

SN-A Hybrid Approach for Recommender Systems Based on Alternating Least Squares and CatBoost

Fiddin Yusfida A'la*¹, Hartatik², Berliana Kusuma Riasti³

^{1,2,3}Applied Data Science and AI Research Group, Universitas Sebelas Maret, Indonesia

Email: ¹fiddinyusfida@gmail.com

Received: Jun 30, 2025; Revised: Sep 2, 2025; Accepted: Sep 2, 2025; Published: Sep 24, 2025

Abstract

This study aims to improve the accuracy of movie rating predictions by applying and combining collaborative filtering and machine learning techniques in a hybrid recommender system. The research utilizes the MovieLens dataset to implement two distinct approaches: the Alternating Least Squares (ALS) matrix factorization model and the CatBoost gradient boosting model. The ALS model is trained to capture latent user-item interactions, while CatBoost leverages nonlinear relationships using user and item features. A simple hybrid strategy averages the predictions from both models to evaluate potential performance gains. Experimental results show that the hybrid approach achieves lower error metrics compared to either model individually, with Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) values of 0.828 and 0.666, respectively. This demonstrates that combining latent factor models with tree-based learning can effectively reduce prediction errors by exploiting complementary strengths. The novelty of this research lies in its efficient yet effective hybridization strategy that improves recommendation quality without complex ensembling techniques. The findings suggest that even lightweight model fusion can significantly enhance predictive accuracy in recommender systems and may be adapted for other domains where combining linear and nonlinear modeling is beneficial. This research contributes to the field of Informatics and Computer Science by demonstrating that a lightweight hybridization of latent factor models and tree-based learning can significantly improve recommender system accuracy while offering practical implications for real-world digital applications.

Keywords: *ALS, CatBoost, collaborative filtering, hybrid model, MovieLens, recommender system.*

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



1. PENDAHULUAN

Perkembangan pesat layanan streaming film mendorong kebutuhan akan sistem rekomendasi yang semakin akurat dan personal [1], [2]. Sistem rekomendasi berperan penting dalam membantu pengguna menemukan konten yang sesuai preferensi, sehingga meningkatkan kepuasan dan retensi pengguna. Namun, membangun sistem rekomendasi yang akurat bukanlah tugas mudah. Tantangan utama muncul dari masalah *sparsity* (data rating yang jarang), *cold-start* (pengguna atau item baru), serta kesulitan memanfaatkan fitur kontekstual atau metadata film [3]. Selain itu, model *deep learning* yang kompleks sering kali menuntut data besar dan komputasi tinggi [4].

Untuk mengatasi hal tersebut, para peneliti banyak mengeksplorasi metode klasik seperti *Alternating Least Squares* (ALS) untuk *collaborative filtering* berbasis *matrix factorization* [5], algoritma *boosting* seperti CatBoost untuk memanfaatkan fitur side-information [6] serta pendekatan *hybrid* yang menggabungkan keunggulan keduanya. Banyak pembahasan tentang teknik *collaborative filtering* untuk menyelesaikan permasalahan, diantaranya mengatasi *sparsity* [7], rekomendasi film [8], rekomendasi *social network* [9], [10], [11], [12], rekomendasi produk [13], belanja online [14], [15], [16], [17], [18], dan rekomendasi wisata [19]. Sedangkan penggunaan teknik *boosting* diantaranya untuk

klasifikasi website jurnal [20], meningkatkan akurasi klasifikasi pada kasus *multiclass* [21], prediksi penyakit stroke [22], dan prediksi skor peserta didik [23].

Dataset MovieLens adalah salah satu dataset *benchmark* yang sangat populer dan sering digunakan untuk menguji metode-metode rekomendasi. Dengan struktur data yang kaya (rating, tag, genre), MovieLens memungkinkan penelitian pada algoritma rekomendasi dan evaluasi performa secara terukur. Beberapa penelitian menggunakan dataset MovieLens, diantaranya adalah penelitian yang menggabungkan *Genetic Algorithm* (GA) dengan *content-based filtering* untuk mengoptimalkan bobot fitur genre, pemeran, dan kru dalam menghasilkan rekomendasi yang lebih baik. Hasil eksperimen menunjukkan bahwa GA mampu meningkatkan metrik Precision@k dan F-Measure sekitar 58%, sehingga menghasilkan rekomendasi yang lebih akurat dan relevan dibandingkan metode lain [24]. Penelitian lain yang menggunakan dataset MovieLens yaitu membandingkan teknik klasifikasi seperti BayesNet, Decision Table, Logistic, k-NN, Random Forest, dan lainnya berdasarkan enam metrik evaluasi, serta meningkatkan akurasi melalui teknik *resampling*. Sepuluh algoritma pelatihan dibandingkan menggunakan alat WEKA pada dataset MovieLens, dengan hasil menunjukkan bahwa Random Forest mencapai akurasi tertinggi sebesar 99,39%. Penelitian ini diharapkan membantu peneliti lain dalam memilih teknik klasifikasi terbaik untuk meningkatkan akurasi sistem rekomendasi [25].

Penelitian ini dirancang untuk menjawab permasalahan terkait efektivitas berbagai pendekatan dalam sistem rekomendasi film, khususnya dalam konteks prediksi rating pengguna. Dengan semakin besarnya jumlah konten pada layanan streaming, diperlukan metode yang mampu memprediksi preferensi pengguna secara lebih akurat dan personal.

