



Pemodelan Sistem Rekomendasi Pada E-Commerce Berbasis Collaborative Filtering dan Machine Learning

Recommendation System Modeling in E-Commerce Based on Collaborative Filtering and Machine Learning

Novita Rahmayuna¹⁾; Cicik Harini²⁾

¹⁾ Sistem Informasi, Fakultas Sistem Informasi, Universitas Bina Nusantara, Indonesia

²⁾ Manajemen, Fakultas Ekonomi, Universitas Pandanaran, Indonesia

Email: ¹⁾ cici@unpand.ac.id

How to Cite :

Rahmayuna, N. Harini, C. (2025). Pemodelan Sistem Rekomendasi Pada E-Commerce Berbasis Collaborative Filtering dan Machine Learning. *Sinta Journal* ,6 (2), 593–600 DOI: <https://doi.org/10.37638/sinta.6.2.593-600>

ARTICLE HISTORY

Received [24 October 2025]

Revised [15 November 2025]

Accepted [02 December 2025]

KEYWORDS

sistem rekomendasi,
e-commerce,
collaborative filtering,
machine learning

ABSTRAK

Sistem rekomendasi saat ini telah dikembangkan untuk memberikan sekumpulan item yang cocok berdasarkan preferensi pengguna. Pada e-commerce, sistem rekomendasi merupakan alat yang efektif untuk menawarkan produk yang sesuai dengan minat pengguna. Membangun sekumpulan produk yang tepat untuk ditawarkan kepada pengguna merupakan tantangan dalam sistem rekomendasi. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi menggunakan pendekatan collaborative filtering dan algoritma machine learning. Penelitian ini menggunakan data matriks dari rating yang diberikan oleh pengguna terhadap produk, sehingga pemilihan algoritma machine learning perlu disesuaikan. Hasil terbaik ditunjukkan dari kinerja evaluasi yang diukur menggunakan root mean square error (RMSE), dimana algoritma SVD++ memberikan kinerja yang paling baik yaitu sebesar 0.9846.

ABSTRACT

Currently a recommendation system has been developed to provide a suitable set of items based on user preferences. In e-commerce, recommendation systems are an effective tool for offering products that match user interests. Building the right set of products to offer users is a challenge in recommendation systems. This study aims to develop a recommendation system using a collaborative filtering approach and machine learning algorithms. This study uses a data matrix from the ratings given by users to the product, so that the selection of machine learning algorithms needs to be adjusted. The best results are shown from the performance

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



evaluation which is measured using the root mean square error (RMSE), where SVD++ shows the best performance, which is equal to 0.9846.

PENDAHULUAN

Fenomena big data mengakibatkan ketersediaan informasi terus bertambah. Ketersediaan informasi yang berlebihan menyebabkan sulitnya pengguna dalam menemukan sesuatu yang mereka butuhkan. Sistem rekomendasi menjadi salah satu alat yang dikembangkan untuk membantu pengguna dalam menemukan hal-hal yang sesuai dengan preferensi mereka. Sistem rekomendasi bekerja dengan mempelajari kemiripan fitur dari suatu konten dan riwayat interaksi yang dilakukan oleh pengguna. Dengan demikian, sistem rekomendasi menjadi salah satu komponen yang sangat penting untuk platform online seperti pada bidang *e-commerce*, musik, film, dokumen dan lain sebagainya (Walek & Fajmon, 2023).

Perkembangan bisnis saat ini telah memanfaatkan *e-commerce* sebagai alat untuk memasarkan produk. *E-commerce* sebagai *platform* untuk menjalankan aktivitas penjualan hingga pembelian produk memiliki beberapa komponen untuk memberikan penawaran produk kepada pengguna (Yu, dkk, 2005). Sistem rekomendasi merupakan salah satu fitur didalam *e-commerce* yang bekerja dengan memberikan penawaran produk terbaik bagi pengguna. Pengembangan sistem rekomendasi pada dasarnya memberikan penawaran berdasarkan preferensi pengguna yang bisa didapatkan melalui pengalaman transaksi pengguna. Hal ini dapat menimbulkan permasalahan ketika tidak ditemukan riwayat preferensi pengguna karena status pengguna masih baru, maupun ketika produk belum memiliki ulasan (Yalcin dan Bilge, 2022).

