

Skincare Product Recommendation System Using Hybrid Ensemble Learning

Yulviani Puteri Puspita Sari ^{1*}, Ika Nur Fajri ^{2*}, Anggit Dwi Hartanto ^{3*}

* Sistem Informasi, Universitas Amikom Yogyakarta

yulvianipps@students.amikom.ac.id¹, fajri@amikom.ac.id², anggit@amikom.ac.id³

Article Info

Article history:

Received 2025-07-24

Revised 2026-02-05

Accepted 2026-02-27

Keyword:

*Ensemble Learning,
Hybrid Recommendation System,
Random Forest,
TF-IDF,
SVD Matrix Factorization.*

ABSTRACT

The global beauty and personal care products market reached USD 557.24 billion in 2023 and is projected to reach USD 937.13 billion by 2030, growing at a CAGR of 7.7%. This growth is driven by increasing consumer awareness of personal appearance and the proliferation of e-commerce platforms, particularly in the Asia-Pacific region. However, the wide variety of skincare products creates information overload, making it difficult for consumers to identify products that suit their specific needs. Traditional recommendation systems generally rely on a single approach, either content-based filtering, which struggles to capture user preferences, or collaborative filtering, which faces cold-start and data sparsity issues. This study proposes a hybrid ensemble learning approach that integrates three complementary techniques: (1) TF-IDF content-based filtering to analyze product similarities based on brand, category, and product attributes; (2) SVD matrix factorization collaborative filtering to capture latent patterns of user-product interactions through synthetic user data generation; and (3) Random Forest as a meta-learner to intelligently combine the outputs of the two methods. The proposed system was evaluated using a dataset of more than 7,500 skincare products from Sociolla via the Kaggle repository, covering more than 300 brands with detailed product attributes including ratings, reviews, prices, and customer engagement metrics. The hybrid ensemble approach showed strong predictive performance with an R^2 score of 0.830, explaining 83.0% of the variance in product ratings. The system successfully recommended five products from five different brands (Biyu, True-to-skin, Jacqueline, The-aubree, and Biore) with ensemble scores ranging from 0.872 to 0.996, demonstrating cross-brand recommendation capabilities. Feature importance analysis shows relatively equal contributions from $\log_wishlist$ (29.7%), $brand_encoded$ (25.6%), $\log_reviews$ (24.1%), and $category_encoded$ (20.6%). These findings indicate that this hybrid approach effectively overcomes the limitations of single-method recommendation systems and can be adapted to various e-commerce product categories, providing a more relevant and personalized shopping experience for consumers.



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

I. PENDAHULUAN

Pasar kecantikan dan *skincare* global menunjukkan pertumbuhan yang sangat pesat, tercatat mencapai \$557.24 miliar pada tahun 2023, dan diperkirakan akan terus berkembang dengan tingkat pertumbuhan tahunan gabungan (CAGR) sebesar 7,7% dari tahun 2024 hingga 2030, hingga mencapai \$937,13 miliar [1]. Pertumbuhan ini dipicu oleh

meningkatnya kesadaran konsumen akan pentingnya penampilan, yang ditambah dengan berkembangnya transformasi digital yang semakin meluas, membuka peluang baru bagi pasar, dimana *platform* digital menjadi saluran utama dalam penemuan dan pembelian produk *skincare* [2]. Tren ini semakin relevan bagi generasi millennial dan Gen Z

yang mengandalkan *platform* online untuk mencari informasi dan menilai produk sebelum melakukan pembelian [3], [4].

Dalam konteks ini, *platform* e-commerce kecantikan menghadapi tantangan yang kompleks dalam membantu konsumen memilih produk dari ribuan pilihan *skincare*, *makeup*, dan produk perawatan tubuh dengan harga dan merek yang bervariasi [5], [6]. Konsumen sering kesulitan memilih produk yang sesuai kebutuhan, terutama ketika mencari alternatif produk dengan karakteristik serupa dari *brand* yang berbeda [7]. Situasi ini menciptakan kelebihan informasi yang dapat menghambat pengalaman berbelanja, menurunkan *conversion rate*, dan mengurangi kepuasan konsumen secara keseluruhan. Oleh karena itu, sistem rekomendasi efektif menjadi salah satu pilar strategis untuk meningkatkan kualitas personalisasi dan pengalaman berbelanja dalam pengambilan keputusan pembelian [6].

Namun, penerapan sistem rekomendasi pada *platform* e-commerce kecantikan saat ini umumnya masih bergantung pada pendekatan tunggal, baik *content-based* filtering maupun *collaborative filtering*, yang masing-masing memiliki keterbatasan dalam merekomendasikan produk yang beragam [8]. Pendekatan *content-based filtering* yang mengandalkan kesamaan atribut produk seperti kategori dan merek cenderung menghasilkan rekomendasi yang terbatas pada produk serupa dalam hal variasi, sehingga mengurangi eksplorasi konsumen terhadap alternatif produk yang potensial [6], [7]. Sistem ini juga tidak mampu menangkap preferensi pengguna yang kompleks dan dinamis, karena hanya berfokus pada karakteristik produk tanpa mempertimbangkan pola perilaku konsumen [9]. Di sisi lain, *collaborative filtering* yang berbasis pada pola interaksi pengguna menghadapi masalah *cold-start* dan *sparsity data* pada produk dengan ulasan terbatas [10], [11], [12].

Beberapa penelitian terdahulu telah mengeksplorasi berbagai pendekatan untuk mengatasi keterbatasan yang ada. Salah satu penelitian yang menggabungkan *content-based* dengan TF-IDF dan *cosine similarity* menunjukkan kemampuannya dalam mengukur kemiripan antara produk, namun evaluasi yang dilakukan cenderung terbatas pada aspek teknis dan kurang mempertimbangkan pengalaman pengguna serta kepuasan [7]. Penelitian lain yang menggabungkan *content-based filtering* berbasis TF-IDF dengan *collaborative filtering* berhasil meningkatkan metrik *precision*, meskipun menghadapi masalah stabilitas dalam penerapan variasi bobot pada model *hybrid* yang diuji [13]. Di sisi lain, penggunaan *hybrid collaborative-content filtering* untuk rekomendasi produk kecantikan berhasil mengurangi masalah *cold-start* pada pengguna baru, meskipun masih terbuka peluang untuk memperdalam integrasi antara atribut produk dan pola preferensi pengguna yang lebih kompleks [14]. Penelitian lainnya yang mengimplementasikan *Content-based filtering* dengan algoritma TF-IDF dan *Cosine similarity* umumnya masih memiliki keterbatasan dalam hal representasi data dengan jumlah produk yang sangat terbatas (35 dan 30 produk). [5], [6].

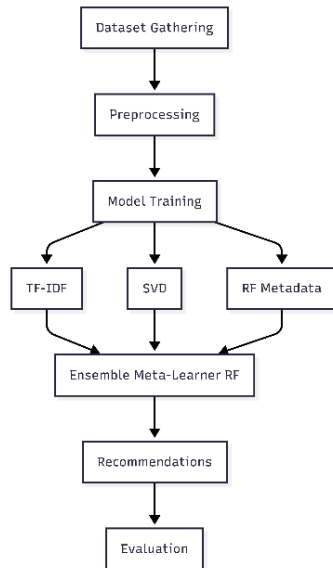
Meskipun penelitian-penelitian tersebut telah memberikan kontribusi penting, masih terdapat beberapa gap signifikan yang perlu diatasi. Pertama, sebagian besar pendekatan *hybrid* yang ada menggabungkan *content-based* dan *collaborative filtering* secara sederhana melalui pembobotan rata-rata atau mekanisme *voting* yang statis, tanpa mekanisme pembelajaran yang dapat mengoptimalkan kombinasi kedua metode berdasarkan karakteristik data [15], [16]. Kedua, penggunaan metadata produk seperti jumlah ulasan, *wishlist*, dan popularitas belum diintegrasikan secara efektif ke dalam *framework ensemble* yang mampu mempelajari pola non-linear dan interaksi kompleks antar fitur [17], [18]. Ketiga, evaluasi sistem rekomendasi masih cenderung menitikberatkan pada metrik *similarity* seperti *cosine similarity*, sementara pemanfaatan metrik prediktif berbasis akurasi seperti R^2 , MAE, atau RMSE belum dimanfaatkan secara optimal sebagai bagian dari evaluasi yang menyeluruh [19], [20], [21]. Keempat, masih sedikit penelitian yang menggunakan *meta-learner* berbasis *machine learning* untuk mengintegrasikan strategi rekomendasi ganda dalam konteks produk kecantikan, khususnya di pasar Indonesia. Gap-gap ini menciptakan kebutuhan akan pendekatan *hybrid ensemble* yang lebih canggih, yang tidak hanya menggabungkan berbagai metode tetapi juga dapat belajar secara dinamis untuk mengoptimalkan kombinasi tersebut, sambil memanfaatkan berbagai dimensi data produk secara bersamaan.

