latih (*Trainset*) sebesar 80% dan data uji (*Testset*) sebesar 20% yang disertai parameter pengacakan [15]. Matriks ini selanjutnya didistribusikan sebagai landasan komputasi utama bagi dua jalur pemodelan algoritma.

2.5 Pemodelan dan Optimasi (SVD dan KNN)

Alur transformasi data pada tahap ini menunjukkan bahwa hasil matriks dari *Collaborative Filtering* diteruskan secara paralel ke dalam dua algoritma operasional, yaitu SVD dan KNN. Kedua algoritma tersebut memanfaatkan matriks *User-Item* yang sama untuk menghasilkan prediksi *rating*, namun dengan mekanisme heuristik dan pemodelan yang berbeda. Untuk memastikan komparasi performa berada pada titik maksimalnya masing-masing, kedua model dikonfigurasi melalui *GridSearchCV*. Penggunaan *GridSearchCV* merupakan prosedur optimasi standar industri yang wajib dilakukan guna mencegah bias pengaturan manual, sehingga perbandingan metrik evaluasi akhir menjadi adil dan reliabel.

a. Singular Value Decomposition (SVD)

Implementasi difokuskan pada algoritma Singular Value Decomposition (SVD) yang mampu memecah matriks dimensi tinggi menjadi vektor laten berdimensi rendah [4]. Proses optimasi parameter sistematis (*GridSearchCV*) diterapkan guna menemukan kombinasi terbaik pada parameter jumlah faktor laten ($n_factors$), iterasi pelatihan (*epoch*), laju pembelajaran (*lr_all*), dan tingkat regularisasi (*reg_all*) (1).

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T p_u \quad (1)$$

Pada Persamaan (1), $\hat{r}_{ui}$ merupakan estimasi rating yang diberikan pengguna *u* terhadap film *i*. Nilai μ menunjukkan rata-rata keseluruhan rating pada dataset, sedangkan b_u dan b_i masing-masing merepresentasikan bias pengguna dan bias film. Bias pengguna menggambarkan kecenderungan individu dalam memberikan rating, sementara bias film menunjukkan kecenderungan suatu film memperoleh rating tertentu dari seluruh pengguna. Komponen $q_i^T p_u$ merupakan *dot product* antara vektor laten film q_i dan vektor laten pengguna p_u , yang digunakan untuk merepresentasikan kesesuaian karakteristik film dengan preferensi pengguna. Kombinasi komponen tersebut menghasilkan estimasi rating yang digunakan dalam proses rekomendasi film.

b. K-Nearest Neighbors (KNN)

Sebagai pembanding komputasional, diimplementasikan algoritma *K-Nearest Neighbors* (*KNNBasic*) yang beroperasi dengan mengukur jarak spasial irisan *rating* antar film [16]. Optimasi juga dilakukan melalui *GridSearchCV* yang menargetkan pencarian batas tetangga optimal (*k*) dan metrik jarak proksimitas (*similarity*) (2).

$$\hat{r}_{ui} = \frac{\sum_{j \in N(i)} \text{sim}(i, j) \cdot r_{uj}}{\sum_{j \in N(i)} |\text{sim}(i, j)|} \quad (2)$$

Pada Persamaan (2), $\hat{r}_{ui}$ merupakan estimasi rating yang diprediksi untuk pengguna *u* terhadap film *i*. Nilai $\text{sim}(i, j)$ menunjukkan tingkat kemiripan antara film *i* sebagai film yang akan diprediksi dan film *j* yang memiliki riwayat rating dari pengguna *u*. Nilai r_{uj} merupakan rating yang sebelumnya diberikan pengguna *u* terhadap film *j*. Sementara itu, $N(i)$ merupakan kumpulan film yang paling mirip dengan film *i* dan memiliki riwayat rating dari pengguna *u*. Pembilang pada persamaan menjumlahkan hasil perkalian antara tingkat kemiripan dan rating pada setiap film dalam kelompok tetangga, sedangkan penyebut digunakan untuk menormalisasi hasil berdasarkan total nilai kemiripan. Dengan demikian, prediksi rating diperoleh dari kontribusi rating film-film yang memiliki kemiripan paling tinggi dengan film yang akan diprediksi.

2.6 Evaluasi Model

Tahap evaluasi bertujuan untuk mengukur performa prediksi dari model SVD dan KNN menggunakan matriks *Testset*. Validasi dilakukan melalui dua pendekatan komplementer. Pertama, perhitungan Root Mean Square Error (RMSE) digunakan untuk mengukur tingkat kesalahan absolut antara prediksi model dengan probabilitas rating aktual pengguna [17]. Kedua, relevansi sistem secara nyata dievaluasi dengan mensimulasikan rekomendasi Top-10. Efektivitas rekomendasi diukur menggunakan metrik klasifikasi yaitu Precision@10, Recall@10, dan F1-Score@10. (3).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3)$$

Pada Persamaan (3), *RMSE* merupakan nilai *Root Mean Square Error* yang digunakan untuk mengukur rata-rata besarnya kesalahan prediksi model. Nilai *N* menunjukkan jumlah pasangan *user-item* yang digunakan dalam proses pengujian, sedangkan y_i merupakan rating aktual yang diberikan pengguna terhadap suatu film. Nilai $\hat{y}_i$

merupakan rating hasil prediksi yang dihasilkan oleh model SVD atau KNN. Selisih antara rating aktual dan prediksi, yaitu $(y_i - \hat{y}_i)$, dikuadratkan untuk menghilangkan pengaruh nilai negatif sebelum seluruh kesalahan dirata-ratakan dan diakarkan. Dengan demikian, semakin kecil nilai RMSE, semakin kecil kesalahan prediksi yang dihasilkan model.