Collaborative filtering merupakan salah satu pendekatan dalam sistem rekomendasi yang memanfaatkan informasi yang muncul dari pengguna, contohnya ulasan dan penilaian (rating) produk (Ba, dkk, 2013; Anwar dan Uma, 2021). Informasi ini menjadi penting karena rekomendasi yang akan dikembangkan memperhatikan penilaian dari pengguna, tidak hanya menganalisis informasi dari konten item (Xia, dkk, 2015). Namun demikian, transaksi-transaksi yang hanya melibatkan satu jenis pembelian produk menjadi suatu kendala karena menyebabkan kesulitan dalam menganalisis rekomendasi produk lain yang relevan dengan pilihan pengguna (De Basio, dkk, 2023). Dengan demikian, perlu diperhatikan terkait popularitas item. Selain itu, pada beberapa kasus, permasalahan skalabilitas yang tinggi menjadi suatu kendala dalam membangun sistem rekomendasi. Beberapa penelitian telah dilakukan dengan menggunakan pendekatan kemiripan antara user dan item seperti *cosine similarity* dan pendekatan data mining seperti KNN (Kuo, dkk, 2021; Fkih, 2022). Pemodelan dilakukan dengan memanfaatkan fitur-fitur yang didapat dari *collaborative filtering* maupun *content-based filtering* (Walek & Fajmon, 2023).

Sistem rekomendasi berbasis *collaborative filtering* dapat dilakukan berdasarkan rating produk yang diberikan oleh pengguna. Matriks *user-item rating* dimanfaatkan untuk memberikan representasi terhadap data. Terdapat beberapa pendekatan yang dapat digunakan untuk data berbentuk matriks, diantaranya pengukuran kesamaan dan *matrix factorization*. Penelitian ini bertujuan untuk melakukan komparasi terhadap kinerja yang dihasilkan pada pemodelan sistem rekomendasi menggunakan pendekatan pengukuran kesamaan maupun *machine learning*.

METODE PENELITIAN

Pada penelitian ini serangkaian langkah akan dilakukan untuk menghasilkan sistem rekomendasi. Beberapa langkah yang dilakukan meliputi tahapan *preprocessing*, pemodelan sistem dan evaluasi.

Preprocessing dilakukan untuk mendapatkan data yang diperlukan dengan menganalisis kekurangan informasi pada data. Pada tahapan ini proses *filtering* untuk tidak melibatkan data-data yang mengandung missing value dilakukan. Selain itu, untuk membangun rekomendasi yang tepat, popularitas suatu produk dipertimbangkan. Data yang akan digunakan adalah produk yang memiliki ting-kat popularitas tinggi, yaitu produk yang sudah memiliki sejumlah transaksi pada waktu yang berbeda. Selain itu, terdapat tiga atribut yang digunakan pada penelitian ini mengacu pada *userid*, *itemid* dan *rating*.

Pemodelan dilakukan setelah data telah melalui tahapan *preprocessing*. Pada tahapan ini, beberapa algoritma akan diaplikasikan untuk diukur kinerjanya. Beberapa algoritma yang akan digunakan pada pemodelan sistem diantaranya tertera pada Tabel 1.

Tabel 1. Algoritma yang digunakan

Pendekatan	Algoritma
Pengukuran Kesamaan	<i>Pearson Similarity</i> <i>Cosine Similarity</i> <i>Mean Squared Difference (MSD) Similarity</i> <i>Pearson-baseline Similarity</i>
<i>Matrix Factorization</i>	<i>Singular Value Decomposition (SVD)</i> <i>Singular Value Decomposition ++ (SVD++)</i> <i>Non-negative Matrix Factorization (NMF)</i>

Data dengan struktur matriks *user-item rating* digunakan pada penelitian ini, sehingga pemilihan pendekatan untuk melakukan pemodelan sistem rekomendasi perlu diperhatikan. Pada data yang memiliki struktur matriks, sistem rekomendasi dapat diselesaikan menggunakan beberapa pendekatan yaitu pendekatan pengukuran kesamaan dan *matrix factorization*. Penelitian ini akan melakukan pemodelan dan komparasi pada beberapa algoritma yang telah disajikan pada Tabel 1.