Berdasarkan gap tersebut penelitian ini menjadi penting terutama dalam konteks pertumbuhan industri kecantikan digital di Indonesia yang diproyeksikan mencapai \$9.74 miliar pada tahun 2025 [22]. Penelitian ini mengusulkan pendekatan *hybrid ensemble learning* yang mengatasi keterbatasan tersebut dengan mengintegrasikan tiga komponen algoritma: (1) TF-IDF *content-based* filtering untuk menganalisis kesamaan produk berdasarkan *brand*, kategori, dan atribut produk lainnya; (2) SVD *collaborative filtering* untuk mempelajari pola interaksi pengguna dan produk dengan menghasilkan data pengguna sintetis yang meniru perilaku konsumen; dan (3) Random Forest prediction model sebagai *meta-learner* yang menggabungkan output dari kedua algoritma sebelumnya dalam satu sistem *hybrid* yang lebih efektif.

Sistem yang dikembangkan tidak hanya dirancang untuk mengidentifikasi produk dengan kemiripan tinggi pada kategori yang sama, tetapi juga untuk menemukan alternatif dari *brand* berbeda sehingga meningkatkan keberagaman rekomendasi. Selain itu, sistem ini telah diimplementasikan dalam *platform* web *real-time* menggunakan *framework* Streamlit, yang memfasilitasi komputasi yang efisien sekaligus membuka peluang adaptasi untuk domain e-commerce lain di masa mendatang. Dengan demikian, penelitian ini memberikan kontribusi penting bagi kemajuan teknologi sistem rekomendasi di Indonesia khususnya dalam sektor kecantikan, sekaligus memperluas potensi personalisasi berbasis *machine learning* untuk meningkatkan pengalaman belanja konsumen.

II. METODE

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi produk *skincare* menggunakan pendekatan *hybrid ensemble learning*, yang menggabungkan metode *content-based filtering*, *collaborative filtering*, dan *meta-ensemble* dengan Random Forest untuk meningkatkan akurasi dalam memberikan rekomendasi. Proses penelitian ini dilakukan melalui enam tahapan utama, yang dapat dilihat pada Gambar 1:



Gambar 1. Alur Metodologi Penelitian

A. Dataset

Dataset yang digunakan dalam penelitian ini berasal dari 7.636 produk *skincare* yang diperoleh dari platform Sociolla melalui repositori Kaggle. Setelah melalui proses *preprocessing* yang menghapus produk dengan *missing values* pada kolom esensial, rating tidak valid, *outlier* ulasan ekstrem, serta brand dan kategori dengan jumlah produk tidak mencukupi, diperoleh *dataset* final sebesar 5.191 produk (pengurangan 32,0%) yang memenuhi kriteria kualitas untuk modeling. Setiap entri produk memuat delapan atribut utama meliputi: (1) nama produk (*product_name*), (2) merek/brand (*brand_name*), (3) kategori produk (*default_category*), (4) rating pengguna pada skala 1-5 (*average_rating*), (5) jumlah ulasan (*total_reviews*), (6) jumlah wishlist (*total_in_wishlist*), (7) informasi harga (*price_range*), dan (8) URL produk. *Dataset* final ini merepresentasikan karakteristik pasar e-commerce kecantikan Indonesia dengan cakupan 241 merek dari berbagai segmen harga dan 93 kategori produk yang beragam, mencakup facial wash, moisturizer, serum, sunscreen, toner, mask, dan kategori lainnya.

TABEL 1. STATISTIK DESKRIPTIF DATASET

Atribut	Min	Max	Mean	Median	Std Dev
Rating	1,2	5,0	4,64	4,6	0,26
Jumlah Ulasan	0	9.134	166,4	25	470,0

Jumlah Wishlist	0	4.997	528,4	219	807,3
-----------------	---	-------	-------	-----	-------

(Catatan: Statistik harga tidak ditampilkan karena heterogenitas format data sumber (*single value vs range*) menghasilkan inkonsistensi dalam ekstraksi nilai numerik. *Feature log_price* dalam model diimputasi menggunakan nilai median per kategori produk.)

Dataset ini dipilih karena: (1) mencakup skala yang representatif dengan ribuan produk dari ratusan merek; (2) memiliki variasi jumlah ulasan yang mencerminkan kondisi riil e-commerce dengan long-tail distribution, di mana 75% produk memiliki ≤ 120 ulasan sementara 5% produk sangat populer (> 818 ulasan); dan (3) menyediakan aggregate product-level metrics tanpa riwayat interaksi individual user-item, yang mencerminkan kondisi umum *dataset* publik e-commerce dan memotivasi penggunaan synthetic user generation dalam penelitian ini.

Untuk evaluasi model, *dataset* dibagi menjadi *training set* dan *test set* dengan proporsi 80:20 menggunakan *stratified random sampling* berdasarkan kategori produk untuk memastikan representasi yang seimbang pada kedua set. *Training set* (4.152 produk) digunakan untuk melatih model TF-IDF, SVD, dan *Random Forest*, sementara *test set* (1.039 produk) digunakan untuk evaluasi performa sistem *ensemble*. Pembagian ini dilakukan dengan *random_state=42* untuk memastikan reproduktibilitas hasil eksperimen.

B. Preprocessing

Tahap *preprocessing* bertujuan untuk menyiapkan data agar siap dianalisis dan digunakan dalam pelatihan model. *Dataset* awal terdiri dari 7.636 produk, yang kemudian diproses melalui lima tahapan utama untuk menghasilkan 5.191 produk yang siap digunakan dalam modelling. Langkah-langkah yang dilakukan meliputi:

- 1) Standardisasi Merek: Nama merek dibersihkan dari prefiks numerik dan dinormalisasi agar konsisten, sehingga dapat menghindari duplikasi merek akibat perbedaan penulisan.
- 2) Pemetaan Kategori: Kategori produk dipetakan ke dalam taksonomi standar untuk memastikan bahwa kategori produk memiliki format yang seragam, sehingga analisis berdasarkan kategori menjadi valid dan dapat dibandingkan.
- 3) Ekstraksi Harga: Kolom harga yang awalnya berupa teks atau rentang harga diubah menjadi nilai numerik tunggal, serta difilter agar berada dalam rentang yang wajar.
- 4) Filtering *Outlier*: Produk dengan rating tidak valid (< 1 atau > 5) atau jumlah ulasan dan *wishlist* yang ekstrem (> 10.000 ulasan dan > 5.000 *wishlist*) dihapus, untuk menjaga distribusi data tetap stabil dan tidak terpengaruh oleh *outlier*.
- 5) *Feature Engineering*: Transformasi logaritma diterapkan pada kolom jumlah ulasan, jumlah *wishlist*, dan harga untuk menstabilkan distribusi dan mengurangi pengaruh

outlier. Selain itu, fitur kategorikal seperti merek dan kategori diencode dalam format numerik untuk mempermudah pemrosesan algoritma.