Permasalahan utama yang diangkat dalam penelitian ini adalah bagaimana performa tiga pendekatan berbeda, yaitu ALS sebagai metode *collaborative filtering* berbasis faktor matriks, CatBoost sebagai model *gradient boosting* yang mampu memanfaatkan fitur tambahan, serta metode *Hybrid* yang dirancang untuk menggabungkan keunggulan kedua pendekatan tersebut dalam memprediksi rating film pada dataset MovieLens. Penelitian ini bertujuan untuk mengevaluasi seberapa baik masing-masing metode mampu menghasilkan prediksi rating yang mendekati nilai sebenarnya, serta menilai apakah strategi *hybrid* dapat memberikan peningkatan akurasi yang signifikan dibandingkan model individual.

Kesenjangan penelitian (*research gap*) yang diangkat adalah minimnya kajian yang secara langsung membandingkan pendekatan faktor laten tradisional seperti ALS dengan model pembelajaran mesin nonlinier modern seperti CatBoost, serta menggabungkannya dalam sebuah strategi hybrid yang teruji pada dataset rekomendasi populer. Dengan demikian, orisinalitas penelitian ini terletak pada kontribusi dalam menghadirkan kerangka evaluasi yang komprehensif sekaligus menunjukkan potensi penggabungan kedua pendekatan sebagai solusi baru yang lebih akurat untuk sistem rekomendasi.

2. METODE PENELITIAN

Proses diawali dengan mendefinisikan tujuan analisis atau sistem yang akan dibangun, yaitu membuat model prediksi atau sistem rekomendasi berbasis algoritma ALS, CatBoost, dan pendekatan hybrid. Tahap ini mencakup penyiapan lingkungan kerja di mana penelitian ini menggunakan platform *Google Colab* sebagai media eksperimen dan Python sebagai Bahasa Pemrograman.

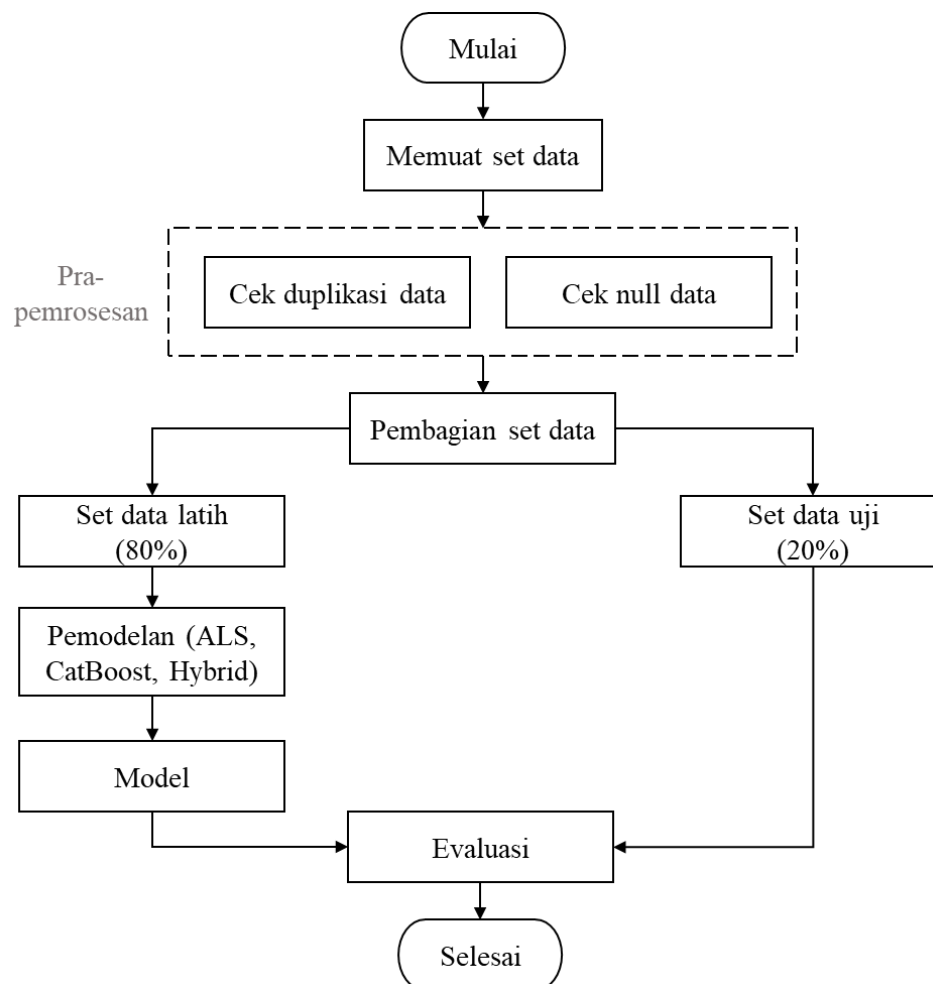
Data yang akan digunakan diimpor dari sumber dataset publik MovieLens 1M dalam bentuk .dat. Proses ini memastikan data tersedia dalam format yang sesuai untuk analisis. Pada tahap ini dilakukan eksplorasi awal untuk melihat dimensi data dan tipe fitur. Pembahasan lebih detail akan disajikan pada subbab selanjutnya.

