

IMPLEMENTASI XGBOOST UNTUK REKOMENDASI SMARTPHONE MENGUNAKAN CONTENT-BASED FILTERING

Muhhamad Juna Edi¹, Aryanto^{2, *}

^{1,2}Progam Studi Statistika Falkutas Matematika dan Ilmu Pengetahuan Alam Universitas Cenderawasih

*Correspondance author: aryanto.dandan@gmail.com

ABSTRACT

The demand for recommendation systems has increased as a result of advances in information technology. These systems must assist customers in selecting electronic products, particularly smartphones, according to their preferences. The aim of this research is to develop a content-based recommendation system supported by the Extreme Gradient Boosting (XGBoost) algorithm, which will enhance prediction accuracy and recommendation relevance. The dataset used is sourced from the public platform Kaggle and includes user ratings and specification data for various Realme and Samsung smartphone products. Data processing involves data cleaning, conversion of numerical and categorical data types, and the selection of key features such as price, RAM, processor, camera, battery, and screen size. Cosine similarity is used to measure the degree of similarity between products in order to generate relevant recommendations, while the XGBoost model predicts product rating values based on their features. To evaluate the model, the Root Mean Squared Error (RMSE) metric is used. The evaluation results, which indicate excellent predictive performance, show an RMSE of 0.1529. In addition, the system was tested through simulated recommendation scenarios. In these scenarios, products with high similarity levels and high predicted ratings were identified as viable alternatives. This study demonstrates that the combination of XGBoost and content-based filtering is highly effective in creating intelligent and accurate recommendation systems that provide added value to users during the purchasing decision process. To further enhance the quality of recommendation results, this system could be developed further by integrating hybrid methods or user interaction data.

Keywords: Recommendation System, Content-Based Filtering, XGBoost, Smartphone, Cosine Similarit

ABSTRAK

Kebutuhan akan sistem rekomendasi telah meningkat sebagai akibat dari kemajuan teknologi informasi. Sistem rekomendasi ini harus dapat membantu pelanggan memilih barang elektronik, khususnya smartphone, sesuai dengan preferensi mereka. Tujuan dari penelitian ini adalah untuk membuat sistem rekomendasi berbasis konten yang dibantu oleh algoritma Extreme Gradient Boosting (XGBoost), yang akan meningkatkan akurasi prediksi dan relevansi rekomendasi. Dataset yang digunakan berasal dari platform publik Kaggle dan mencakup rating dan data spesifikasi pengguna untuk berbagai produk smartphone Realme dan Samsung. Pengolahan data mencakup pembersihan data, konversi tipe data numerik dan kategorikal, dan memilih fitur penting seperti harga, RAM, prosesor, kamera, baterai, dan



ukuran layar. Metode cosine similarity digunakan untuk mengukur tingkat kemiripan antarproduk untuk menghasilkan rekomendasi yang relevan, sementara model XGBoost memprediksi nilai rating produk berdasarkan fiturnya. Untuk menilai model, metrik Root Mean Squared Error (RMSE) digunakan. Hasil evaluasi, yang menunjukkan performa prediksi yang sangat baik, adalah 0.1529. Selain itu, sistem diuji melalui simulasi skenario rekomendasi. Dalam skenario ini, produk dengan tingkat kemiripan yang tinggi dan prediksi rating yang tinggi diidentifikasi sebagai alternatif yang layak. Studi ini menunjukkan bahwa kombinasi XGBoost dan filter berbasis konten sangat efektif untuk membuat sistem rekomendasi yang pintar, akurat, dan mampu memberikan nilai tambahan kepada pengguna selama proses pengambilan keputusan pembelian. Untuk meningkatkan kualitas hasil rekomendasi, sistem ini dapat dikembangkan lebih lanjut dengan menggabungkan metode hybrid atau data interaksi pengguna.

Kata kunci: Sistem Rekomendasi, Content-Based Filtering, XGBoost, Smartphone, Cosine Similarity

ARTICLE INFO

Submission received: 12 June 2025

Accepted: 31 August 2025

Revised: 15 August 2025

Published: 31 August 2025

Available on: <https://doi.org/10.32493/sm.v7i2.xxxx>

StatMat: Jurnal Statistika dan Matematika is licenced under a Creative Commons Attribution-ShareAlike 4.0 International License.