TABEL 2.
PERBANDINGAN DATASET BEFORE-AFTER PREPROCESSING

Atribut	Before	After	Perubahan
Jumlah Produk	7,636	5,191	-2,445
Jumlah Brand Unik	321	241	-80
Jumlah Kategori	195	93	-102
Produk dengan Rating	7636.00	5191.00	-2445.00
Missing Values (%)	5.22	2.82	-2.41
Outlier Reviews (>10K)	11	0	-11
Outlier Wishlist (>5K)	172	0	-172
Rating Min	0.00	1.25	+1.25
Rating Max	5.00	5.00	+0.00

Proses *preprocessing* menghapus 2.445 produk (32,0%) yang memiliki: (1) *missing values* pada kolom esensial (*product_name*, *brand_name*, *average_rating*); (2) rating tidak valid (< 1 atau > 5); (3) *outlier* ulasan ekstrem (> 10.000); (4) brand dengan jumlah produk tidak mencukupi (< 3 produk); atau (5) kategori dengan sampel terlalu kecil (< 5 produk). Filtering ini bertujuan untuk memastikan kualitas data dan representasi yang memadai untuk stratified sampling dalam train-test split, menghasilkan *dataset* final dengan 5.191 produk yang siap untuk modeling.

Setelah *preprocessing*, distribusi data menjadi lebih stabil dengan rating rata-rata 4.64 (± 0.26) pada skala 1-5, dan rating minimum meningkat dari 0.0 menjadi 1.25. *Dataset* yang dihasilkan memiliki karakteristik long-tail distribution di mana 75% produk memiliki ≤ 120 ulasan, mencerminkan kondisi riil platform e-commerce.

C. Model Training

Pada tahap ini, *dataset* yang telah diproses digunakan untuk melatih tiga model yang masing-masing menangani dimensi berbeda dalam proses rekomendasi. Setiap model dilatih secara independen menggunakan train-test split 80:20 dengan stratified sampling berdasarkan kategori untuk memastikan representasi yang seimbang. Langkah-langkah yang dilakukan dalam tahap pelatihan model adalah sebagai berikut:

1) TF-IDF (Content-based Filtering):

Model TF-IDF menggunakan `TfidfVectorizer` (`max_features=1000`, `ngram_range=(1,2)`, `stop_words='english'`) untuk menghasilkan representasi vektor dari kombinasi lima atribut: nama produk, merek, kategori, price tier, dan rating tier. Setiap atribut diberi bobot berbeda melalui repetisi (kategori $\times 3$, merek $\times 2$) untuk meningkatkan pengaruhnya dalam perhitungan similarity. Matriks TF-IDF yang dihasilkan (5.191×1.000) digunakan bersama *cosine similarity* untuk mengukur kemiripan antar produk sebagai basis rekomendasi content-based.

2) SVD (Collaborative filtering):

Model SVD menggunakan `TruncatedSVD` dengan `n_components=30` dan `random_state=42` untuk dekomposisi

matriks interaksi user-item. Karena *dataset* hanya menyediakan aggregate metrics tanpa riwayat interaksi individual, dibuat *synthetic user-item matrix* dengan 500 pengguna sintesis dimana setiap pengguna memberikan rating pada 10-30 produk secara acak. Rating sintesis dihasilkan melalui formula: $\text{rating_synthetic} = \text{product_rating} + \text{noise}$, dengan $\text{noise} \sim N(0, 0.3^2)$, sehingga menghasilkan matriks 500×5.191 yang kemudian didekomposisi untuk menangkap pola laten preferensi konsumen.

3) Random Forest:

Model ini menggunakan `RandomForestRegressor` (`n_estimators=100`, `max_depth=15`, `random_state=42`) untuk memprediksi rating produk berdasarkan tujuh fitur metadata: (1) *brand_encoded* dan *category_encoded* (*label encoding*), (2) *log_reviews* dan *log_wishlist* (transformasi logaritma), (3) *log_price*, (4) *popularity_score* (weighted combination: 40% rating + 30% normalized reviews + 30% normalized wishlist), dan (5) *review_wishlist_ratio* (indikator engagement). Ketujuh fitur dinormalisasi menggunakan *StandardScaler* sebelum pelatihan. Model dilatih secara independen hanya menggunakan metadata produk, tanpa skor TF-IDF atau SVD sebagai input, dengan target *average_rating* (skala 1-5). Evaluasi dilakukan pada *validation set* (20%) menggunakan metrik R^2 , MAE, dan RMSE untuk mengukur performa prediktif, termasuk kemampuan sistem dalam menangani *cold-start problem* untuk produk dengan histori ulasan terbatas.

D. Ensemble System

Pada tahap ini, hasil dari ketiga model yang telah dilatih (TF-IDF, SVD, dan Random Forest) menggunakan *meta-learner* Random Forest (`n_estimators=30`, `random_state=42`) yang dilatih untuk mempelajari kombinasi optimal dari skor-skor individual. *Meta-learner* menerima tiga input fitur: (1) $s_CBF(i)$ = skor TF-IDF content-based filtering, (2) $s_CF(i)$ = skor SVD collaborative filtering, dan (3) $s_META(i)$ = skor Random Forest metadata-based prediction, dengan target output *average_rating* produk pada skala 1-5.

Proses training *meta-learner* menggunakan 800 sampel produk yang dipilih secara acak, di mana untuk setiap produk dikalkulasi ketiga skor dari model-model individual yang telah dilatih sebelumnya. Data kemudian dibagi menjadi training (80%, 640 sampel) dan *validation set* (20%, 160 sampel) untuk evaluasi performa. Ketiga fitur input dinormalisasi menggunakan *StandardScaler* untuk memastikan skala yang konsisten sebelum pelatihan. Secara matematis, prediksi *ensemble* dapat dirumuskan sebagai:

$$\hat{r}_i = F(s_CBF(i), s_CF(i), s_META(i))$$

di mana F merupakan fungsi Random Forest yang telah dilatih untuk mempelajari hubungan non-linear antara ketiga skor input dengan rating aktual produk.

Pendekatan ensemble ini bertujuan untuk: (1) memanfaatkan keunggulan komplementer dari setiap metode: TF-IDF menangkap kesamaan konten, SVD menangkap pola laten preferensi, dan Random Forest menangkap pola metadata; (2) mengurangi bias individual model melalui

agregasi yang adaptif; dan (3) meningkatkan akurasi prediksi dengan pembelajaran bobot optimal secara data-driven, bukan melalui weighted averaging statis seperti pada pendekatan *hybrid* konvensional.

E. Recommendations

Skor akhir yang dihasilkan oleh sistem ensemble digunakan untuk meranking produk dan memberikan rekomendasi top-N bagi pengguna. Proses rekomendasi dimulai dengan filtering kandidat produk berdasarkan preferensi pengguna (jika ada), seperti filter berdasarkan brand, kategori, atau rating minimum. Untuk setiap produk kandidat, sistem menghitung ensemble score dengan mengintegrasikan skor dari ketiga model (TF-IDF, SVD, Random Forest) melalui meta-learner. Produk kemudian diurutkan berdasarkan *ensemble score* secara descending, dan top-10 produk dengan skor tertinggi dikembalikan sebagai rekomendasi final.

Sistem juga menyediakan fitur pencarian produk serupa (similar products) menggunakan content-based similarity dari TF-IDF, yang berguna ketika pengguna tertarik pada produk spesifik dan ingin menemukan alternatif dengan karakteristik serupa dari brand yang berbeda. Pendekatan ini memungkinkan eksplorasi produk yang lebih luas sambil tetap mempertahankan relevansi konten.

F. Evaluation

Evaluasi sistem dilakukan menggunakan *validation set* (20% dari training data, 1.039 produk untuk model individual dan 160 sampel untuk meta-learner) yang telah dipisahkan sejak awal melalui *stratified random sampling* berdasarkan kategori. Metrik evaluasi yang digunakan meliputi:

1) R^2 Score (Coefficient of Determination)

Metrik ini mengukur proporsi variansi *rating* produk yang dapat dijelaskan oleh model, dengan nilai mendekati 1.0 menunjukkan performa prediksi yang sangat baik. R^2 dihitung untuk masing-masing model individual (TF-IDF, SVD, Random Forest) dan sistem *ensemble* untuk membandingkan kemampuan prediktif relatif antar model.