Tahap pra-pemrosesan bertujuan untuk membersihkan dan menyiapkan data agar layak digunakan dalam pemodelan. Pada tahap ini dilakukan beberapa langkah penting. Pertama adalah pengecekan duplikasi data. Langkah ini bertujuan mengidentifikasi baris-baris data yang benar-benar identik. Baris duplikat perlu dihapus agar tidak mengganggu distribusi data, mencegah bias, dan

memastikan model tidak mempelajari pola yang salah atau berlebihan dari data yang tidak unik. Kedua adalah pengecekan terhadap data yang kosong atau hilang (*null values*). Berdasarkan pengecekan, dataset ini tidak ditemukan data duplikat atau data null.

Data yang sudah dibersihkan kemudian dibagi menjadi dua bagian, yaitu set data latih dan set data uji. Set data latih (80%) digunakan untuk melatih model dan menemukan pola. Sementara itu, set uji (20%) digunakan di akhir untuk menilai performa model pada data yang benar-benar baru.

Tahap selanjutnya adalah inti pemodelan, di mana algoritma dipakai membangun model prediktif. Terdapat tiga algoritma yang digunakan, yaitu ALS untuk memfaktorkan matriks interaksi user-item dalam sistem rekomendasi, CatBoost sebagai algoritma boosting yang unggul untuk data kategorikal, dan pendekatan Hybrid yang menggabungkan prediksi dari keduanya. Model yang telah dioptimasi kemudian dievaluasi pada set uji untuk mengukur kinerjanya secara objektif. Proses evaluasi ini biasanya menggunakan metrik seperti RMSE (*Root Mean Squared Error*) atau MAE (*Mean Absolute Error*). Alur dari eksperimen ditunjukkan pada Gambar 1.



Gambar 1. Diagram alir eksperimen

2.1. Dataset

Dataset ini terdiri dari 1.000.209 rating anonim terhadap sekitar 3.900 film, diberikan oleh 6.040 pengguna yang mendaftar di layanan MovieLens pada tahun 2000 [26]. Dataset ini banyak digunakan sebagai *benchmark* dalam penelitian sistem rekomendasi berbasis *collaborative filtering* dan *content-based filtering*. Contoh dataset setelah proses *merge* ditunjukkan pada Tabel 1.

Tabel 1. Contoh data MovieLens 1M

User ID	Movie ID	Rating	Timestamp	Gender	Age	Occupation	Zip-code	Title	Genres
1	1193	5	978300760	F	1	10	48067	One Flew Over the Cuckoo's Nest (1975)	Drama
1	661	3	978302109	F	1	10	48067	James and the Giant Peach (1996)	Animation
1	914	3	978301968	F	1	10	48067	My Fair Lady (1964)	Musical
...	...	...	...	...	...	...	...	...	...

2.2. Alternating Least Squares (ALS)

Alternating Least Squares (ALS) adalah algoritma *matrix factorization* yang umum dipakai untuk sistem rekomendasi. ALS memfaktorkan matriks interaksi pengguna-item menjadi dua matriks faktor laten (user dan item), yang dikalibrasi bergantian untuk meminimalkan *squared error*. Kelebihannya adalah skalabilitas dan kemudahan paralelisasi, cocok untuk data besar [27] [28] [29].

Model *Singular Value Decomposition* (SVD) diimplementasikan menggunakan *library Surprise*, yang terkenal luas dalam penelitian sistem rekomendasi. Data pelatihan dikonversi ke format *Surprise* dengan rentang rating 1–5, lalu model dilatih pada keseluruhan data pelatihan. Parameter SVD yang digunakan adalah *default* di *Surprise* meliputi pengaturan jumlah faktor laten (*n_factors*=100), jumlah epoch (20), pilihan penggunaan bias (*True*), skema inisialisasi faktor, laju pembelajaran (*lr_all*=0.005) dan regularisasi (*reg_all*=0.02), serta opsi granular untuk pengaturan *learning rate* dan regularisasi masing-masing komponen. Hal ini memungkinkan penyesuaian model secara mendetail untuk menangkap pola interaksi pengguna–item dan mengendalikan risiko overfitting.

2.3. CatBoost

CatBoost adalah algoritma *gradient boosting* berbasis pohon keputusan yang dioptimasi untuk menangani fitur kategorikal secara langsung tanpa banyak pra-pemrosesan. Algoritma ini dirancang untuk mengurangi *overfitting* dan bias prediksi lewat teknik seperti *ordered boosting*. CatBoost sering digunakan untuk masalah prediksi tabular karena akurat, cepat, dan mudah dipakai [30] [31] [6].

Nilai parameter yang digunakan pada eksperimen ini merupakan nilai *default* dari *library*. Parameter beserta nilainya yang digunakan adalah sebagai berikut: iterasi bernilai 500, jumlah boosting rounds (atau *trees*) yang akan dibangun. *Learning_rate*=0.05 menunjukkan ukuran langkah untuk pembaruan di setiap boosting step. *Depth*=6 untuk menentukan kedalaman maksimum untuk setiap pohon keputusan. *Verbose*=100 menunjukkan kontrol output log selama pelatihan. *Verbose*=100 artinya akan mencetak progress setiap 100 iterasi.