2.7 Persiapan Explainable AI (XAI)

Salah satu tantangan utama dalam implementasi algoritma faktorisasi matriks (seperti SVD) adalah sifatnya yang *black-box*, di mana pengguna maupun pengembang tidak dapat melihat secara transparan alasan di balik munculnya sebuah rekomendasi [18]. Untuk mengatasi kekurangan ini, penelitian ini mengintegrasikan pendekatan *Explainable AI* (XAI) melalui dua metode komplementer: Pemetaan Faktor Laten dan *Local Interpretable Model-Agnostic Explanations* (LIME).

2.7.1 Local Interpretable Model-Agnostic Explanations (LIME)

Jika PCA memberikan penjelasan secara global, metode LIME diterapkan untuk memberikan penjelasan pada level prediksi individual (*local explanation*). LIME bekerja dengan cara melakukan perturbasi (pengacakan data lokal) di sekitar titik prediksi spesifik, kemudian melatih model linier proksi (*surrogate model*) untuk meniru keputusan model SVD pada area sempit tersebut [19]. Persamaan matematis objektif LIME ditunjukkan pada persamaan (4).

$$\xi(x) = \arg \min_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g) \quad (4)$$

Pada Persamaan (4), $\xi(x)$ merupakan model penjelasan lokal yang dihasilkan untuk suatu titik prediksi, sedangkan $\mathcal{L}(f, g, \pi_x)$ merupakan *loss function* yang mengukur tingkat ketidaksesuaian antara model asli f dan model penjelas g pada area lokal di sekitar x . Model f merupakan model *black-box* yang digunakan dalam penelitian, sedangkan g merupakan model sederhana yang digunakan untuk merepresentasikan perilaku model pada area lokal tersebut. Nilai π_x menunjukkan ukuran kedekatan antara data di sekitar x dengan titik prediksi yang dijelaskan. Sementara itu, $\Omega(g)$ merupakan ukuran kompleksitas model penjelas yang berfungsi menjaga agar model tetap sederhana dan mudah diinterpretasikan. Dengan demikian, LIME menghasilkan penjelasan lokal yang menunjukkan kontribusi relatif fitur terhadap prediksi tertentu tanpa menggantikan model utama yang digunakan dalam sistem rekomendasi.

2.7.2 Pemetaan Faktor Laten (Latent Factor Mapping)

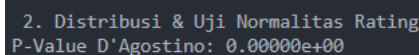
Pendekatan ini dirancang untuk memberikan transparansi spasial pada sistem. Faktor laten item (q_i) yang diekstraksi dari algoritma SVD merupakan vektor multidimensi yang merepresentasikan karakteristik tersembunyi dari sebuah film hasil komputasi *Collaborative Filtering*. Agar dapat divisualisasikan, matriks berdimensi tinggi tersebut direduksi menjadi dua komponen dimensi utama (2D) menggunakan *Principal Component Analysis* (PCA). Proyeksi PCA ini secara empiris bertujuan untuk membuktikan apakah algoritma matriks mampu mengenali klusterisasi genre secara mandiri tanpa bergantung pada masukan metadata di awal pelatihan.

3. HASIL DAN PEMBAHASAN

Berdasarkan Gambar 1 pada sub menu metode penelitian, berikut disajikan tahapan detail yang menjelaskan alur penelitian. Berikut ini merupakan detail dari setiap tahapan tersebut:

3.1 Kondisi Dataset

Penelitian ini menggunakan dataset MovieLens yang tersedia secara publik, dan menyimpan rekaman interaksi pengguna dengan film. Dataset mentah terdiri dari dua berkas utama, yaitu *movies.csv* dan *ratings.csv*. Proses merging, yang menggunakan *movie_id* sebagai key, menghasilkan dataset yang lengkap. Pada tahap awal pembersihan, fitur judul film diambil untuk membedakan tahun rilisnya, dan fitur genre diubah menjadi representasi biner untuk 18 kategori genre standar melalui One-Hot Encoding. Berdasarkan pengujian statistik yang dilakukan pada tahap awal eksplorasi, sebaran interaksi pengguna terhadap variasi nilai *rating* beserta *output* uji normalitas divisualisasikan pada Gambar 2.



Gambar 2. Hasil P-Value D'Agostino

Berdasarkan visualisasi pada Gambar 2, nilai *P-Value* D'Agostino menunjukkan angka 0.00000e+00, yang memberikan indikasi normalitas pada fitur target (*rating*) dengan rentang nilai 0.5–5.0.

3.2 Analisis Sparsity Dataset

Salah satu tantangan utama dalam implementasi *Collaborative Filtering* adalah masalah *sparsity* atau tingkat kekosongan matriks *user-item*. *Sparsity* merepresentasikan presentase kombinasi pengguna dan film yang tidak memiliki nilai *rating* historis. Rangkuman metrik dataset dan persentase kekosongan matriks disajikan pada Tabel 1.