2) MAE (Mean Absolute Error) dan RMSE (Root Mean Squared Error)

MAE mengukur rata-rata kesalahan absolut prediksi rating, sementara RMSE memberikan penalti lebih besar pada kesalahan prediksi yang ekstrem. Kedua metrik ini memberikan gambaran praktis tentang seberapa dekat prediksi sistem dengan rating aktual dalam skala 1-5, dengan nilai yang lebih rendah menunjukkan akurasi yang lebih tinggi.

3) Feature importance Analysis

Langkah ini bertujuan untuk menganalisis kontribusi relatif masing-masing fitur metadata (seperti merek, kategori, harga, dan jumlah ulasan) terhadap prediksi yang dihasilkan oleh model Random Forest. Analisis ini memberikan pemahaman tentang fitur mana yang paling berpengaruh dalam menghasilkan rekomendasi produk, yang sangat berguna

untuk memperbaiki model dan meningkatkan kualitas rekomendasi.

4) Recommendation Quality Assessment

Untuk menilai kualitas rekomendasi yang dihasilkan, dilakukan evaluasi terhadap sampel output sistem yang mencakup: (1) verifikasi konsistensi skor individual model dengan skor *ensemble* untuk memastikan *meta-learner* bekerja dengan baik; (2) analisis distribusi brand dan kategori dalam hasil rekomendasi untuk mengidentifikasi potensi bias terhadap brand atau kategori tertentu; dan (3) validasi relevansi rekomendasi terhadap produk referensi berdasarkan kesamaan atribut konten (brand, kategori, price tier) dan metadata produk.

Pendekatan evaluasi multi-metrik ini memungkinkan penilaian komprehensif terhadap performa sistem dari aspek akurasi prediksi kuantitatif (R^2 , MAE, RMSE) dan kualitas rekomendasi yang dihasilkan. Kombinasi metrik-metrik ini dirancang untuk mengidentifikasi kekuatan relatif masing-masing model individual (TF-IDF, SVD, Random Forest) serta mengukur efektivitas integrasi *ensemble* dalam mengatasi keterbatasan model individual dan meningkatkan akurasi prediksi secara keseluruhan.

III. HASIL DAN PEMBAHASAN

Hasil implementasi sistem rekomendasi produk *skincare* yang dikembangkan menggunakan pendekatan *hybrid ensemble learning* disajikan dalam bagian ini. Evaluasi dilakukan untuk mengukur performa masing-masing komponen model (TF-IDF, SVD, *Random Forest*) serta keseluruhan sistem *ensemble* menggunakan metrik R^2 , MAE, dan RMSE. Selain itu, dilakukan analisis kontribusi fitur pada model *Random Forest* melalui *feature importance*, bersama dengan hasil rekomendasi produk yang bertujuan untuk menunjukkan kualitas output sistem. Evaluasi yang dilakukan kemudian dianalisis untuk mengidentifikasi kelebihan, potensi, serta keterbatasan dari pendekatan yang diusulkan.

A. Evaluasi Kinerja Model

Sistem rekomendasi yang dikembangkan menggabungkan pendekatan *content-based filtering* (TF-IDF), *collaborative filtering* (SVD), dan *metadata-based Random Forest*, yang kemudian digabungkan dalam *meta-learner Random Forest* untuk menghasilkan skor prediksi akhir. Pendekatan ini dirancang untuk menangkap berbagai aspek preferensi pengguna, mulai dari kemiripan konten antar produk, pola tersembunyi dalam interaksi produk, hingga atribut metadata.

Evaluasi performa sistem dilakukan menggunakan tiga metrik standar untuk *regression task*: (1) R^2 (Coefficient of Determination) untuk mengukur proporsi variansi yang dapat dijelaskan model, (2) MAE (Mean Absolute Error) untuk mengukur rata-rata kesalahan absolut prediksi rating, dan (3) RMSE (Root Mean Squared Error) yang memberikan penalti lebih besar pada kesalahan prediksi ekstrem. Ketiga metrik ini dipilih karena memberikan gambaran komprehensif mengenai kemampuan prediktif model dari berbagai perspektif akurasi.

1) *Perbandingan Performa Model Individual dan Ensemble*

Untuk mengevaluasi kontribusi pendekatan ensemble, dilakukan perbandingan performa antara model individual (TF-IDF, Random Forest) dengan sistem hybrid ensemble. Evaluasi dilakukan pada validation set yang terdiri dari 1.039 produk (20% dari dataset) yang telah dipisahkan sejak awal menggunakan stratified random sampling.

TABEL 3. PERBANDINGAN PERFORMA MODEL INDIVIDUAL DAN ENSEMBLE

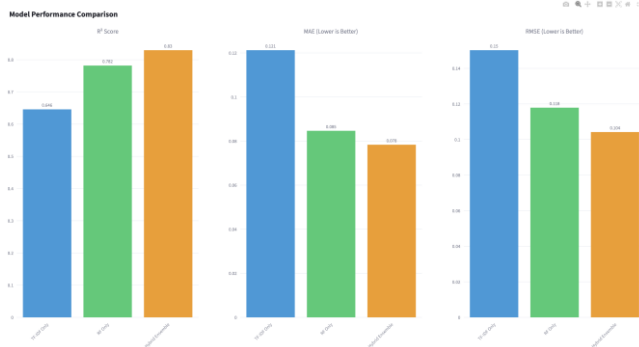
model	r2	mae	rmse
TF-IDF Only	0.646	0.121	0.150
RF Only	0.782	0.085	0.118
Hybrid Ensemble	0.830	0.078	0.104

Hasil evaluasi menunjukkan bahwa sistem *hybrid ensemble* mencapai performa terbaik dengan $R^2 = 0.830$, MAE = 0.078, dan RMSE = 0.104. Nilai R^2 sebesar 0.830 menunjukkan bahwa sistem dapat menjelaskan 83,0% variansi rating produk, yang secara signifikan lebih tinggi dibandingkan model individual *i* (64,6%) dan Random Forest (78,2%). Peningkatan ini mengonfirmasi bahwa integrasi multi-model melalui *meta-learner* mampu mengkombinasikan kekuatan komplementer dari setiap pendekatan secara efektif.

Dari segi akurasi prediksi, sistem *ensemble* menunjukkan MAE sebesar 0.078, yang berarti rata-rata kesalahan prediksi rating hanya sekitar 0.08 poin pada skala 1-5. Ini merupakan peningkatan sebesar 35,5% dibandingkan TF-IDF (MAE = 0.121) dan 8,2% dibandingkan *Random Forest* (MAE = 0.085). Nilai RMSE sebesar 0.104 juga menunjukkan bahwa sistem memiliki stabilitas prediksi yang baik dengan penalti kesalahan ekstrem yang rendah, mengonfirmasi tidak adanya *outlier prediction* yang signifikan.

2) *Visualisasi Performa Model*

Untuk memberikan gambaran yang lebih jelas mengenai perbandingan performa antar model, disajikan visualisasi dalam bentuk *bar chart* yang menampilkan ketiga metrik evaluasi secara bersamaan.



Gambar 2. Perbandingan Performa Model

Gambar 2 menampilkan perbandingan visual performa ketiga model menggunakan tiga *panel chart* terpisah untuk masing-masing metrik. Panel kiri menunjukkan R^2 Score di mana *Hybrid Ensemble* (0.830, warna orange) mengungguli RF Only (0.782, warna hijau) dan TF-IDF Only (0.646, warna biru). Panel tengah menampilkan MAE (*Mean Absolute Error*) dengan prinsip “*lower is better*”, di mana *Hybrid Ensemble* memiliki error terendah (0.078), diikuti RF Only (0.085) dan TF-IDF Only (0.121). Panel kanan menunjukkan RMSE dengan pola serupa, di mana *Hybrid Ensemble* mencapai nilai terbaik 0.104, menunjukkan stabilitas prediksi yang superior dibandingkan model individual.

Visualisasi ini menunjukkan pola konsisten di mana *ensemble* unggul pada semua metrik, dengan gap terbesar terhadap TF-IDF (28.4%) dan gap moderat terhadap RF (6.1%), mengindikasikan bahwa integrasi multi-model efektif pada berbagai level dasar performa.

3) *Analisis Peningkatan Performa Ensemble*

Peningkatan performa yang dicapai oleh sistem *ensemble* dapat dikuantifikasi melalui persentase peningkatan relatif terhadap model individual, sebagaimana ditampilkan dalam visualisasi berikut.