2.4. Hybrid

Hybrid adalah pendekatan gabungan yang memanfaatkan kelebihan dua model berbeda [32], [33], [34]. Dalam hal ini, prediksi dari ALS dan CatBoost digabungkan dengan cara memberi bobot yang sama pada hasil masing-masing model, sehingga dapat menghasilkan prediksi akhir yang lebih akurat dan seimbang dengan memanfaatkan kekuatan keduanya. Persamaan untuk menghitung hasil hybrid ditunjukkan pada Persamaan 1.

$$\hat{y}_{Hybrid} = \frac{\hat{y}_{ALS} + \hat{y}_{CatBoost}}{2} \quad (1)$$

Di mana:

$\hat{y}_{ALS}$ = prediksi dari model ALS

$\hat{y}_{CatBoost}$ = prediksi dari model CatBoost

2.5. Evaluation

MAE adalah rata-rata dari nilai absolut selisih antara nilai sebenarnya (*actual*) dan nilai prediksi (*predicted*). Dalam sistem rekomendasi, MAE mengukur seberapa jauh prediksi rating dari nilai rating sebenarnya secara rata-rata, tanpa memperhitungkan arah error (positif atau negatif). Formula matematika untuk menghitung MAE ditunjukkan pada Persamaan 2.

$$MAE = \frac{1}{n} \sum_{i=1}^n |c| \quad (2)$$

Di mana:

n = jumlah data (contoh rating yang diuji)

y_i = nilai sebenarnya (*true rating*)

$\hat{y}_i$ = nilai prediksi (*predicted rating*)

RMSE adalah akar kuadrat dari rata-rata kuadrat selisih antara nilai sebenarnya dan nilai prediksi. RMSE memberi bobot lebih besar pada *error* yang besar karena kuadrat, sehingga lebih sensitif terhadap *outlier*. Persamaan matematika untuk menghitung nilai RMSE ditunjukkan pada Persamaan 3.

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

Di mana:

n = jumlah data (contoh rating yang diuji)

y_i = nilai sebenarnya (*true rating*)

$\hat{y}_i$ = nilai prediksi (*predicted rating*)

3. HASIL DAN PEMBAHASAN

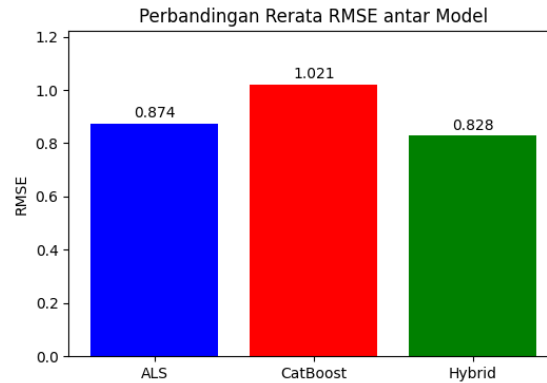
3.1. Hasil Eksperimen

Hasil eksperimen ketiga model dievaluasi menggunakan RMSE. Evaluasi performa model menggunakan metrik RMSE memberikan gambaran kuantitatif mengenai tingkat akurasi prediksi dari masing-masing pendekatan yang diuji. Model ALS menghasilkan nilai RMSE sebesar 0.874, menunjukkan kemampuan yang cukup baik dalam pemodelan.

Sementara itu, model CatBoost mencatat RMSE sebesar 1.021, yang relatif lebih tinggi. Hal ini mengindikasikan bahwa meskipun CatBoost unggul dalam menangani data kategorikal dan hubungan nonlinier, pada konfigurasi saat ini model tersebut belum mencapai tingkat akurasi prediksi yang setara dengan ALS.

Pendekatan Hybrid yang menggabungkan prediksi kedua model berhasil menurunkan nilai RMSE menjadi 0.828. Hasil ini menunjukkan bahwa strategi *ensembling* dapat memanfaatkan keunggulan komplementer dari masing-masing model, sehingga mampu menghasilkan prediksi yang lebih akurat secara keseluruhan. Dengan demikian, model Hybrid dapat dipertimbangkan sebagai alternatif yang

lebih optimal dalam sistem rekomendasi berbasis data ini, karena menunjukkan performa prediksi terbaik di antara ketiga pendekatan yang dievaluasi. Grafik perbandingan hasil evaluasi RMSE ditunjukkan pada Gambar 2.