Evaluasi dilakukan untuk mengukur kinerja sistem. Pada artikel ini, nilai *root mean square error* (RMSE) dihitung untuk mengukur tingkat kesalahan dalam memprediksi produk yang akan direkomendasikan. Formula (1) menunjukkan rumus untuk mengukur RMSE. Selain itu, *mean absolute error* (MAE) digunakan untuk mengukur sejauh mana perbedaan antara nilai prediksi dan nilai sebenarnya pada data uji. Perhitungan MAE ditunjukkan pada Formula (2) (Gadzar dan Hidri, 2020).

$$RMSE = \sqrt{\frac{1}{n} \sum_{u,i} (\hat{r}_{ui} - r_{ui})^2} \quad (1)$$

$$MAE = \sqrt{\frac{1}{n} (\hat{r}_{ui} - r_{ui})} \quad (2)$$

Dimana, $\hat{r}_{ui}$ menunjukkan nilai rating pada item yang diprediksi, r_{ui} nilai rating pada data asli, dan n menunjukkan ukuran dari data rekomendasi.

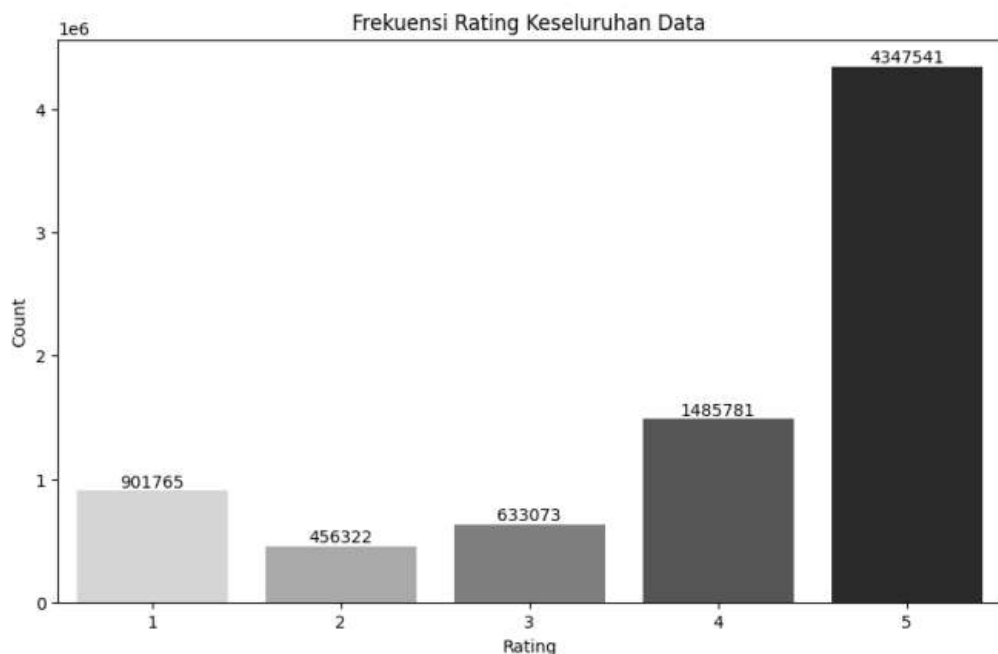
HASIL DAN PEMBAHASAN

Dataset yang digunakan merupakan data transaksi pengguna yang melakukan penilaian pada produk elektronik Amazon yang bersumber dari Kaggle (<https://www.kaggle.com/datasets/saurav9786/amazon-productreviews?resource=download>).