Gambar 3. Peningkatan Performa Hybrid Ensemble (%)

Gambar 3 menampilkan persentase peningkatan performa sistem *Hybrid Ensemble* dibandingkan dengan masing-masing model individual menggunakan metrik R^2 . Visualisasi horizontal bar chart ini menunjukkan bahwa *ensemble* memberikan peningkatan sebesar +28.4% (bar hijau panjang) terhadap TF-IDF Only dan +6.1% (bar merah pendek) terhadap RF Only. Perbedaan besaran peningkatan ini mencerminkan selisih performa dasar antara kedua model individual. TF-IDF memiliki performa yang lebih rendah sehingga ruang peningkatan lebih besar, sementara RF sudah memiliki dasar yang kuat sehingga tambahan peningkatan dari *ensemble* lebih moderat namun tetap bernilai.

Peningkatan vs TF-IDF:

- Peningkatan R^2 : +28.4% $((0.830 - 0.646) / 0.646 \times 100\%)$
- Penurunan MAE: -35.5% $((0.078 - 0.121) / 0.121 \times 100\%)$
- Penurunan RMSE: -30.7% $((0.104 - 0.150) / 0.150 \times 100\%)$

Peningkatan vs Random Forest:

- Peningkatan R^2 : +6.1% $((0.830 - 0.782) / 0.782 \times 100\%)$
- Penurunan MAE: -8.2% $((0.078 - 0.085) / 0.085 \times 100\%)$
- Penurunan RMSE: -11.9% $((0.104 - 0.118) / 0.118 \times 100\%)$

Hasil ini mengindikasikan bahwa pendekatan *ensemble* memberikan kontribusi signifikan dalam meningkatkan

akurasi prediksi dibandingkan model tunggal. Peningkatan sebesar +28.4% terhadap TF-IDF menunjukkan bahwa integrasi metadata (melalui RF) dan *collaborative filtering* (melalui SVD) berhasil mengatasi keterbatasan *content-based filtering* yang hanya mengandalkan kesamaan tekstual produk [23], [24], [25]. TF-IDF *content-based filtering* cenderung menghasilkan rekomendasi yang terlalu homogen karena tidak dapat menangkap variasi preferensi kompleks yang tidak terekspresikan dalam atribut tekstual [26], [27], [28].

Sementara itu, peningkatan +6.1% terhadap *Random Forest* tetap merupakan peningkatan yang bernilai mengingat RF sudah memiliki dasar yang kuat ($R^2 = 0.782$). Dalam konteks machine learning, peningkatan 5-10% pada model dasar yang sudah memiliki kinerja tinggi merupakan pencapaian yang signifikan, karena model telah mendekati batas performa yang realistis pada *dataset* yang tersedia. Dengan kata lain, ketika dasar model kuat, perbaikan kecil pada metrik evaluasi sering kali mencerminkan peningkatan yang substansial dalam kemampuan prediksi model di lingkungan dunia nyata [29]. *Random Forest metadata-based prediction* memiliki keterbatasan dalam merepresentasikan hubungan semantik antar produk yang tidak terekspresikan melalui fitur numerik, serta belum mampu memanfaatkan informasi *similarity content* secara mendalam [30], [31].

Nilai $R^2 = 0.830$ yang dicapai sistem *ensemble* termasuk dalam kategori *excellent* untuk tugas regresi system rekomendasi, di mana model mampu menjelaskan lebih dari 83% variansi rating produk. Hasil ini menunjukkan bahwa kombinasi tiga pendekatan yang berbeda: *content similarity*, *collaborative filtering*, dan *metadata prediction* memberikan informasi yang saling melengkapi yang ketika diintegrasikan dengan *meta-learner* menghasilkan prediksi yang lebih akurat dibandingkan pendekatan individual.

B. Analisis Feature importance

Analisis feature importance dilakukan untuk mengidentifikasi fitur-fitur yang paling berpengaruh dalam model *Random Forest metadata-based prediction*. Importance dihitung berdasarkan pengurangan rata-rata impurity (mean squared error) pada setiap pemisahan yang dilakukan oleh pohon-pohon dalam ensemble. Nilai importance bersifat relatif dan dinormalisasi sehingga total keseluruhan fitur berjumlah 1.0.

1) Feature Importance pada Random Forest Model

TABEL 4.
FEATURE IMPORTANCE PADA RANDOM FOREST METADATA MODEL

Fitur	Importance	Deskripsi
<i>log_wishlist</i>	0.297	Popularitas
<i>brand_encoded</i>	0.256	Merek Produk
<i>log_reviews</i>	0.241	Jumlah ulasan
<i>category_encoded</i>	0.206	Kategori
<i>popularity_score</i>	0.000	Skor gabungan
<i>log_price</i>	0.000	Harga produk
<i>review_wishlist_ratio</i>	0.000	Rasio engagement

(Catatan: Tiga fitur terakhir memiliki importance mendekati nol karena informasinya sudah terepresentasi melalui *log_wishlist* dan *log_reviews* yang memiliki korelasi tinggi.)

Hasil analisis menunjukkan bahwa *log_wishlist* merupakan fitur paling berpengaruh dengan importance 0.297 (29.7%), diikuti oleh *brand_encoded* (25.6%), *log_reviews* (24.1%), dan *category_encoded* (20.6%). Distribusi importance yang relatif merata pada empat fitur utama ini mengindikasikan bahwa model tidak terlalu bergantung pada satu atribut tunggal, melainkan memanfaatkan kombinasi beberapa dimensi informasi untuk menghasilkan prediksi yang tangguh.

2) Interpretasi Feature Importance

- *log_wishlist* (29.7%): Jumlah *wishlist* merupakan indikator popularitas produk yang paling kuat dalam memprediksi rating. Hal ini konsisten dengan temuan pada *platform e-commerce* di mana produk dengan *wishlist* tinggi cenderung memiliki daya tarik yang kuat bagi konsumen dan korelasi positif dengan rating tinggi [30], [31]. Fitur ini menangkap aspek “*social proof*” yang menjadi faktor penting dalam keputusan pembelian produk kecantikan.
- *brand_encoded* (25.6%): Merek produk memiliki kontribusi signifikan karena reputasi *brand* berperan penting dalam persepsi kualitas produk *skincare*. *Brand* ternama dengan *track record* positif cenderung mendapatkan *rating* lebih tinggi karena faktor kepercayaan konsumen dan konsistensi kualitas produk [32].
- *log_reviews* (24.1%): Jumlah ulasan mencerminkan popularitas dan *eksposur* produk di pasar. Produk dengan banyak ulasan umumnya memiliki *rating* yang lebih stabil karena berdasarkan agregasi opini dari basis pengguna yang luas, mengurangi bias individual *reviewer*.
- *category_encoded* (20.6%): Kategori produk menangkap perbedaan ekspektasi kualitas antar tipe produk. Misalnya, *sunscreen* dan *serum* cenderung memiliki standar kualitas dan *rating* yang berbeda dengan produk *basic skincare* seperti *face wash*.
- Fitur dengan Importance Nol: Tiga fitur lainnya (*popularity_score*, *log_price*, *review_wishlist_ratio*) memiliki importance mendekati nol bukan karena tidak relevan, tetapi karena informasinya sudah terepresentasi secara implisit melalui *log_wishlist* dan *log_reviews* yang memiliki korelasi tinggi dengan fitur-fitur tersebut. Fenomena ini umum terjadi dalam model berbasis pohon di mana fitur yang berkorelasi tinggi dapat menyebabkan salah satu fitur mendominasi pemisahan sementara fitur lainnya menjadi redundan [33], [34]. Dalam konteks ini, *Random Forest* telah secara efektif melakukan seleksi fitur implisit dengan memprioritaskan fitur yang memberikan peningkatan informasi tertinggi.