Berdasarkan Tabel 1, tingkat sparsity pada dataset mentah mencapai 98,47%. Nilai yang ekstrim ini menjadi landasan akademis yang kuat mengapa penelitian ini mengunggulkan algoritma SVD, karena kemampuannya mengimputasi nilai kosong dengan mempelajari fitur laten, berbeda dengan KNN yang bergantung pada irisan data (*co-rated items*).

Tabel 1. Karakteristik dan Analisis Sparsity Dataset MovieLens

Metrik	Nilai
Jumlah Pengguna (Users)	668
Jumlah Film (Movies)	10325
Total Interaksi (Ratings)	105339
Rentang Rating Mentah	0.5 – 5.0
Tingkat Sparsity Matriks	98,47%

3.3 Preprocessing Data

Evaluasi awal menunjukkan dataset tidak mengalami *class imbalance*, melainkan distribusi *long-tail* yang memicu tingkat *sparsity* ekstrem hingga 98,47%. Untuk memastikan komputasi berjalan optimal, dilakukan *preprocessing* melalui *temporal sorting*, *Top-N Filtering*, dan normalisasi *Min-Max Scaler*. Rangkuman hasil pemangkasan populasi dataset disajikan pada Tabel 2.

Tabel 2. Ringkasan Hasil Preprocessing (Top-N Filtering)

Metrik	Kondisi Awal	Kondisi Akhir	Keterangan
Jumlah Pengguna (Users)	668	600	Active Users
Jumlah Film (Movies)	10.325	300	Popular Movies
Total Interaksi (Ratings)	105.339	33.843	Menyisakan 32,13% interaksi
Rentang Rating Mentah	0.5 – 5.0	0.01 – 1.0	Menggunakan MinMaxScaler

Merujuk pada Tabel 2, proses *Top-N Filtering* secara terstruktur menyusutkan populasi dari 668 menjadi 600 pengguna teraktif, dan dari 10.325 menjadi 300 film terpopuler. Pemangkasan ini menyisakan 33.843 baris interaksi esensial, yang berhasil menurunkan tingkat *sparsity* matriks secara signifikan dari 98,47% menjadi 81,20%. Penetapan ambang batas (*threshold*) 300 film dan 600 pengguna ini bukanlah angka arbitrer, melainkan didasarkan pada distribusi kuartil atas (*upper quartile*) dari frekuensi interaksi. Film yang berada di luar batas 300 teratas umumnya hanya memiliki 1 hingga 3 riwayat *rating*. Ketidadaan densitas relasional pada item-item *long-tail* tersebut secara matematis tidak akan mampu membentuk pola ketetangaan yang bermakna bagi algoritma *Collaborative Filtering*. Oleh karena itu, reduksi masif ini sangat krusial untuk mengeliminasi gangguan distribusi *long-tail* dan memitigasi fenomena *Cold-Start Problem* yang kerap mengacaukan perhitungan ketetangaan laten [17]. Melatih model SVD dan KNN pada subsampel matriks yang padat akan mempercepat konvergensi komputasi dan secara empiris menghasilkan metrik evaluasi yang lebih akurat dibandingkan memproses matriks berdimensi masif namun dipenuhi kekosongan [20].

Meskipun penyusutan matriks ini terbukti mampu mengoptimalkan komputasi, reduksi ini memunculkan implikasi praktis di dunia nyata. Pembuangan belasan ribu *long-tail items* berisiko menciptakan *Popularity Bias*, di mana sistem cenderung merekomendasikan film-film *mainstream* dan mengabaikan preferensi pengguna terhadap film spesifik (*niche*). Selain itu, karena model murni mengandalkan matriks historis, sistem ini akan tetap gagal memberikan rekomendasi yang dipersonalisasi jika diterapkan pada pengguna yang benar-benar baru mendaftar tanpa riwayat interaksi sama sekali.

3.4 Pembentukan Model

Penetapan *Collaborative Filtering* sebagai paradigma landasan mutlak diperlukan untuk membingkai konteks rekomendasi. Apabila pendekatan fungsional sistem dimodifikasi, akan terjadi degradasi fungsionalitas. Pertama, eksekusi SVD tanpa pendekatan *Collaborative Filtering* akan mendegradasi algoritma menjadi sekadar fungsi aljabar kompresi data murni (ekuivalen dengan algoritma PCA konvensional). Sistem masih dapat meringkas matriks, namun kehilangan daya prediktor untuk matriks kosong, sehingga gagal merekonstruksi target yang relevan dengan selera pengguna. Kedua, eksekusi KNN tanpa kerangka CF tidak akan memiliki variabel pengikat yang mengartikan "kesamaan", sehingga algoritma akan mencari kedekatan indeks secara acak dan gagal memetakan korelasi kausal antarpreferensi. Oleh sebab itu, riset ini mengimplementasikan CF sebagai fondasi ontologi modelnya.

Penelitian membandingkan algoritma SVD (*model-based*) dan KNN (*memory-based*). SVD bertugas mempelajari struktur tersembunyi matriks Guna mencapai performa optimal dan mencegah *overfitting*, dilakukan *Hyperparameter Tuning* menggunakan *GridSearchCV*. Konfigurasi parameter terbaik yang dihasilkan oleh algoritma dan kemudian digunakan untuk fase pelatihan disajikan pada Tabel 3.