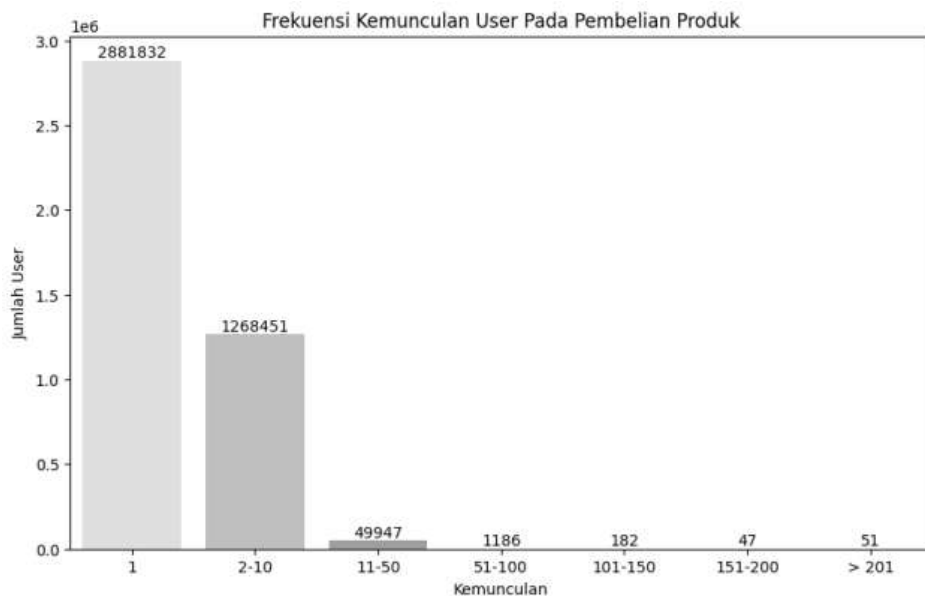
Data terdiri atas 7.824.482 record transaksi dengan rentang penilaian (rating) antara 1 sampai 5. Atribut yang digunakan pada penelitian ini diantaranya *userid*, *itemid* dan *rating* produk. Data akan dibagi secara acak menjadi dua bagian yaitu data latih dan data uji 70 : 30. Data latih digunakan sebagai acuan untuk membangun model rekomendasi dan data uji digunakan sebagai acuan untuk mengukur kinerja rekomendasi.

Analisis, Eksplorasi dan Persiapan Data

Dataset yang digunakan merupakan data transaksi pembelian produk makanan yang merupakan yang telah dinilai oleh pengguna e-commerce pada produk elektronik. Atribut yang digunakan untuk membangun sistem rekomendasi diantaranya *userid*, *itemid* dan *rating*. Dataset yang digunakan memiliki rentang rating dengan penilai 1-5, dimana digambarkan pada Gambar 1. Selain itu, frekuensi kemunculan user terhadap suatu produk juga dianalisis dan ditunjukkan pada Gambar 2.



