

ANALISIS KESESUAIAN PRODUK SKINCARE BERDASARKAN KOMPOSISI BAHAN MENGGUNAKAN METODE JACCARD SIMILARITY

Nicholas Leonardo^{1*}, Ahmad Zidane Arrasyid², Muhammad Arif Billah Natagama³, Risqi Adriyanto Adi Putro⁴, Vitri Tundjungsari⁵

^{1,2,3,4,5}Prodi Teknik Informatika, Universitas Esa Unggul, Jl. Harapan Indah Boulevard No.2, Pusaka Rakyat, Kec. Tarumajaya, Kabupaten Bekasi, Jawa Barat 17214

Keywords:

Sistem Rekomendasi;

Skincare;

Jaccard Similarity;

Content-Based Filtering;

Corespondent Email:

nicholasmgl678@student.esaunggul.ac.id

Abstrak. Pemilihan produk skincare yang sesuai kebutuhan kulit merupakan permasalahan yang sering dihadapi oleh pengguna karena banyaknya variasi produk dengan komposisi bahan yang beragam. Kesalahan dalam memilih produk dapat menimbulkan efek yang tidak diinginkan pada kulit. Oleh karena itu, diperlukan sebuah sistem rekomendasi yang mampu membantu pengguna dalam menentukan produk skincare berdasarkan kesamaan kandungan bahan. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem rekomendasi produk skincare menggunakan metode *Jaccard Similarity*. Dataset yang digunakan berasal dari Kaggle yang berisi informasi produk skincare beserta komposisi bahan penyusunnya. Proses penelitian meliputi tahap pengumpulan data, *preprocessing* data, tokenisasi bahan, serta perhitungan tingkat kemiripan antar produk menggunakan metode Jaccard Similarity. Implementasi sistem dilakukan menggunakan bahasa pemrograman Python pada platform Google Colab. Hasil penelitian menunjukkan bahwa sistem mampu memberikan rekomendasi produk skincare yang relevan berdasarkan tingkat kesamaan komposisi bahan, sehingga dapat membantu pengguna dalam memilih produk yang sesuai. Sistem ini diharapkan dapat menjadi solusi awal dalam pengembangan sistem rekomendasi skincare berbasis konten.



Copyright © **JITET** (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstract. Choosing skincare products that suit individual skin needs is a common problem faced by users due to the wide variety of products with diverse ingredient compositions. Choosing the wrong product can cause unwanted effects on the skin. Therefore, a recommendation system is necessary to assist users in selecting skincare products based on ingredient similarities. This study aims to design and implement a skincare product recommendation system using the Jaccard Similarity method. The dataset used comes from Kaggle, which contains information on skincare products and their ingredient compositions. The research process includes data collection, data preprocessing, ingredient tokenization, and calculation of the similarity level between products using the Jaccard Similarity method. The system was implemented using the Python programming language on the Google Colab platform. The study's results demonstrate that the system is capable of providing relevant skincare product recommendations based on the similarity of ingredient compositions, thereby helping users select suitable products. This system is expected to be an initial solution in the development of content-based skincare recommendation systems.

1. PENDAHULUAN

Perkembangan industri kecantikan, khususnya produk skincare, mengalami peningkatan yang sangat pesat dalam beberapa tahun terakhir, didorong oleh tren global dan masuknya berbagai merek internasional seperti produk *skincare* asal Korea Selatan [1]. Beragam merek dan jenis produk skincare dengan komposisi bahan yang berbeda-beda tersedia di pasaran, sehingga memberikan banyak pilihan bagi konsumen. Namun, kondisi ini juga menimbulkan permasalahan baru, yaitu kesulitan pengguna untuk menentukan produk skincare yang sesuai dengan kebutuhan dan kondisi kulitnya [2]. Pemilihan produk yang tidak tepat dapat menyebabkan iritasi, alergi, atau ketidakefektifan perawatan kulit. Pemilihan produk yang tidak tepat dapat menyebabkan iritasi, alergi, atau ketidakefektifan perawatan kulit, yang pada akhirnya berdampak buruk bagi kesehatan wajah [3].

Pada pemilihan produk skincare, kandungan bahan (ingredients) merupakan faktor terpenting yang perlu diperhatikan. Setiap bahan memiliki fungsi dan karakteristik tertentu, serta bisa menimbulkan efek yang berbeda pada setiap jenis kulit. Oleh karena itu, analisis kesamaan komposisi bahan antar produk dapat dimanfaatkan untuk membantu pengguna dalam menemukan produk skincare yang memiliki karakteristik serupa dengan produk yang telah cocok digunakan sebelumnya.