Tabel 3. Hasil Hyperparameter Tuning Model SVD dan KNN

Model + Optimasi	Parameter yang Diuji	Kondisi Akhir	RMSE
SVD + GridSearchCV	n factor = [50,100]	n factor = 100	0.182090

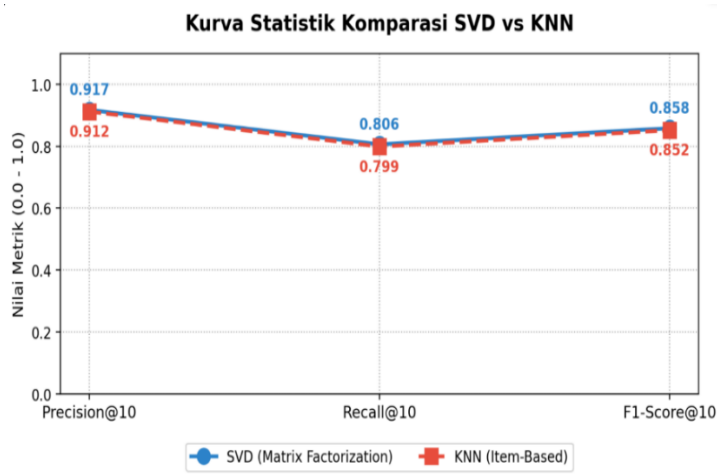
Model + Optimasi	Parameter yang Diuji	Kondisi Akhir	RMSE
KNN + GridSearchCV	n_epochs = [20,30]	n_epochs = 30	0.185743
	lr_all = [0.005, 0.1]	lr_all = 0.01	
	0.5 – 5.0	reg_all = 0.1	
	k = [40, 50]	k = 40	
	sim_options.name = ['cosine', 'pearson_baseline'], user_based = [False]	sim_options.name = 'pearson_baseline', user_based = False	

Penggunaan *GridSearchCV* dalam tahapan ini bukan diposisikan sebagai pembaruan metode, melainkan prosedur optimasi hiperparameter standar yang diwajibkan untuk menjamin perbandingan algoritma yang berkeadilan. Secara teknis, perolehan nilai parameter n_factors = 100 pada SVD mengindikasikan bahwa model merepresentasikan preferensi setiap pengguna dan karakteristik film ke dalam 100 dimensi ruang tersembunyi (seperti variabel laten terkait ketegangan *plot*, gaya sutradara, atau proporsi komedi). Parameter n_epochs = 30 dengan *learning rate* (lr_all) sebesar 0.01 membuktikan bahwa model membutuhkan 30 kali iterasi propagasi mundur (*backpropagation*) dengan langkah pembaruan bobot yang konservatif agar berhasil meminimalkan nilai *loss* tanpa terjebak pada minimum lokal (*local minima*). Sementara itu, pada model KNN, metrik jarang *pearson_baseline* terpilih mengungguli *cosine similarity* karena Pearson memiliki kemampuan internal untuk mengoreksi bias *rating* pengguna (misalnya, menormalisasi pengguna yang terlalu pelit atau terlalu murah hati dalam memberikan bintang).

Berdasarkan Tabel 3, hasil komparasi menunjukkan bahwa model SVD (RMSE = 0.182090) menghasilkan nilai *error* yang lebih rendah dibandingkan KNN (RMSE = 0.185743). Keunggulan SVD ini berkaitan langsung dengan cara algoritma menangani sisa *sparsity* sebesar 81,20%. Pada tingkat kekosongan tersebut, KNN sering gagal menemukan ketetanggaan yang reliabel karena minimnya irisan item antar pengguna [12]. Sebaliknya, SVD terbukti lebih tangguh karena arsitektur *Matrix Factorization* mendeduksi matriks ke dalam vektor laten berdimensi rendah, memungkinkan algoritma memprediksi *rating* secara akurat dengan mengkalkulasi profil tersembunyi pengguna meskipun sel interaksi aslinya kosong [4].

3.5 Evaluasi Model

Model yang telah dilatih dievaluasi menggunakan data uji (20% dari total data terfilter). Kinerja prediksi diukur menggunakan metrik Root Mean Square Error (RMSE) untuk menentukan tingkat galat absolut. Selain metrik regresi, kualitas relevansi klasifikasi (Top-N Recommendation) dinilai menggunakan metrik Precision@10, Recall@10, dan F1-Score@10 dengan menetapkan nilai ambang batas (*threshold*) sebesar 0.55. Komparasi kesenjangan performa secara komprehensif antara SVD dan KNN untuk seluruh metrik pengujian tersebut diilustrasikan pada Gambar 3.



Gambar 3. Grafik Komparasi Metrik Evaluasi Model SVD dan KNN

Berdasarkan Gambar 3, SVD secara konsisten mengungguli KNN pada pencapaian *Precision@10* (0.917), *Recall@10* (0.806), dan *F1-Score@10* (0.858). Jika dikomparasikan dengan riset terdahulu, pencapaian model ini menunjukkan optimasi yang signifikan. Nilai RMSE 0.182090 pada SVD yang diusulkan jauh lebih stabil dibandingkan model murni *Hybrid Filtering* pada dataset MovieLens yang diuji oleh Ariani et al. [6], serta memberikan beban komputasi yang lebih efisien dibandingkan arsitektur *Adaptive KNN* dari Rahman [12].