Gambar 2. Perbandingan RMSE antar model

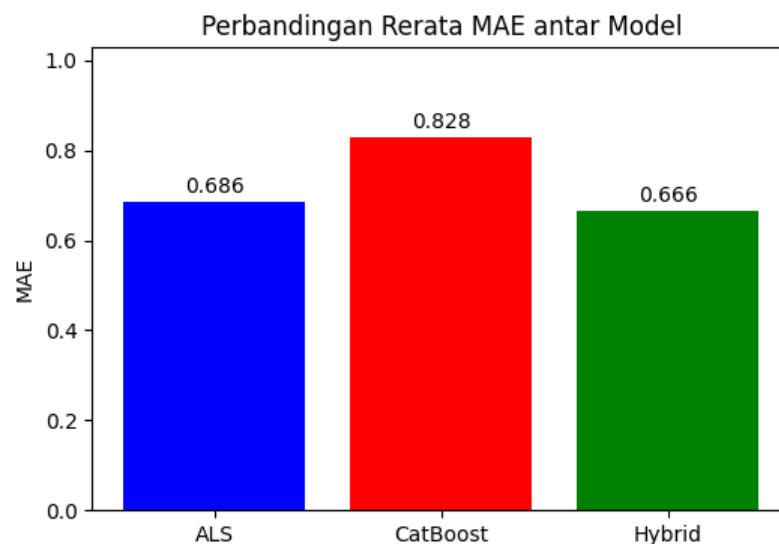
Tabel 2 menunjukkan hasil pengujian sebanyak 30 kali dan hasil menunjukkan bahwa pendekatan hybrid mampu secara signifikan meningkatkan akurasi.

Tabel 2. Hasil 30 kali pengujian menggunakan RMSE

Iterasi	ALS	CatBoost	Hybrid
1	0.874	1.021	0.827
2	0.876	1.02	0.828
3	0.876	1.021	0.828
4	0.874	1.021	0.828
5	0.873	1.02	0.827
6	0.874	1.021	0.828
7	0.874	1.021	0.828
8	0.873	1.021	0.827
9	0.871	1.021	0.827
10	0.871	1.021	0.827
11	0.872	1.021	0.827
12	0.874	1.021	0.828
13	0.874	1.021	0.827
14	0.873	1.021	0.828
15	0.874	1.021	0.827
16	0.873	1.02	0.827
17	0.873	1.021	0.828
18	0.874	1.02	0.828
19	0.874	1.021	0.828
20	0.873	1.021	0.828
21	0.875	1.021	0.828
22	0.875	1.021	0.827
23	0.874	1.02	0.827
24	0.875	1.021	0.828

25	0.874	1.021	0.828
26	0.871	1.022	0.827
27	0.872	1.021	0.828
28	0.873	1.021	0.828
29	0.875	1.021	0.828
30	0.873	1.022	0.828

Hasil eksperimen ketiga model kemudian dievaluasi menggunakan MAE. Hasil evaluasi model menunjukkan variasi kinerja yang jelas pada metrik MAE. Model ALS memperoleh nilai MAE sebesar 0.686, yang menandakan tingkat kesalahan rata-rata yang relatif rendah dalam memprediksi rating pengguna. Sementara itu, model CatBoost menghasilkan MAE sebesar 0.828, yang merupakan nilai tertinggi di antara ketiga model. Hal ini menunjukkan bahwa, pada data yang sama, CatBoost memiliki kesalahan prediksi rata-rata yang lebih besar dibandingkan ALS. Grafik perbandingan MAE ditunjukkan pada Gambar 3.



Gambar 3. Perbandingan MAE antar model

Tabel 3 menunjukkan hasil pengujian sebanyak 30 kali, hasil ini menunjukkan bahwa pendekatan hybrid mampu secara signifikan meningkatkan akurasi.

Tabel 3. Hasil 30 kali pengujian menggunakan MAE

Iterasi	ALS	CatBoost	Hybrid
1	0.874	1.021	0.827
2	0.876	1.02	0.828
3	0.876	1.021	0.828
4	0.874	1.021	0.828
5	0.873	1.02	0.827
6	0.874	1.021	0.828
7	0.874	1.021	0.828
8	0.873	1.021	0.827
9	0.871	1.021	0.827
10	0.871	1.021	0.827

11	0.872	1.021	0.827
12	0.874	1.021	0.828
13	0.874	1.021	0.827
14	0.873	1.021	0.828
15	0.874	1.021	0.827
16	0.873	1.02	0.827
17	0.873	1.021	0.828
18	0.874	1.02	0.828
19	0.874	1.021	0.828
20	0.873	1.021	0.828
21	0.875	1.021	0.828
22	0.875	1.021	0.827
23	0.874	1.02	0.827
24	0.875	1.021	0.828
25	0.874	1.021	0.828
26	0.871	1.022	0.827
27	0.872	1.021	0.828
28	0.873	1.021	0.828
29	0.875	1.021	0.828
30	0.873	1.022	0.828

Contoh perhitungan untuk pendekatan hybrid adalah sebagai berikut

$$\begin{array}{ll}
 \text{Prediksi dari ALS} & 3.7 \\
 \text{Prediksi dari CatBoost} & 4.1 \\
 \text{Hybrid (ALS + CatBoost)} & \frac{3.7+4.1}{2} = \frac{7.8}{2} = 3.9
 \end{array}$$

Pendekatan hybrid yang menggabungkan prediksi dari ALS dan CatBoost secara sederhana melalui rata-rata, memberikan hasil MAE paling rendah yaitu 0.666. Capaian ini mengindikasikan bahwa metode gabungan mampu memanfaatkan keunggulan masing-masing model sekaligus mereduksi kelemahan yang ada pada model individual. Dengan demikian, strategi hybrid menunjukkan potensi yang lebih baik dalam meningkatkan akurasi prediksi rating pengguna pada sistem rekomendasi yang dibangun. Hybrid secara signifikan meningkatkan akurasi prediksi dibandingkan ALS maupun CatBoost. Peningkatan terhadap CatBoost sangat besar, sedangkan terhadap ALS peningkatannya moderat tapi konsisten dan signifikan.