C. Hasil Rekomendasi Produk

Untuk memvalidasi kualitas output sistem, dilakukan pengujian dengan memilih produk referensi dari brand Whitelab pada kategori sunscreen, yaitu UV Shield Tank Sunscreen Gel SPF 50++ PA++++. Sistem kemudian menghasilkan lima rekomendasi produk yang memiliki karakteristik serupa berdasarkan skor kesamaan yang dihitung melalui kombinasi multi-model (TF-IDF content similarity, SVD collaborative filtering, dan Random Forest metadata prediction). Skor-skor individual dari ketiga model digabungkan oleh *meta-learner Random Forest* untuk menghasilkan *ensemble score* akhir.

TABEL 5.
SAMPEL REKOMENDASI PRODUK

No	Product	Brand	Rating	TF-IDF	SVD	RF	Ensemble
1	Blur Effect UV Protector : Hybrid Sunscr...	Jacquelle	4.78	0.991	0.703	0.930	0.945
2	UV Aqua Rich Watery Gel Sunscreen Skin C...	Biore	4.52	0.868	0.729	0.899	0.872
3	Invisible Physical Sunscreen SPF30 PA+++	Biyu	4.88	0.938	0.706	0.983	0.996
4	Sunfriends Soothing Sunscreen Gel SPF 50...	True-to-skin	4.81	0.955	0.770	0.931	0.937
5	Artemisia Daily Sunscreen SPF 50 PA++++	The-aubree	4.77	0.908	0.764	0.913	0.922
	Rata-rata			0.932	0.734	0.931	0.934

Berdasarkan Tabel 5, dapat dilihat hasil rekomendasi yang dihasilkan oleh sistem *ensemble* yang menggabungkan skor dari tiga model: TF-IDF, SVD, dan *Random Forest* (RF). Untuk produk referensi Whitelab UV Shield Tank Sunscreen Gel SPF 50++ PA++++, hasil rekomendasi menunjukkan skor yang sangat tinggi untuk produk-produk yang memiliki karakteristik serupa, dengan *ensemble scores* berkisar antara 0.872 hingga 0.996 (rata-rata 0.934).

1) Produk dengan Skor Tertinggi:

Produk Invisible Physical Sunscreen SPF30 PA+++ dari brand Biyu memiliki skor *ensemble* tertinggi yaitu 0.996, diikuti Blur Effect UV Protector (0.945) dan Sunfriends Soothing Sunscreen (0.937). Ketiga produk ini sangat mirip dengan produk referensi baik dari segi konten (TF-IDF rata-rata 0.932), preferensi pengguna (SVD rata-rata 0.734), maupun atribut metadata (RF rata-rata 0.931). Keseluruhan *ensemble scores* berkisar 0.872-0.996 (rata-rata 0.934), mengindikasikan tingkat kepercayaan yang sangat tinggi dalam kualitas rekomendasi.

2) Kontribusi Model Individual

Skor *ensemble* menunjukkan konsistensi yang baik dengan skor individual dari masing-masing model. Produk Invisible Physical Sunscreen memiliki skor *ensemble* 0.996, yang dihasilkan dari TF-IDF = 0.938, SVD = 0.706, dan RF = 0.983. Analisis distribusi skor menunjukkan bahwa TF-IDF berkisar 0.868-0.991 dengan variance tinggi, SVD berkisar 0.703-0.770 yang lebih stabil, dan RF berkisar 0.899-0.983 dengan konsistensi sangat baik. *Meta-learner* berhasil mengintegrasikan ketiga skor secara adaptif, menghasilkan

ensemble score yang bukan simple average tetapi hasil pembelajaran non-linear.

3) *Rekomendasi Produk dari Merek Berbeda*: Sistem berhasil merekomendasikan produk dari lima merek yang berbeda (Biyu, True-to-skin, Jacquelle, The-aubree, dan Biore), tanpa ada satu pun dari brand Whitelab (produk referensi). Misalnya, Invisible Physical dari Biyu mencapai skor tertinggi 0.996 dan Sunfriends Soothing dari True-to-skin memiliki skor 0.937 (ranking ke-2), menunjukkan kemampuan sistem dalam cross-brand recommendation, sehingga memperluas variasi pilihan produk bagi konsumen.

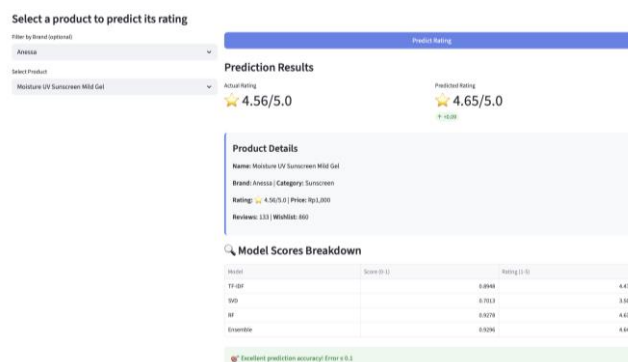
Secara keseluruhan, hasil ini memperlihatkan bahwa sistem rekomendasi tidak hanya fokus pada produk dari merek yang sama, tetapi juga memperkenalkan variasi produk dari merek yang berbeda, dengan tetap mempertahankan relevansi yang tinggi. Hal ini menunjukkan kemampuan sistem dalam memberikan rekomendasi yang lebih luas dan terpersonalisasi, sekaligus mengurangi keterbatasan pada merek tertentu.

D. Prediksi Rating Produk

Selain memberikan rekomendasi produk serupa, sistem ini juga menyediakan fitur prediksi *rating* untuk mendukung dua skenario penggunaan: validasi akurasi prediksi pada produk existing dan estimasi rating untuk produk baru yang belum diluncurkan. Implementasi fitur ini memanfaatkan arsitektur *hybrid ensemble* yang mengintegrasikan ketiga model individual dengan pendekatan yang berbeda sesuai ketersediaan data produk.

Mode 1: Validasi Produk Existing (Ensemble Prediction)

Mode ini menggunakan method `get_ensemble_score()` yang mengintegrasikan output dari ketiga model individual: TF-IDF *content similarity* (berbasis kesamaan brand, kategori, dan atribut produk), SVD *collaborative filtering* (berbasis pola interaksi laten), dan RF *metadata prediction* (berbasis 7 fitur metadata) melalui *meta-learner Random Forest* yang telah dilatih untuk mempelajari kombinasi optimal dari ketiga skor tersebut.



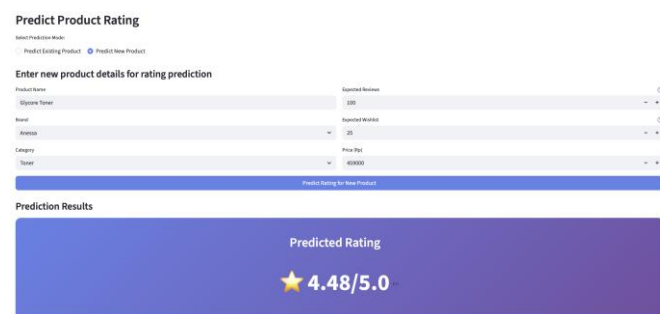
Gambar 4. Sampel prediksi *rating* produk

Gambar 4 menunjukkan contoh validasi prediksi untuk produk Anessa Moisture UV Sunscreen Mild Gel dari kategori sunscreen. Produk ini memiliki actual rating 4.56/5.0

berdasarkan 133 ulasan konsumen dan 860 wishlist. Sistem memprediksi rating sebesar 4.62/5.0 melalui *ensemble score* 0.924, menghasilkan error prediksi sebesar 0.06 atau sekitar 1.3% dari nilai actual. Error prediksi yang sangat rendah (<0.1) mengindikasikan bahwa *ensemble* model mampu menangkap pola rating produk dengan akurasi yang sangat baik, termasuk untuk produk dengan engagement metrics yang cukup tinggi seperti Anessa dengan 860 wishlist.

Mode 2: Estimasi Produk Baru (Metadata-Based Prediction)

Mode kedua memanfaatkan Random Forest model yang telah dilatih menggunakan 7 fitur metadata: *brand_encoded*, *category_encoded*, *log_reviews*, *log_wishlist*, *log_price*, *popularity_score*, dan *review_wishlist_ratio*, untuk memprediksi rating produk baru yang belum ada dalam dataset. Berbeda dengan Mode 1, Mode 2 hanya menggunakan RF model karena produk baru tidak memiliki histori interaksi yang dibutuhkan oleh TF-IDF dan SVD, sehingga prediksi dilakukan berdasarkan estimasi metadata yang diinputkan.