Gambar 1. Frekuensi Jumlah Rating pada Dataset

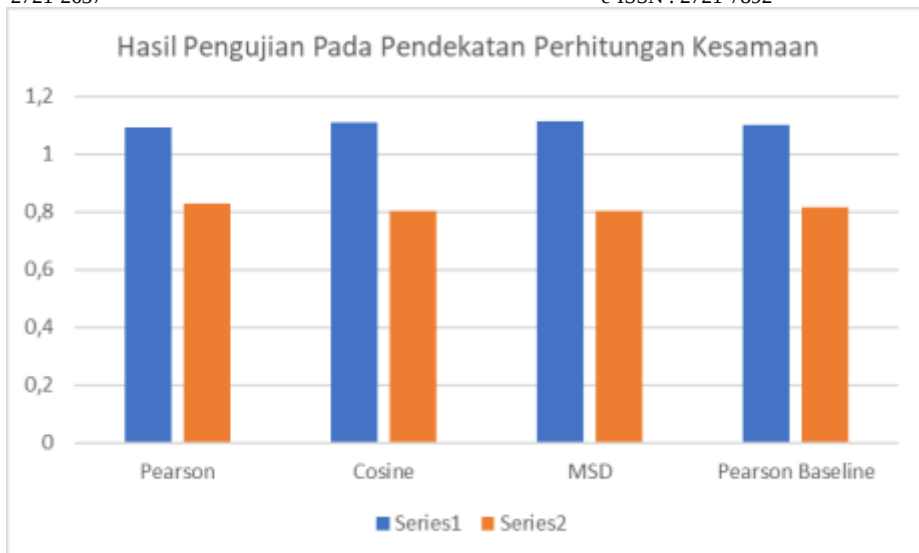


Gambar 2. Frekuensi Kemunculan Pengguna Pada Pembelian Produk

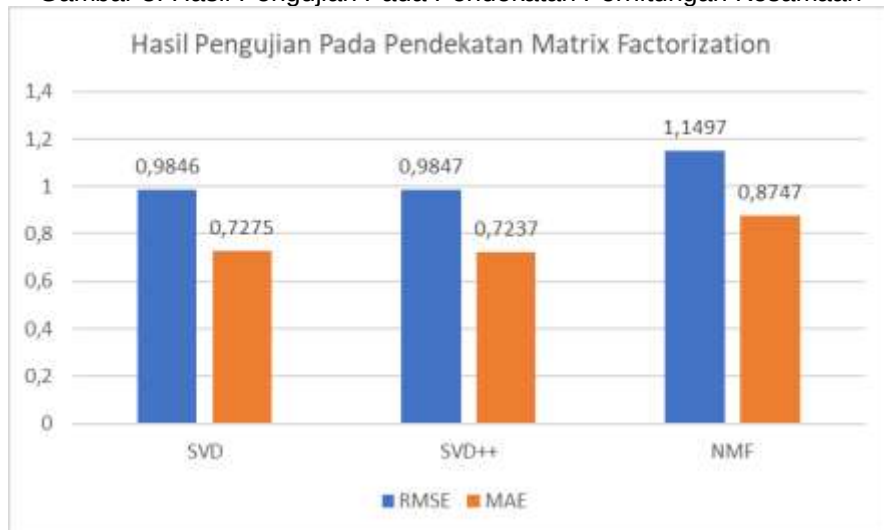
Berdasarkan Gambar 2, dapat dilihat bahwa terdapat banyak pengguna baru yang telah memberikan rating pada produk, dilihat dari kemunculan pengguna yang hanya sekali dalam memberikan penilaian. Penelitian ini melakukan penyaringan pada data yang akan digunakan, dimana data yang akan digunakan adalah data dari pengguna yang telah melakukan penilaian terhadap produk lebih dari 50 kali. Hal ini dilakukan untuk mengurangi permasalahan skalabilitas dan meningkatkan nilai penilaian suatu produk yang telah dinilai oleh pengguna yang berpengalaman. Selanjutnya, data yang telah disaring akan digunakan sebagai masukan dalam pemodelan sistem rekomendasi. Pada penelitian ini, algoritma machine learning digunakan untuk menyelesaikan permasalahan sistem rekomendasi dengan data yang berbentuk matriks user-item.

Hasil Pengujian

Pengujian dilakukan menggunakan beberapa algoritma machine learning yang menggunakan pendekatan pengukuran kesamaan dan *matrix factorization*. Gambar 3 dan Gambar 4 menunjukkan hasil pengujian dari algoritma yang menggunakan pendekatan perhitungan kesamaan dan *matrix factorization*.



Gambar 3. Hasil Pengujian Pada Pendekatan Perhitungan Kesamaan



Gambar 4. Hasil Pengujian Pada Pendekatan Matrix Factorization

Berdasarkan hasil pengujian, kinerja yang dihasilkan pada menggunakan pendekatan *matrix factorization* lebih baik dibandingkan pendekatan pengukuran kesamaan. Hal ini ditunjukkan pada nilai RMSE dan MAE pada pendekatan *matrix factorization* yang lebih kecil dibandingkan pada pendekatan pengukuran kesamaan. Semakin kecil nilai RMSE dan MAE, maka semakin baik sistem rekomendasi yang dibangun bekerja. Berdasarkan hasil pengujian, algoritma SVD memiliki kinerja 0.0001 lebih baik dibandingkan SVD++ dari sisi RMSE. Sedangkan, SVD++ memiliki kinerja yang lebih baik sebesar 0.0038 dibandingkan SVD. Dengan demikian, dapat disimpulkan bahwa kinerja SVD++ memberikan kinerja yang lebih baik dibandingkan yang lainnya.