3.2. Analisis Hasil

Evaluasi performa ketiga model (ALS, CatBoost, dan Hybrid) dilakukan menggunakan dua metrik utama: RMSE dan MAE. Kedua metrik ini digunakan untuk menilai seberapa jauh prediksi model menyimpang dari nilai rating sebenarnya, dengan RMSE memberikan penalti lebih besar pada kesalahan besar dibanding MAE.

ALS menunjukkan performa yang cukup baik dengan nilai RMSE dan MAE yang relatif rendah. RMSE di bawah 0.9 menunjukkan bahwa rata-rata error kuadrat yang diakarnya masih terjaga pada kisaran yang moderat. MAE sebesar 0.686 menunjukkan bahwa, secara rata-rata absolut, deviasi prediksi ALS dari nilai sebenarnya adalah sekitar 0.7 poin rating. ALS cukup baik dalam mempelajari

hubungan antara pengguna dan item, tetapi tetap memiliki keterbatasan dalam menangani faktor-faktor nonlinier atau fitur side-information yang lebih kompleks.

Sebaliknya, CatBoost menampilkan nilai RMSE dan MAE yang lebih tinggi dibanding ALS. RMSE di atas 1 mengindikasikan bahwa model cenderung membuat prediksi dengan error yang lebih besar secara kuadrat, sedangkan MAE mendekati 0.83 mengonfirmasi bahwa secara rata-rata kesalahan prediksinya juga lebih tinggi. Hal ini bisa disebabkan oleh karakteristik data rating yang memiliki sparsity tinggi atau kurangnya fitur tambahan yang informatif, yang membuat model gradient boosting berbasis pohon kesulitan menangkap interaksi kompleks antar pengguna dan item dengan akurasi tinggi.

Pendekatan hybrid, yang menggabungkan prediksi dari kedua model dengan cara rata-rata sederhana, menunjukkan hasil yang paling baik pada kedua metrik. RMSE turun menjadi 0.828, lebih rendah dari RMSE kedua model individual. Hal yang sama terlihat pada MAE, yang menjadi 0.666, lebih rendah dari MAE pada ALS maupun CatBoost. Penurunan error ini mengindikasikan bahwa kombinasi model berhasil mengompensasi kelemahan masing-masing: prediksi CatBoost yang cenderung lebih bias dilunakkan oleh prediksi ALS, sementara prediksi ALS yang kadang over-smooth diperbaiki sebagian oleh sifat nonlinier CatBoost.

4. DISKUSI

4.1. Interpretasi Hasil

Analisis hasil evaluasi menunjukkan bahwa penggunaan pendekatan hybrid sederhana berhasil memberikan prediksi dengan kesalahan lebih rendah dibandingkan penggunaan salah satu model secara terpisah. Hal ini menegaskan potensi metode ensembling dalam sistem rekomendasi: dengan menggabungkan pendekatan berbasis faktor laten (ALS) dan pembelajaran nonlinier (CatBoost), sistem dapat memanfaatkan keunggulan masing-masing untuk menghasilkan prediksi yang lebih akurat dan stabil. Strategi penggabungan model seperti ini layak dipertimbangkan sebagai pendekatan yang lebih unggul dalam pengembangan sistem rekomendasi berbasis rating.

4.2. Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Data yang digunakan masih terbatas pada informasi rating tanpa mempertimbangkan faktor kontekstual lain seperti waktu atau preferensi dinamis pengguna. Selain itu, model yang digunakan belum dioptimasi secara mendalam melalui penyesuaian *hyperparameter* atau validasi silang yang lebih luas. Pendekatan hybrid yang digunakan juga masih sederhana, hanya berupa rata-rata prediksi, sehingga potensinya untuk kombinasi yang lebih kompleks belum dieksplorasi sepenuhnya.

5. KESIMPULAN

Berdasarkan hasil evaluasi, pendekatan hybrid yang menggabungkan model ALS dan CatBoost terbukti memberikan performa terbaik dengan nilai RMSE dan MAE yang lebih rendah dibandingkan masing-masing model secara terpisah. Hal ini menunjukkan bahwa strategi penggabungan model mampu memanfaatkan keunggulan metode faktor laten ALS serta kemampuan nonlinier CatBoost untuk menghasilkan prediksi rating yang lebih akurat. Temuan ini mendukung penggunaan teknik ensemble sebagai pendekatan yang efektif dalam meningkatkan kualitas sistem rekomendasi berbasis rating. Selain itu, penelitian ini memiliki urgensi dan signifikansi penting dalam bidang Ilmu Informatika/Ilmu Komputer karena dapat berkontribusi pada pengembangan metode rekomendasi yang lebih adaptif, akurat, dan aplikatif, sehingga mendukung inovasi dalam sistem informasi, e-commerce, maupun personalisasi layanan digital.

UCAPAN TERIMA KASIH

Karya ini didukung oleh Grup riset **Applied Data Science and AI** di bawah skema Hibah Penelitian Group Riset **Universitas Sebelas Maret** dengan Nomor Kontrak Penelitian: **371/UN27.22/PT.01.03/2025**.