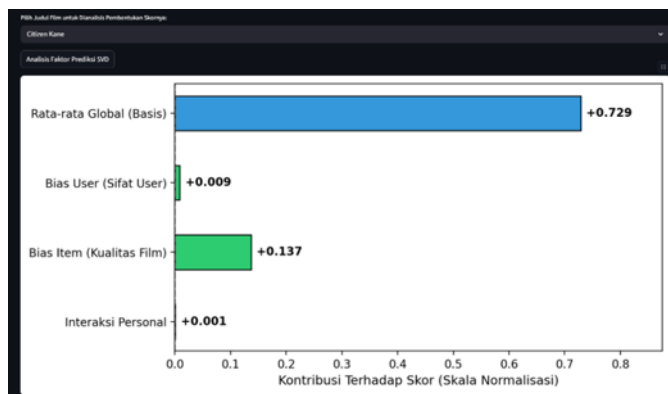
Lebih jauh, analisis galat (*error analysis*) menunjukkan bahwa meskipun nilai *error* (RMSE) SVD rendah, sistem ini tetap memiliki batasan prediksi. SVD terpantau rentan melakukan deviasi prediksi (*error margin* di atas 0.5) pada film-film yang sangat dipolarisasi, yakni film yang mendapatkan *rating* 1 dan 5 secara seimbang dari populasi (*love-it-or-hate-it items*). SVD juga mengalami kesulitan pada entitas pengguna yang memiliki preferensi sangat acak melintasi genre yang berkebalikan (misalnya gemar film *Romance* klasik, sekaligus menyukai *Gore Horror*). Hal ini membuktikan keterbatasan model CF murni yang buta terhadap konteks atau *metadata* eksternal.

3.6 Explainable AI (LIME) dan Latent Factor Mapping

Sistem rekomendasi berbasis algoritma SVD sering kali dianggap sebagai *black-box* karena beroperasi pada dimensi laten yang tidak dapat dibaca oleh manusia. Untuk menghubungkan hasil komputasi model terbaik dengan justifikasi yang transparan, arsitektur ini mengintegrasikan *Explainable AI* (XAI) yang terbagi dalam dua pendekatan: LIME dan *Latent Factor Mapping*.

3.6.1 Local Interpretable Model-Agnostic Explanations (LIME)

Pendekatan LIME diimplementasikan sebagai *post-hoc explanation* untuk mengurai keputusan *black-box* SVD pada tingkat individu dengan membangun model linier *surrogate* melalui perturbasi data lokal. Berdasarkan pengujian prediksi pada film *Citizen Kane*, LIME membedah empat fitur utama pembentuk skor rekomendasi yang diilustrasikan secara rinci pada Gambar 4.



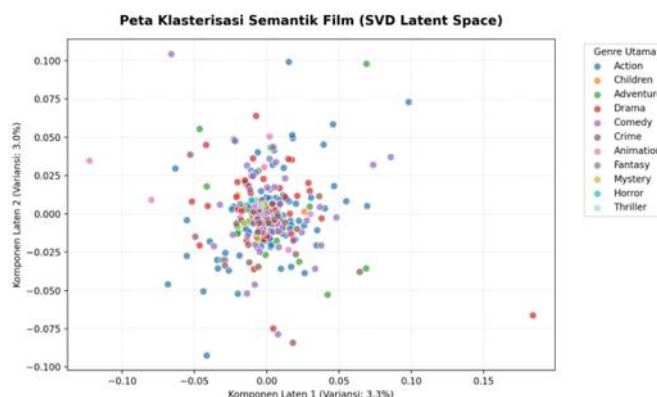
Gambar 4. Visualisasi LIME

Hasil komputasi pada Gambar 4 menunjukkan bahwa untuk *Citizen Kane*, atribut *Rata-rata Global (Basis)* memberikan kontribusi positif terbesar (+0.729), diikuti oleh *Bias Item* (+0.137). Sebaliknya, fitur *Interaksi Personal* memberikan kontribusi sangat kecil (+0.001). Terkait dengan perilaku pengguna nyata, metrik ini membuktikan secara rasional bahwa rekomendasi film klasik tersebut sangat didorong oleh opini prestise arus utama dan tingginya kualitas film secara makro, alih-alih dilandasi oleh kecocokan spesifik terhadap preferensi historis personal penggunanya.

Sebagai perbandingan komprehensif, apabila perturbasi LIME diaplikasikan pada film dengan genre yang jauh lebih tersegmentasi dan berkarakter spesifik (misalnya, film fiksi ilmiah laga seperti *The Matrix* atau film *indie* beralur lambat), distribusi pembobotan fitur diproyeksikan akan bergeser drastis. Pada kasus film *niche*, komponen *Interaksi Personal* (*dot product* antara vektor laten pengguna dan item) akan mendominasi dan mengambil alih proporsi *Rata-rata Global*. Fenomena ini mengonfirmasi kapabilitas unik dari LIME yang mampu secara dinamis membedah logika inferensi SVD kapan algoritma menyarankan sebuah item karena item tersebut memang populer (rekomendasi pasif), dan kapan algoritma benar-benar "bekerja keras" mencocokkan irisan laten item dengan selera unik (*niche*) pengguna (rekomendasi aktif).