KESIMPULAN DAN SARAN

Kesimpulan

Sistem rekomendasi pada penelitian ini dibangun untuk dengan pendekatan collaborative filtering dan machine learning. Data yang digunakan adalah data transaksi pada produk e-commerce. Atribut yang digunakan untuk membangun sistem rekomendasi diukur berdasarkan userid, itemid dan rating. Beberapa algoritma yang sesuai dengan struktur data matriks digunakan pada penelitian ini, yaitu berdasarkan pendekatan pengukuran kesamaan dan matrix factorization. Kinerja sistem diukur menggunakan mengkalkulasi RMSE. Hasil menunjukkan bahwa kinerja terbaik dihasilkan pada model menggunakan algoritma SVD++ dengan nilai error yang paling rendah. Sistem rekomendasi telah berhasil dilakukan dengan menggunakan beberapa algoritma machine learning. Namun penelitian ini hanya menggunakan informasi rating dari riwayat transaksi. Kedepannya, informasi-informasi penting lainnya seperti informasi dari konten dapat digunakan untuk memodelkan sistem rekomendasi. Kedepannya, perlu dilakukan eksplorasi lebih dalam terkait penggunaan fitur-fitur implisit yang ada pada item maupun user, sehingga model rekomendasi yang dibangun bisa lebih sesuai dengan preferensi pengguna.

DAFTAR PUSTAKA

- Anwar, T., & Uma, V. (2021). Comparative study of recommender system approaches and movie recommendation using collaborative filtering. *International Journal of System Assurance Engineering and Management*, 12(3), 426–436. <https://doi.org/10.1007/s13198-021-01087-x>
- Ba, Q., Li, X., & Bai, Z. (2013). Clustering Collaborative Filtering Recommendation System Based on SVD Algorithm. 2013 IEEE 4th International Conference on Software Engineering and Service Science, 963–967.
- Beel, J., Gipp, B., Langer, S., & Breiteringer, C. (2016). Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries*, 17(4), 305–338. <https://doi.org/10.1007/s00799-015-0156-0>
- De Biasio, A., Montagna, A., Aioli, F., & Navarin, N. (2023). A systematic review of value-aware recommender systems. *Expert Systems with Applications*, 226. <https://doi.org/10.1016/j.eswa.2023.120131>
- Fkih, F. (2022). Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison. *Dalam Journal of King Saud University - Computer and Information Sciences* (Vol. 34, Nomor 9, hlm. 7645–7669). King Saud bin Abdulaziz University. <https://doi.org/10.1016/j.iksuci.2021.09.014>
- Gazdar, A., & Hidri, L. (2020). A new similarity measure for collaborative filtering based recommender systems ☆. 188, 105058. <https://doi.org/10.1016/j.knosys>
- Kuo, R. J., Chen, C. K., & Keng, S. H. (2021). Application of hybrid metaheuristic with perturbation-based K-nearest neighbors algorithm and densest imputation to collaborative filtering in recommender systems. *Information Sciences*, 575, 90–115. <https://doi.org/10.1016/j.ins.2021.06.026>
- Walek, B., & Fajmon, P. (2023). A hybrid recommender system for an online store using a fuzzy expert system. *Expert Systems with Applications*, 212. <https://doi.org/10.1016/j.eswa.2022.118565>
- Wang, J., Han, P., Miao, Y., & Zhang, F. (2019). A Collaborative Filtering Algorithm Based on SVD and Trust Factor.
- Xia, H., Fang, B., Gao, M., Ma, H., Tang, Y., & Wen, J. (2015). A novel item anomaly detection approach against shilling attacks in collaborative recommendation

- systems using the dynamic time interval segmentation technique. *Information Sciences*, 306, 150–165. <https://doi.org/10.1016/j.ins.2015.02.019>
- Yalcin, E., & Bilge, A. (2022). Evaluating unfairness of popularity bias in recommender systems: A comprehensive user-centric analysis. *Information Processing and Management*, 59(6). <https://doi.org/10.1016/j.ipm.2022.103100>
- Yu, L., Liu, L., & Li, X. (2005). A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce. *Expert Systems with Applications*, 28(1), 67–77. <https://doi.org/10.1016/j.eswa.2004.08.013>
- Zhang, W., Zhang, X., & Chen, D. (2021). Causal neural fuzzy inference modeling of missing data in implicit recommendation system. *Knowledge-Based Systems*, 222. <https://doi.org/10.1016/j.knosys.2020.106678>
- Zhu, Z., Yan, M., Deng, X., & Gao, M. (2022). Rating prediction of recommended item based on review deep learning and rating probability matrix factorization. *Electronic Commerce Research and Applications*, 54. <https://doi.org/10.1016/j.eierap.2022.101160>