DAFTAR PUSTAKA

- [1] F. A. Darmawan and Z. K. A. Baizal, "An Approach to a Group Movie Recommender System using Matrix Factorization-based Collaborative Filtering," *JURNAL INFOTEL*, vol. 16, no. 2, pp. 243–254, May 2024, doi: 10.20895/infotel.v16i2.1126.
- [2] B. T. Imani and E. B. Setiawan, "Recommender System Based on Matrix Factorization on Twitter Using Random Forest (Case Study: Movies on Netflix)," *International Journal on Information and Communication Technology (IJoICT)*, vol. 8, no. 2, pp. 11–21, Dec. 2022, doi: 10.21108/ijoict.v8i2.655.
- [3] S. Kulkarni and S. F. Rodd, "Context Aware Recommendation Systems: A review of the state of the art techniques," *Comput Sci Rev*, vol. 37, p. 100255, Aug. 2020, doi: 10.1016/j.cosrev.2020.100255.
- [4] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A Survey on Bias and Fairness in Machine Learning," *ACM Comput Surv*, vol. 54, no. 6, pp. 1–35, Jul. 2022, doi: 10.1145/3457607.
- [5] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative Filtering for Implicit Feedback Datasets," in *2008 Eighth IEEE International Conference on Data Mining*, IEEE, Dec. 2008, pp. 263–272. doi: 10.1109/ICDM.2008.22.
- [6] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," in *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems*, ACM, 2018.
- [7] H. Al-Bashiri, M. A. Abdulgaber, A. Romli, and F. Hujainah, "Collaborative Filtering Recommender System: Overview and Challenges," *Adv Sci Lett*, vol. 23, no. 9, pp. 9045–9049, Sep. 2017, doi: 10.1166/asl.2017.10020.
- [8] S. Annisa, D. P. Rini, and A. Abdiansah, "Collaborative Filtering Recommendation System Using A Combination of Clustering and Association Rule Mining," *Journal of Information Systems and Informatics*, vol. 6, no. 3, pp. 1499–1516, Sep. 2024, doi: 10.51519/journalisi.v6i3.802.
- [9] A. Fareed, S. Hassan, S. B. Belhaouari, and Z. Halim, "A collaborative filtering recommendation framework utilizing social networks," *Machine Learning with Applications*, vol. 14, p. 100495, Dec. 2023, doi: 10.1016/j.mlwa.2023.100495.
- [10] J. Xu, "Analysis of Social Media Algorithm Recommendation System," *Studies in Social Science & Humanities*, vol. 1, no. 3, Oct. 2022, doi: 10.56397/SSSH.2022.10.06.
- [11] Y. A. Bekeneva and T. U. Eze, "Improved Social Network User Recommendation System—The Machine Learning Approach," in *INTELS'22*, Basel Switzerland: MDPI, Jun. 2023, p. 13. doi: 10.3390/engproc2023033013.
- [12] A. Anandhan, M. A. Ismail, L. Shuib, W. S. N. Aiza, and M. M. Elaish, "Social Media Recommender Systems (SMRS): A Bibliometric Analysis Study 2000–2021," *IEEE Access*, vol. 10, pp. 35479–35497, 2022, doi: 10.1109/ACCESS.2022.3161497.
- [13] A. A. Patoulia, A. Kiourtis, A. Mavrogiorgou, and D. Kyriazis, "A Comparative Study of Collaborative Filtering in Product Recommendation," *Emerging Science Journal*, vol. 7, no. 1, pp. 1–15, Oct. 2022, doi: 10.28991/ESJ-2023-07-01-01.