1. PENDAHULUAN

Pandangan konsumen tentang produk elektronik, khususnya smartphone, telah berubah sebagai akibat dari kemajuan teknologi informasi. Banyak pilihan dan kompleksitas spesifikasi sering menyulitkan untuk menemukan produk terbaik. Sistem rekomendasi adalah solusi penting. Menurut Jannach (2010) alat penyaringan informasi yang dikenal sebagai sistem rekomendasi berfungsi untuk memprediksi preferensi pengguna dengan melihat data interaksi sebelumnya, seperti pembelian, ulasan, dan aktivitas di keranjang belanja. Sistem ini sangat membantu pengguna dalam membuat keputusan, khususnya dalam hal e-commerce yang memiliki banyak barang yang berbeda. Burke (2007) membagi sistem rekomendasi menjadi lima pendekatan utama: *collaborative filtering*, *content-based filtering*, *knowledge-based*, *hybrid*, dan *demographic filtering*.

Content-Based Filtering (CBF) merupakan salah satu pendekatan sistem rekomendasi yang bekerja dengan informasi fitur dari item itu sendiri, seperti spesifikasi teknis dan deskripsi produk. (Shahbazi 2019) Dengan metode ini, sistem dapat memberikan rekomendasi yang relevan berdasarkan kemiripan karakteristik antar produk. Algoritma *Extreme Gradient Boosting* (XGBoost) digunakan untuk meningkatkan akurasi prediksi sistem rekomendasi ini.

XGBoost adalah algoritma pembelajaran mesin berbasis peningkatan yang terkenal karena skalabilitas, efisiensi, dan kemampuan untuk menangani data besar dan kompleks. XGBoost membantu sistem rekomendasi memprediksi interaksi pengguna terhadap item tertentu dengan melihat fitur perilaku pengguna seperti waktu akses dan riwayat pembelian.

Dari penjelasan tersebut, penelitian ini bertujuan untuk mengembangkan sistem rekomendasi produk smartphone berbasis content-based filtering yang diperkuat oleh model prediktif XGBoost untuk membantu pengguna memilih produk dengan prediksi rating terbaik dan kemiripan spesifikasi yang sesuai dengan preferensi mereka.

2. METODOLOGI

Jenis Penelitian

Penelitian ini menggunakan pendekatan data science kuantitatif dan eksperimen komputasional untuk membangun dan menguji sistem rekomendasi produk smartphone berbasis konten yang diperkuat oleh model XGBoost.

Sumber Data

Data yang digunakan berasal dari data sekunder yang diperoleh dari platform publik Kaggle. Dataset ini berisi informasi tentang berbagai fitur smartphone Samsung dan Realme, termasuk harga, nama produk, kapasitas RAM, prosesor, tampilan layar, kamera, baterai, dan rating.

Teknik Pengumpulan Data

Data dikumpulkan melalui proses pengunduhan dataset dalam format CSV. Kemudian, proses integrasi dan penggabungan antara dua dataset berdasarkan atribut yang sepadan digunakan. Hasil integrasi disimpan dan diolah menggunakan Python dan pustaka seperti Pandas, SciKit-Learn, dan XGBoost.

Teknik Pengolahan Data

Langkah-langkah pengolahan data dilakukan sebagai berikut:

1. Preprocessing
Menghapus data duplikat dan baris kosong, mengubah format harga dan RAM ke tipe numerik, serta mengubah semua data kategorikal menjadi bentuk numerik dengan LabelEncoder.
2. Penggabungan Data
Menggabungkan dua dataset dari masing-masing merek menjadi satu dataset utama.
3. Feature Selection
Memilih fitur-fitur yang relevan untuk digunakan dalam model seperti price, storage_ram, os_processor, camera, battery, dan display.
4. Pembentukan Target
Rating pengguna digunakan sebagai target prediksi oleh model.
5. Split Data
Data dibagi menjadi data pelatihan dan pengujian dengan rasio 80:20.
6. Pemodelan
Model utama penelitian ini adalah XGBoost Regressor, yang digunakan melalui pustaka XGBoost. Model ini dilatih untuk memprediksi nilai rating berdasarkan fitur. Selain itu, sebagai bagian dari filter berbasis konten, *cosine similarity*

digunakan untuk menghitung seberapa mirip produk satu sama lain.

7. Evaluasi

Sistem dievaluasi dengan menggunakan metrik *Root Mean Squared Error* (RMSE) untuk mengetahui seberapa akurat prediksi model terhadap data rating pengguna. Selain itu, sistem juga diuji dengan simulasi skenario rekomendasi, yaitu memberikan satu input produk yang disukai pengguna dan melihat lima produk teratas yang memiliki kemiripan yang paling besar dan diprediksi memiliki rating tinggi. Penggabungan kedua dataset setelah penyamaan nama kolom.