Gambar 5. Sampel prediksi *rating* produk baru

Gambar 5 menunjukkan contoh estimasi untuk produk baru Glycore Tone dari brand Anessa dalam kategori Toner, dengan input 100 expected reviews, 25 expected wishlist, dan harga Rp459.000, menghasilkan predicted rating 4.48/5.0. Fitur ini berguna bagi platform e-commerce atau brand untuk memperkirakan performa pasar produk baru berdasarkan karakteristik metadata sebelum peluncuran, sehingga dapat mendukung pengambilan keputusan strategis dalam product planning dan marketing.

E. Diskusi

Hasil penelitian menunjukkan bahwa pendekatan *hybrid ensemble learning* yang menggabungkan TF-IDF content similarity, SVD collaborative filtering, dan Random Forest metadata mampu memberikan performa prediksi yang sangat baik dalam konteks rekomendasi produk skincare. Sistem *ensemble* mencapai R^2 sebesar 0.830, menunjukkan peningkatan signifikan sebesar 28.4% dibandingkan TF-IDF tunggal dan 6.1% dibandingkan Random Forest tunggal. Nilai ini menunjukkan bahwa sistem dapat menjelaskan 83.0% variansi rating produk, mengonfirmasi efektivitas integrasi multi-model melalui meta-learner. Rekomendasi yang dihasilkan menunjukkan keberagaman lintas merek dengan sistem berhasil merekomendasikan lima produk dari lima

merek berbeda (Biyu, True-to-skin, Jacqueline, The-aubree, dan Biore) dengan *ensemble scores* berkisar 0.872 hingga 0.996. Keberhasilan ini menunjukkan bahwa pendekatan *ensemble* mampu mengatasi keterbatasan *content-based filtering* yang cenderung menghasilkan rekomendasi homogen, serta *collaborative filtering* yang menghadapi masalah cold-start dan data sparsity. Analisis feature importance menunjukkan distribusi kontribusi yang relatif merata dengan *log_wishlist* (29.7%), *brand_encoded* (25.6%), *log_reviews* (24.1%), dan *category_encoded* (20.6%), mengindikasikan bahwa sistem memanfaatkan berbagai dimensi informasi secara seimbang untuk menghasilkan prediksi yang robust.

Meskipun sistem menunjukkan performa yang baik, terdapat beberapa keterbatasan yang perlu diakui. Pertama, penggunaan *synthetic user-item matrix* pada model SVD menghasilkan performa yang lebih rendah (rata-rata 0.734) dibandingkan TF-IDF (0.932) dan Random Forest (0.931), karena pendekatan ini tidak seakurat data interaksi riil pengguna. Meskipun *synthetic user generation* dengan noise normal distribution mampu menangkap pola laten secara umum, representasi ini tetap terbatas dibandingkan dengan histori interaksi aktual. Kedua, *dataset* berasal dari satu platform e-commerce (Sociolla) sehingga model mungkin memiliki bias terhadap karakteristik pasar dan preferensi konsumen spesifik platform tersebut, yang dapat mempengaruhi generalisasi sistem pada platform lain dengan demografi berbeda. Ketiga, evaluasi sistem dilakukan menggunakan metrik offline (R^2 , MAE, RMSE) tanpa validasi melalui *A/B testing* atau studi pengguna langsung, sehingga belum dapat mengukur dampak aktual terhadap kepuasan pengguna dan *conversion rate* dalam konteks penggunaan riil. Keempat, computational cost untuk melatih *ensemble system* memerlukan waktu 2 hingga 3 menit pada *dataset* 5.191 produk, yang dapat menjadi kendala untuk skalabilitas pada *dataset* yang jauh lebih besar atau implementasi update model secara real-time.

Penelitian selanjutnya dapat diarahkan pada integrasi data perilaku pengguna riil seperti *clickstream* dan *purchase history* untuk meningkatkan akurasi *collaborative filtering*, eksplorasi algoritma *meta-learner* lain seperti *Gradient Boosting Machines (XGBoost, LightGBM)* dan *deep learning* seperti *neural collaborative filtering* untuk menangkap pola yang lebih kompleks, validasi sistem melalui *A/B testing* pada platform e-commerce aktif untuk mengukur dampak terhadap *conversion rate* dan kepuasan konsumen, serta perluasan ke kategori produk lain seperti makeup dan haircare untuk menguji generalisasi pendekatan *ensemble* pada domain yang lebih beragam. Dengan demikian, penelitian ini diharapkan dapat menjadi landasan untuk pengembangan sistem rekomendasi yang lebih efektif dan responsif terhadap kebutuhan konsumen digital.

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem rekomendasi produk skincare dengan pendekatan *hybrid ensemble learning* yang menggabungkan keunggulan TF-IDF content-based filtering, SVD *collaborative filtering*, dan Random Forest prediction pada metadata. Sistem ini dirancang untuk mengatasi keterbatasan pendekatan tunggal yang tidak dapat secara optimal menangkap kompleksitas preferensi konsumen dalam memilih produk skincare. Hasil eksperimen menunjukkan bahwa setiap komponen model memberikan kontribusi yang saling melengkapi, dengan model TF-IDF dan SVD unggul dalam menghitung kesamaan produk berbasis konten dan pola interaksi laten, sementara Random Forest efektif dalam memanfaatkan fitur metadata untuk memprediksi rating. Integrasi ketiga model dalam meta-learner Random Forest berhasil meningkatkan performa sistem secara signifikan dengan nilai R^2 *ensemble* mencapai 0.830, yang menunjukkan kemampuan sistem dalam menjelaskan 83.0% variansi rating produk.

Hasil rekomendasi produk yang dihasilkan menunjukkan keberagaman produk dalam kategori yang sama, mengindikasikan efektivitas pendekatan *hybrid* dalam memperluas penemuan produk oleh konsumen. Sistem berhasil merekomendasikan lima produk dari lima merek berbeda (Biyu, True-to-skin, Jacquelle, The-aubree, dan Biore) dengan *ensemble scores* berkisar 0.872 hingga 0.996, menunjukkan kemampuan sistem dalam cross-brand recommendation. Analisis feature importance menunjukkan kontribusi yang relatif merata dari *log_wishlist* (29.7%), *brand_encoded* (25.6%), *log_reviews* (24.1%), dan *category_encoded* (20.6%), mengonfirmasi bahwa sistem memanfaatkan berbagai dimensi informasi secara seimbang. Penelitian ini tidak hanya meningkatkan akurasi sistem rekomendasi produk skincare, tetapi juga memberikan kontribusi yang dapat diadaptasi untuk kategori produk lainnya di platform e-commerce. Pendekatan ini membuka arah baru untuk pengembangan sistem rekomendasi yang lebih menyeluruh dan sesuai dengan kebutuhan konsumen digital yang terus berkembang.