-
- [14] I. T. Afolabi, O. S. Makinde, and O. O. Oladipupo, "Semantic Web mining for Content-Based Online Shopping Recommender Systems," *International Journal of Intelligent Information Technologies*, vol. 15, no. 4, pp. 41–56, Oct. 2019, doi: 10.4018/IJIT.2019100103.
- [15] W.-E. Kong, T.-E. Tai, P. Naveen, and H. A. Santoso, "Performance Evaluation on E-Commerce Recommender System based on KNN, SVD, CoClustering and Ensemble Approaches," *Journal of Informatics and Web Engineering*, vol. 3, no. 3, pp. 63–76, Oct. 2024, doi: 10.33093/jiwe.2024.3.3.4.
- [16] F. T. Abdul Hussien, A. M. S. Rahma, and H. B. Abdul Wahab, "Recommendation Systems For E-commerce Systems An Overview," *J Phys Conf Ser*, vol. 1897, no. 1, p. 012024, May 2021, doi: 10.1088/1742-6596/1897/1/012024.
- [17] S. Gosh, N. Nahar, M. A. Wahab, M. Biswas, M. S. Hossain, and K. Andersson, "Recommendation System for E-commerce Using Alternating Least Squares (ALS) on Apache Spark," Springer, 2021, pp. 880–893. doi: 10.1007/978-3-030-68154-8_75.
- [18] M. O. Puspitaningtyas, K. Puspita, Y. Alkhalifi, and Y. A. Wardani, "Implementation of Support Vector Machine, Particle Swarm Optimization, And Naïve Bayes Algorithms in Sentiment Analysis of Product Reviews: A Case Study of e-Commerce Lazada," *Jurnal Riset Informatika*, vol. 7, no. 2, pp. 30–37, Mar. 2025, doi: 10.34288/jri.v7i2.362.
- [19] E. Pantano, C.-V. Priporas, N. Stylos, and C. Dennis, "Facilitating tourists' decision making through open data analyses: A novel recommender system," *Tour Manag Perspect*, vol. 31, pp. 323–331, Jul. 2019, doi: 10.1016/j.tmp.2019.06.003.
- [20] N. S. F. Putri, A. P. Wibawa, H. Ar Rasyid, A. Nafalski, and U. R. Hasyim, "Boosting and bagging classification for computer science journal," *International Journal of Advances in Intelligent Informatics*, vol. 9, no. 1, p. 27, Mar. 2023, doi: 10.26555/ijain.v9i1.985.
- [21] Y. Pristyanto, Z. Mukarabiman, and A. F. Nugraha, "Extreme Gradient Boosting Algorithm to Improve Machine Learning Model Performance on Multiclass Imbalanced Dataset," *JOIV: International Journal on Informatics Visualization*, vol. 7, no. 3, pp. 710–715, Sep. 2023, doi: 10.30630/joiv.7.3.1102.
- [22] D. A. Setyarini, A. A. M. D. Gayatri, C. S. K. Aditya, and D. R. Chandranegara, "Stroke Prediction with Enhanced Gradient Boosting Classifier and Strategic Hyperparameter," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 2, pp. 477–490, Mar. 2024, doi: 10.30812/matrik.v23i2.3555.
- [23] D. N. Muhammadiyah, H. A. E. Nugraha, V. R. S. Nastiti, and C. S. K. Aditya, "Students Final Academic Score Prediction Using Boosting Regression Algorithms," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 1, p. 154, Mar. 2024, doi: 10.26555/jiteki.v10i1.28352.
- [24] A. Abdolmaleki and M. H. Rezvani, "An optimal context-aware content-based movie recommender system using genetic algorithm: a case study on MovieLens dataset," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 36, no. 8, pp. 1485–1511, Nov. 2024, doi: 10.1080/0952813X.2022.2153279.
- [25] M. T. Alam *et al.*, "Comparative Analysis of Machine Learning based Filtering Techniques using MovieLens dataset," *Procedia Comput Sci*, vol. 194, pp. 210–217, 2021, doi: 10.1016/j.procs.2021.10.075.
- [26] F. M. Harper and J. A. Konstan, "The MovieLens Datasets," *ACM Trans Interact Intell Syst*, vol. 5, no. 4, pp. 1–19, Jan. 2016, doi: 10.1145/2827872.
- [27] H. A. Adyatma and Z. K. A. Baizal, "Book Recommender System Using Matrix Factorization with Alternating Least Square Method," *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1286–1292, Jul. 2023, doi: 10.47065/josh.v4i4.3816.
-

-
- [28] N. P. Widyasprana, I. M. W. Wirawan, I. G. S. Astawa, and I. D. M. Bayu Atmaja Darmawan, "Analisis Algoritma ALS-MF (Alternating Least Square Matrix Factorization) dengan SVD (Singular Value Decomposition) pada Metode Collaborative Filtering," *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, vol. 13, no. 2, p. 347, Jul. 2024, doi: 10.24843/JLK.2024.v13.i02.p12.
- [29] Y. Zhou, D. Wilkinson, R. Schreiber, and R. Pan, "Large-Scale Parallel Collaborative Filtering for the Netflix Prize," in *Algorithmic Aspects in Information and Management*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 337–348. doi: 10.1007/978-3-540-68880-8_32.
- [30] M. A. Harriz, N. V. Akbariani, H. Setiyowati, and H. Santoso, "Classifying Village Fund in West Java, Indonesia Using Catboost Algorithm," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 2, pp. 691–697, May 2023, doi: 10.35870/jimik.v4i2.269.
- [31] A. F. L. Ptr, M. M. Siregar, and I. Daniel, "Analysis of Gradient Boosting, XGBoost, and CatBoost on Mobile Phone Classification," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 6, no. 2, pp. 661–670, Apr. 2024, doi: 10.47709/cnahpc.v6i2.3790.
- [32] B. F. Azevedo, A. M. A. C. Rocha, and A. I. Pereira, "Hybrid approaches to optimization and machine learning methods: a systematic literature review," *Mach Learn*, vol. 113, no. 7, pp. 4055–4097, Jul. 2024, doi: 10.1007/s10994-023-06467-x.
- [33] J. Paredes, D. Chávez, R. Isa-Jara, and D. Vargas, "A hybrid machine learning algorithm approach to predictive maintenance tasks: A comparison with machine learning algorithms," *Results in Engineering*, vol. 26, p. 105137, Jun. 2025, doi: 10.1016/j.rineng.2025.105137.
- [34] Aditya Gumilar, Sri Suryani Prasetyowati, and Yuliant Sibaroni, "Performance Analysis of Hybrid Machine Learning Methods on Imbalanced Data (Rainfall Classification)," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 3, pp. 481–490, Jul. 2022, doi: 10.29207/resti.v6i3.4142.