3. HASIL DAN PEMBAHASAN

Hasil Nilai RMSE

Data dari merek Samsung dan Realme yang telah diproses dan dibersihkan dikonsolidasi menjadi satu Data Frame dan diolah untuk memastikan keseragaman format dan validitas isi untuk melatih model XGBoost. Setelah melakukan pelatihan model dengan fitur numerik seperti harga, model menghasilkan nilai *Root Mean Squared Error* (RMSE) sebesar 0.1529, yang menunjukkan prediksi yang sangat baik terhadap nilai rating.

Rekomendasi Produk Rating Tertinggi

Berdasarkan hasil prediksi model, berikut adalah lima produk dengan prediksi rating tertinggi:

Tabel 1. Prediksi Rating Tertinggi

Nama Produk	Harga	Prediksi Rating
Samsung Galaxy S23 Ultra 5G (Phantom Black)	111990	4.70
Samsung Galaxy S23 Ultra 5G (Cream)	111000	4.69
Samsung Galaxy J7 (Black)	8999	4.59
Samsung Galaxy M01 (Blue)	8999	4.59
Realme X2 Pro (Neptune Blue)	35999	4.60

Hasil ini menunjukkan bahwa, berdasarkan preferensi pengguna umum, sistem dapat mengidentifikasi produk dengan rating tinggi.

Rekomendasi Berdasarkan Kemiripan (Content-Based)

Sistem melakukan perhitungan kemiripan (*cosine similarity*) jika pengguna menyukai produk "OPPO Reno10 5G (Ice Blue, 256 GB)" berdasarkan fitur produk tersebut dibandingkan dengan semua produk lain dalam dataset. Berikut ini adalah hasil rekomendasi dengan kemiripan konten tertinggi dan prediksi rating yang baik.

Tabel 2. Rekomendasi berdasarkan Kemiripan (Content-Based)

Rekomendasi Produk	Prediksi Rating	Similarity
Samsung Galaxy M53 5G (Mystique)	4.06	0.999999

Rekomendasi Produk	Prediksi Rating	Similarity
Green, 128 GB)		
Samsung Galaxy M53 5G (Deep Ocean Blue, 128 GB)	4.00	0.999998
Samsung Galaxy M53 5G (Emerald Brown, 128 GB)	4.10	0.999998
Samsung Galaxy A23 5G (Light Blue, 128 GB)	3.80	0.999997
Samsung Galaxy M52 5G (Icy Blue, 128 GB)	4.23	0.999997

Sistem ini dapat mengidentifikasi produk dengan spesifikasi yang hampir identik, membantu pelanggan yang mencari alternatif dengan fitur yang sama tetapi harga yang berbeda.

4. Kesimpulan

Penelitian ini menguji algoritma XGBoost dalam sistem rekomendasi smartphone berbasis konten. Sistem dapat memberikan rekomendasi yang akurat dan relevan dengan memanfaatkan fitur utama produk seperti harga, ram, prosesor, baterai, dan lainnya, serta menggunakan *cosine similarity* untuk menghitung kemiripan. Dengan nilai RMSE sebesar 0.1529, hasil evaluasi menunjukkan bahwa model bekerja dengan baik. Sistem ini dapat memprediksi produk dengan rating tinggi dan merekomendasikan produk alternatif yang sebanding, sehingga membantu pelanggan membuat keputusan pembelian yang lebih baik. Untuk meningkatkan kualitas rekomendasi, sistem ini dapat diperluas dengan menambahkan data interaksi pengguna atau menggabungkannya dengan metode filter hibrida.

DAFTAR PUSTAKA

- Burke, R. (2007). Hybrid web recommender systems. In *The Adaptive Web* (pp. 377–408). Springer.
- Gao, M., Wu, Z., & Jiang, F. (2011). UserRank for item-based collaborative filtering recommendation. *Information Processing Letters*, 111(9), 440–446.
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). *Recommender systems: An introduction*. Cambridge University Press.
- Shahbazi, Z., & Byun, Y. C. (2019). Product recommendation based on content-based filtering using XGBoost classifier. *International Journal of Advanced Science and Technology*, 29, 6979–6988.
- Shahbazi, Z., Hazra, D., Park, S., & Byun, Y. C. (2020). Toward improving the prediction accuracy of product recommendation system using extreme gradient boosting and encoding approaches. *Symmetry*, 12(9), 1566. <https://doi.org/10.3390/sym12091566>