3.6.2 Latent Factor Mapping

Pendekatan kedua beroperasi pada tataran global arsitektur SVD. *Latent Factor Mapping* bekerja dengan mengekstrak langsung 100 faktor laten item (berdasarkan konfigurasi parameter terbaik $n_factors=100$ pada SVD) dari atribut *svd_model.qi*. Karena matriks 100 dimensi tidak dapat diinterpretasikan secara visual oleh manusia, algoritma *Principal Component Analysis* (PCA) digunakan untuk mereduksi faktor laten tersebut menjadi 2 komponen dimensi utama (2D). Hasil proyeksi matriks pada ruang laten tersebut diilustrasikan pada Gambar 5.



Gambar 5. Visualisasi Pemetaan Latent Factor Menggunakan PCA 2D

Berdasarkan hasil proyeksi pada Gambar 5, Komponen Utama 1 (*Principal Component 1*) menyumbang *explained variance* sebesar 3,3%, sedangkan Komponen Utama 2 sebesar 3,0%. Dengan demikian, kedua komponen tersebut secara kumulatif hanya menjelaskan 6,3% variasi dari representasi laten yang dihasilkan oleh SVD, sedangkan 93,7% variasi lainnya berada pada komponen-komponen dengan dimensi yang lebih tinggi. Kondisi tersebut menunjukkan bahwa proyeksi dua dimensi hanya merepresentasikan sebagian kecil dari struktur variasi pada ruang laten SVD yang berdimensi 100, sehingga visualisasi tersebut tidak dapat digunakan untuk menggambarkan keseluruhan kompleksitas representasi laten secara utuh. Pemetaan titik-titik film menunjukkan adanya kedekatan spasial pada sejumlah item yang memiliki genre yang sama atau beririsan, meskipun algoritma SVD dilatih hanya menggunakan data *rating* tanpa menjadikan metadata genre sebagai masukan model. Temuan tersebut dapat menjadi indikasi eksploratif bahwa pola preferensi yang dipelajari dari *rating* mungkin memiliki keterkaitan dengan karakteristik genre, tetapi kedekatan visual tersebut belum cukup untuk membuktikan terbentuknya klusterisasi semantik secara organik. Oleh karena itu, hasil PCA dalam penelitian ini diposisikan secara ketat sebagai visualisasi eksploratif terhadap sebagian struktur ruang laten, bukan sebagai bukti bahwa faktor laten SVD secara langsung merepresentasikan genre film atau bahwa telah terbentuk klaster semantik yang tervalidasi secara matematis.

3.7 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma *Singular Value Decomposition* (SVD) secara konsisten mengungguli *K-Nearest Neighbors* (KNN) dalam mengatasi matriks berdimensi renggang, di mana parameter penentu keunggulan model ini tidak didasarkan semata pada margin akurasi (RMSE). Evaluasi komprehensif membuktikan bahwa kerangka kerja yang diusulkan berhasil menjawab celah metodologis (*gap*) yang ditinggalkan oleh penelitian-penelitian terdahulu, khususnya terkait kedalaman metrik evaluasi, rasionalisasi komputasi pada *sparsity* tinggi, dan transparansi keputusan prediksi.

Dari aspek evaluasi metrik, penelitian oleh Ariani et al. [6] yang mengevaluasi *Hybrid Filtering* dan SVD pada dataset MovieLens hanya berhenti pada pengukuran tingkat galat absolut menggunakan *Root Mean Square Error* (RMSE). Penelitian ini berhasil melampaui batasan tersebut dengan tidak hanya menghasilkan tingkat galat yang sangat rendah (RMSE 0,182090), tetapi juga memvalidasi relevansi rekomendasi sistem di dunia nyata melalui metrik klasifikasi. Capaian *Precision@10* (0,917), *Recall@10* (0,806), dan *F1-Score@10* (0,858) membuktikan secara empiris bahwa pemadatan matriks melalui *Top-N Filtering* mampu mempertahankan kualitas prediksi secara holistik tanpa menghilangkan esensi data.

Terkait optimalisasi beban arsitektural pada matriks renggang, kerangka SVD teroptimasi ini memberikan alternatif komputasi yang lebih rasional. Berbeda dengan pendekatan *Adaptive KNN* yang diajukan oleh Rahman [12] yang menuntut sumber daya besar untuk mencari irisan *rating* pada populasi yang didominasi sel kosong, SVD dalam penelitian ini secara efisien langsung mengekstraksi matriks ke dalam vektor laten. Penyusutan *sparsity* menjadi 81,20% bertindak sebagai katalis yang mengakselerasi konvergensi SVD, sebuah efisiensi arsitektural yang menyebabkan KNN sering kali gagal menemukan ketetanggaan yang reliabel.