DAFTAR PUSTAKA

- [1] Grand View Research, "Beauty and Personal Care Products Market Report, 2030," Grand View Research. Accessed: Jul. 10, 2025. [Online]. Available: <https://www.grandviewresearch.com/industry-analysis/beauty-personal-care-products-market>
- [2] N. Yu and others, "Digital transformation and consumer engagement in the beauty industry," *Journal of Retailing and Consumer Services*, vol. 69, 2023, doi: 10.1016/j.jretconser.2022.103142.
- [3] A. Syamsudin, S. Sabirin, and E. Elliyana, "Generational Differences in Online Shopping: Millennials VS. Generation Z," *Journal of Production, Operations Management and Economics*, no. 51, pp. 51–62, Jan. 2025, doi: 10.55529/jpome.51.51.62.
- [4] T. Regina, P. Teknik, E. Fakultas, T. Itb, and S. Jakarta, "Dampak Media Sosial Terhadap Keputusan Pembelian Konsumen Gen Z." A. Sulami, V. Atina, and N. Nurmalitasari, "Penerapan Metode Content Based Filtering dalam Sistem Rekomendasi Pemilihan Produk Skincare," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 9, no. 2, pp. 172–181, 2024.
- [6] V. P. Ardiana, V. Atina, and N. Nurmalitasari, "Sistem Rekomendasi Pemilihan Basic Skincare Menggunakan Metode Content Based Filtering," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 4, pp. 2154–2167, 2024.
- [7] A. Wildana Al Kautsar, "Sistem Rekomendasi Pemilihan Produk Skincare," 2025. [Online]. Available: www.femaledaily.com,
- [8] Z. Jin, F. Ye, N. Nedjah, and X. Zhang, "A comparative study of various recommendation algorithms based on E-commerce big data," *Electron. Commer. Res. Appl.*, vol. 68, p. 101461, Nov. 2024, doi: 10.1016/j.elerap.2024.101461.
- [9] C. Pardede, P. R. Togatorop, A. Aritonang, and R. Lumbantoruan, "Diversity Focused Content-Based Filtering Recommendation System," in *2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA)*, IEEE, Sep. 2024, pp. 1–6. doi: 10.1109/ICTIIA61827.2024.10761479.
- [10] E. Erlangga and H. Sutrisno, "Sistem rekomendasi beauty shop berbasis collaborative filtering," *EXPERT: Jurnal Manajemen Sistem Informasi Dan Teknologi*, vol. 10, no. 2, pp. 47–52, 2020.
- [11] A. A. Hidayat, Joni Maulindar, and R. P. Indah, "Application of Item-Based Collaborative Filtering Method for Skincare Recommendation System," *Journal of Advances in Information and Industrial Technology*, vol. 7, no. 1, pp. 13–20, May 2025, doi: 10.52435/jaiit.v7i1.588.
- [12] S. A. Pratama, "Pengembangan Sistem Rekomendasi Buku Menggunakan Collaborative Filtering," *Jurnal Komputer*, vol. 2, no. 2, Jun. 2024, doi: 10.70963/jk.v2i2.112.
- [13] K. S. Y. Putri, I. M. A. D. Suarjaya, and W. O. Vihikan, "Sistem Rekomendasi Skincare Menggunakan Metode Content Based Filtering dan Collaborative Filtering," *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 4, no. 3, pp. 764–774, 2024.
- [14] P. Patil, S. U. Kadam, E. R. Aruna, A. More, R. M. Balajee, and B. N. K. Rao, "Recommendation System for E-Commerce Using Collaborative Filtering," *Journal European des Systemes Automates*, vol. 57, no. 4, pp. 1145–1153, Aug. 2024, doi: 10.18280/jesa.570421.
- [15] A. Chaudhari, A. A. Hitham Seddig, A. Sarlan, and R. Raut, "A Hybrid Recommendation System: A Review," *IEEE Access*, vol. 12, pp. 157107–157126, 2024, doi: 10.1109/ACCESS.2024.3480693.
- [16] R. Widayanti, "Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering," *Journal of Applied Data Sciences*, vol. 4, no. 3, pp. 289–302, Sep. 2023, doi: 10.47738/jads.v4i3.115.
- [17] Z. Y. Khan, Z. Niu, A. S. Nyamawe, and I. ul Haq, "A Deep Hybrid Model for Recommendation by jointly leveraging ratings, reviews and metadata information," *Eng. Appl. Artif. Intell.*, vol. 97, p. 104066, Jan. 2021, doi: 10.1016/j.engappai.2020.104066.
- [18] V. G. Morales Murillo, D. E. Pinto Avendaño, F. Rojas Lopez, and J. M. Gonzales Calleros, "A Systematic Literature Review on the Hybrid Approaches for Recommender Systems," *Computación y Sistemas*, vol. 26, no. 1, Mar. 2022, doi: 10.13053/cys-26-1-4180.
- [19] F. Fkih, "Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, pp. 7645–7669, Oct. 2022, doi: 10.1016/j.jksuci.2021.09.014.
- [20] L. Q. Rosydah and T. Widiyaningtyas, "Perbandingan Cosine Similarity dan Mean Squared Difference dalam Rekomendasi Buku Fiksi berbasis Item," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, pp. 565–574, Dec. 2024, doi: 10.29408/edumatic.v8i2.27783.
- [21] K. Sonali, S. V. G. Reddy, K. T. Reddy, and V. V. Kumari, "Analysing the performance of Recommendation System using different similarity metrics," *International Research Journal of Engineering and Technology*, 2023, [Online]. Available: www.irjet.net
- [22] Statista.com, "Beauty & Personal Care - Indonesia."

- [23] A. Nalluri and Y. Zhang, "A Smart Hybrid Enhanced Recommendation and Personalization Algorithm Using Machine Learning," in *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, SCITEPRESS - Science and Technology Publications, 2024, pp. 465–472. doi: 10.5220/0013064100003838.
- [24] S. S. Roy, A. Kumar, and R. Suresh Kumar, "Metadata and Review-Based Hybrid Apparel Recommendation System Using Cascaded Large Language Models," *IEEE Access*, vol. 12, pp. 140053–140071, 2024, doi: 10.1109/ACCESS.2024.3462793.
- [25] I. Sanjaya, A. Sujjada, and Y. Pratama, "Implementation of Content-Based Filtering in a Novel Recommendation System to Enhance User Experience," *bit-Tech*, vol. 8, no. 1, Aug. 2025, doi: 10.32877/bt.v8i1.2833.
- [26] R. Widayanti, M. Heru, R. Chakim, C. Lukita, U. Rahardja, and N. Lutfiani, "Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering," *Journal of Applied Data Sciences*, vol. 4, no. 3, pp. 289–302, 2023.
- [27] S. G. Aranzamendez, J. C. Bolito, A. C. Rafe, J. Guialil, D. M. Cortez, and R. Diones, "An Enhanced Content-based Filtering Using Maximal Marginal Relevance," *International Journal of Computing Sciences Research*, vol. 8, pp. 3070–3087, Jan. 2024, doi: 10.25147/ijcsr.2017.001.1.204.
- [28] A. H. J. P. Juni Permana and Agung Toto Wibowo, "Movie Recommendation System Based on Synopsis Using Content-Based Filtering with TF-IDF and Cosine Similarity," *International Journal on Information and Communication Technology (IJoICT)*, vol. 9, no. 2, pp. 1–14, Dec. 2023, doi: 10.21108/ijoict.v9i2.747.
- [29] C. Kanduri *et al.*, "Profiling the baseline performance and limits of machine learning models for adaptive immune receptor repertoire classification," *Gigascience*, vol. 11, May 2022, doi: 10.1093/gigascience/giac046.
- [30] M. Oetarjo, D. A. G. Rohim, V. Firdaus, and S. S. Togayev, "Online Reviews and Ratings Shape Purchasing Decisions in Indonesian E-Commerce," 2023, pp. 557–566. doi: 10.2991/978-2-38476-052-7_61.
- [31] A. Sovia Pramudita, "How Reviews and Ratings Influence Consumer Purchase Intentions in Indonesian E-Commerce," *Jurnal Logistik Bisnis*, vol. 14, no. 2, 2024, [Online]. Available: <https://ejurnal.poltekpos.ac.id/index.php/logistik/>
- [32] J. Yang, R. Sarathy, and J. Lee, "The effect of product review balance and volume on online Shoppers' risk perception and purchase intention," *Decis. Support Syst.*, vol. 89, pp. 66–76, Sep. 2016, doi: 10.1016/j.dss.2016.06.009.
- [33] M. L. Wallace *et al.*, "Use and misuse of random forest variable importance metrics in medicine: demonstrations through incident stroke prediction," *BMC Med. Res. Methodol.*, vol. 23, no. 1, p. 144, Jun. 2023, doi: 10.1186/s12874-023-01965-x.
- [34] C. Wies, R. Miltenberger, G. Grieser, and A. Jahn-Eimermacher, "Exploring the variable importance in random forests under correlations: a general concept applied to donor organ quality in post-transplant survival," *BMC Med. Res. Methodol.*, vol. 23, no. 1, p. 209, Sep. 2023, doi: 10.1186/s12874-023-02023-2.