Pembeda paling radikal sekaligus nilai kebaruan (*novelty*) dari penelitian ini adalah integrasi pada aspek transparansi melalui kerangka *Explainable AI*. Pendekatan *deep learning* modern seperti *Neural Collaborative Filtering* (NCF) yang diimplementasikan oleh Sukmana et al. [1] berisiko mengalami *overfitting* pada tingkat *sparsity* ekstrem dan beroperasi sebagai *black-box* absolut. Penelitian ini secara kritis membuktikan bahwa akurasi prediktif tidak harus mengorbankan interpretabilitas. Melalui demistifikasi LIME, sistem ini berhasil mengungkap secara kuantitatif bahwa skor prediksi pengguna pada entitas film tertentu lebih dominan digerakkan oleh sentimen arus utama (*Rata-rata Global* +0,729) dibandingkan kecocokan riwayat personal (+0,001). Temuan ini memberikan justifikasi logis yang tidak dapat disajikan oleh model rekomendasi konvensional terdahulu.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem rekomendasi film yang tangguh dalam mengatasi masalah *data sparsity* ekstrem sekaligus memberikan penjelasan prediksi yang transparan. Implementasi *Top-N Filtering* pada prapemrosesan terbukti krusial untuk memadatkan interaksi, sehingga *sparsity* matriks menurun signifikan dari 98,47% menjadi 81,20%. Pemodelan *Singular Value Decomposition* (SVD) yang dioptimasi dengan *GridSearchCV* secara empiris mengungguli *K-Nearest Neighbors* (KNN) dengan nilai *Root Mean Square Error* (RMSE) 0,182090 berbanding 0,185743. Pendekatan *Explainable AI* (XAI) terbukti efektif memecahkan masalah *black-box*; metode LIME mengungkap bahwa kontribusi *Rata-rata Global* (+0,729) mendominasi pembentukan skor prediksi dibandingkan *Interaksi Personal* (+0,001), sementara PCA mampu memvisualisasikan reduksi proyektif ruang laten SVD secara global. Secara praktis, sinergi SVD dan XAI ini mengimplikasikan bahwa penyedia *platform streaming* tidak perlu lagi mengorbankan transparansi demi akurasi, sehingga rekomendasi yang presisi dan beralasan logis dapat ditawarkan untuk membangun kepercayaan pengguna (*user trust*). Kendati demikian, penelitian ini memiliki keterbatasan metodologis karena reduksi ekstrem *long-tail items* memunculkan risiko *Popularity Bias* terhadap film *mainstream* serta kelemahan inheren sistem dalam menangani pengguna baru (*Cold-Start Problem*). Oleh karena itu, sejalan dengan tujuan awal untuk menyempurnakan kualitas rekomendasi, penelitian selanjutnya disarankan untuk mengintegrasikan *Content-Based Filtering* menjadi *Hybrid Recommender System* guna mengatasi *Cold-Start* secara

utuh, serta mengeksplorasi arsitektur *Neural Collaborative Filtering* (NCF) agar populasi *long-tail items* yang terbuang dapat diproses secara optimal tanpa mengorbankan stabilitas komputasi.

5. PERNYATAAN PENULIS

5.1 Kontribusi Penulis

Konseptualisasi: R. A. S.;

Metodologi: R. A. S.;

Perangkat Lunak: R. A. S.;

Validasi: R. A. S.;

Analisis Formal: R. A. S.;

Investigasi: R. A. S.;

Sumber Daya: R. A. S.;

Kurasi Data: R. A. S.;

Penulisan Draf Awal: R. A. S.;

Penulisan, Peninjauan, dan Penyuntingan: R. A. S., N. D. S., dan K. L.;

Visualisasi: R. A. S.;

Seluruh penulis telah membaca dan menyetujui versi manuskrip yang telah diterbitkan.

5.2 Ketersediaan Data

Data yang digunakan dalam penelitian ini merupakan dataset publik yang dapat diakses melalui platform Kaggle (Dataset MovieLens).

5.3 Pendanaan

Penelitian ini tidak menerima pendanaan eksternal.

5.4 Pernyataan Dewan Peninjau Institusional

Tidak Berlaku (*Not applicable*).

5.5 Pernyataan Persetujuan Berdasarkan Informasi

Tidak Berlaku (*Not applicable*).

5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan pribadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI dalam Proses Penulisan

Para penulis menggunakan AI Generatif untuk meningkatkan kejelasan penulisan dalam artikel ini. Para penulis telah meninjau dan menyunting konten yang dihasilkan dengan bantuan AI serta bertanggung jawab penuh atas versi akhir publikasi.

REFERENCES

- [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] D. Sartika, F. Elfaladonna, A. Octarina, and F. P. Kesuma, "Analisis Kinerja Sistem Rekomendasi yang Menggunakan Collaborative Filtering Berdasarkan Pengguna dengan Python," *Journal of Global and Multidisciplinary*, vol. 3, pp. 4686–4696, 2025, [Online]. Available: <https://journal.institercom-edu.org/index.php/multiple>
- [3] R. A. Pratama, Y. Safi'i, M. A. Nugraha, A. S. Sobihah, and N. Ifada, "Perbandingan User-Based dan Item-Based pada Sistem Rekomendasi Film Kombinasi Teknik Reduksi Dimensi dan Clustering," *Jurnal Tekno Insentif*, vol. 19, no. 1, pp. 1–14, Jul. 2025, doi: 10.36787/jti.v19i1.1662.
- [4] M. Daffa Ginting and M. Ikhsan, "Analisis dan Implementasi Singular Value Decomposition (SVD) untuk Sistem Rekomendasi Anime Berbasis Collaborative Filtering," *Jurnal Riset Komputer*, vol. 13, no. 1, pp. 2407–389, 2026, doi: 10.30865/jurikom.v13i1.9462.
- [5] Laila Novia Sari and Anita Qoiriah, "Sistem Rekomendasi Drama Korea Berdasarkan Prediksi Rating dan Kemiripan Sinopsis Menggunakan Singular Value Decomposition dan Term Frequency - Inverse Document Frequency," *Journal of Informatics and Computer Science (JINACS)*, vol. 6, no. 03, pp. 804–814, Jan. 2025, doi: 10.26740/jinacs.v6n03.p804-814.
- [6] A. P. Ariani, D. A. Handayani, P. M. Setyowati, D. K. Nugroho, and N. Fauzy, "Evaluasi Komparatif Hybrid Filtering dan Model-Based SVD pada Sistem Rekomendasi Film Menggunakan Dataset MovieLens," *Jurnal Pustaka Data (Pusat Akses Kajian Database, Analisa Teknologi, dan Arsitektur Komputer)*, 2026, doi: 10.55382/jurnalpustakadata.v6i1.1723.

- [7] Devon Vivian Gunawan and I Komang Ari Mogi, “Sistem Rekomendasi Produk pada E-Commerce Menggunakan Metode User-Based Collaborative Filtering,” *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 2, no. 4, pp. 753–760, Aug. 2024, doi: 10.24843/JNATIA.2024.v02.i04.p11.
- [8] A. Sitanggang, “Sistem Rekomendasi Anime Menggunakan Metode Singular Value Decomposition (SVD) dan Cosine Similarity,” *Jurnal Teknologi Informasi*, vol. 2, no. 2, p. 90, Nov. 2023, doi: 10.35308/jti.v2i2.7787.
- [9] A. H. Nababan, R. Nauli, L. Gaol, and F. Rahmadhani, “Pemodelan Biaya Sewa pada Data Pendidikan Internasional Menggunakan Pendekatan Machine Learning dan CRISP-DM,” *Bulletin of Information Technology (BIT)*, vol. 7, no. 1, pp. 22–30, 2026, doi: 10.47065/bit.v5i2.2557.
- [10] Y. Ban, Y. Yan, A. Banerjee, and J. He, “Research and Implementation of Movie Recommendation System Based on the MovieLens Dataset,” in *ICLR 2022 - 10th International Conference on Learning Representations*, International Conference on Learning Representations, ICLR, 2022. Accessed: Sep. 10, 2026. [Online]. Available: <https://online.odu.edu/sites/default/files/2025-04/Research%20Implementation%20Movie%20Recommendation%20System%20MovieLens%20Dataset%20Update.pdf>
- [11] Pratama Reyhan, Richasdy Donni, and Dharayani Ramanti, “Sistem Rekomendasi Collaborative Filtering Pada Smartphone Menggunakan K-Means,” Nov. 2023. Accessed: May 19, 2026. [Online]. Available: <https://garuda.kemdiktisaintek.go.id/documents/detail/3823665>
- [12] S. Rahman, “Extended Collaborative Filtering Recommendation System with Adaptive KNN and SVD,” *International Journal of Engineering and Management Research*, vol. 13, no. 4 (2023) Aug. 24, 2023, doi: 10.2139/ssrn.4550762.
- [13] K. Ramadhan Putra and I. Fathur Rahman, “MIND (Multimedia Artificial Intelligent Networking Database Pemanfaatan Metode Collaborative Filtering dengan Algoritma KNN pada Sistem Rekomendasi Produk,” *Journal MIND Journal | ISSN*, vol. 9, no. 1, pp. 113–123, 2024, doi: 10.26760/mindjournal.v9i1.113-123.
- [14] M. Badouch, N. Elmobark, and M. Boutaounte, “HITRS: A Novel Framework for Intelligent Tourism Recommender Systems based on Reinforcement Learning,” *International Journal of Intelligent Engineering and Systems*, vol. 18, no. 9, pp. 927–939, 2025, doi: 10.22266/ijies2025.1031.59.
- [15] I. Ihsan, L. Alfat, and R. Nurhajati, “Pengembangan Aplikasi Klasifikasi Individu Dengan Gangguan Spektrum Autisme Berdasarkan DSM-5 Menggunakan Pendekatan Naive Bayes,” *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 9, no. 1, pp. 13–24, Feb. 2026, doi: 10.47080/mnx97052.
- [16] C. Y. Hazizah and T. Widiyaningtyas, “Analisis Metode Collaborative Filtering menggunakan KNN dan SVD++ untuk Rekomendasi Produk E-commerce Tokopedia,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, pp. 595–604, Dec. 2024, doi: 10.29408/edumatic.v8i2.27793.
- [17] C. Wibisono, L. S. Haryadi, J. E. Widyaya, and S. L. Liliawati, “Sistem Rekomendasi Suku Cadang Berdasarkan Item Based Filtering,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, Apr. 2021, doi: 10.28932/jutisi.v7i1.3036.
- [18] Raditya Pratama Putra, “Big Data dan Artificial Intelligence dalam Produksi Pengetahuan Ilmiah atas Tantangan Metodologis,” *Jurnal Riset Manajemen Komunikasi*, pp. 113–120, Dec. 2025, doi: 10.29313/jrmk.v5i2.8721.
- [19] L. Brunot, N. Canovas, A. Chanson, N. Labroche, and W. Verdeaux, “Preference-based and local post-hoc explanations for recommender systems,” *Inf. Syst.*, vol. 108, Sep. 2022, doi: 10.1016/j.is.2022.102021.
- [20] T. Keerthika *et al.*, “Enhancing Recommendation Systems through SVD-based collaborative filtering and community detection,” *PLoS One*, vol. 21, no. 4 April, Apr. 2026, doi: 10.1371/journal.pone.0346579.