Salah satu pendekatan yang dapat digunakan untuk menyelesaikan permasalahan tersebut adalah sistem rekomendasi berbasis konten CBF (*content-based filtering*). Sistem ini menganalisis kesamaan fitur suatu item, tanpa bergantung pada data preferensi pengguna lain. Metode *Jaccard Similarity* merupakan salah satu teknik pengukuran kemiripan yang sederhana dan efektif untuk menghitung tingkat kesamaan antara dua himpunan data, sehingga cocok digunakan dalam membandingkan komposisi bahan pada produk skincare.

Beberapa penelitian sebelumnya telah menerapkan sistem rekomendasi pada berbagai domain menggunakan metode kesamaan berbasis fitur. Namun, pemanfaatan metode *Jaccard Similarity* secara khusus untuk menganalisis kesamaan kandungan bahan produk skincare masih relatif terbatas, terutama

dengan menggunakan dataset publik yang dapat direproduksi. Hal ini menunjukkan adanya peluang penelitian untuk mengembangkan sistem rekomendasi skincare yang berfokus pada kesamaan komposisi bahan sebagai dasar rekomendasi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem rekomendasi produk skincare menggunakan metode *Jaccard Similarity*. Dataset yang digunakan berasal dari Kaggle dan berisi informasi produk skincare beserta komposisi bahan penyusunnya. Implementasi sistem dilakukan menggunakan bahasa pemrograman Python pada platform Google Colab. Diharapkan hasil penelitian ini bisa membantu pengguna untuk memilih produk skincare yang sesuai berdasarkan kesamaan kandungan bahan, serta menjadi dasar pengembangan sistem rekomendasi skincare yang lebih lanjut.

2. TINJAUAN PUSTAKA

2.1 Sistem Rekomendasi

Sistem rekomendasi merupakan sebuah sistem yang dirancang untuk membantu pengguna dalam menemukan informasi atau produk yang sesuai dengan kebutuhan dan preferensinya. Sistem ini bekerja dengan menganalisis data dan memberikan saran berdasarkan pola tertentu yang diperoleh dari data tersebut. Sistem rekomendasi banyak diterapkan pada berbagai bidang, seperti e-commerce, hiburan, dan layanan digital, untuk meningkatkan kualitas pengambilan keputusan pengguna. Secara fundamental, sistem ini merupakan salah satu penerapan dari *Machine Learning* yang bekerja dengan mempelajari pola data untuk memprediksi preferensi pengguna terhadap suatu item [4].

2.2 Content-Based Filtering

Content-Based Filtering merupakan salah satu pendekatan dalam sistem rekomendasi yang berfokus pada karakteristik atau fitur dari suatu item. Rekomendasi diberikan berdasarkan tingkat kesamaan antara item yang pernah digunakan atau dipilih pengguna dengan item lainnya. Pendekatan ini tidak bergantung pada data pengguna lain, sehingga cocok digunakan pada sistem dengan keterbatasan data preferensi pengguna. Fleksibilitas metode ini juga terbukti dalam domain lain, seperti penelitian [5] yang

memanfaatkannya untuk merekomendasikan pekerjaan IT berdasarkan kesamaan *skill* dan deskripsi lowongan. Selain itu, Fajriansyah [6] juga menerapkan metode ini untuk memberikan rekomendasi film yang personal bagi pengguna berdasarkan kemiripan *genre* dan sinopsis cerita.

2.3 Produk Skincare dan Komposisi Bahan

Produk skincare merupakan produk perawatan kulit yang diformulasikan dari berbagai bahan dengan fungsi tertentu, seperti melembapkan, melindungi, atau mengatasi permasalahan kulit. Sistem rekomendasi berperan penting dalam memberikan pengetahuan kepada pengguna untuk memilih produk perawatan wajah yang tepat, sehingga dapat meningkatkan efektivitas pemilihan produk sesuai kondisi kulit [7]. Komposisi bahan pada produk skincare memiliki peran penting dalam menentukan efektivitas dan keamanan produk. Analisis terhadap kandungan bahan dapat digunakan sebagai dasar dalam membandingkan dan merekomendasikan produk skincare yang memiliki karakteristik serupa.

2.4 Metode *Jaccard Similarity*

Jaccard Similarity merupakan metode pengukuran kemiripan antara dua himpunan data dengan cara membandingkan jumlah irisan elemen terhadap jumlah gabungan elemen dari kedua himpunan tersebut. Nilai Jaccard Similarity berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 itu tingkat kemiripan yang tinggi. Metode ini banyak digunakan dalam sistem rekomendasi dan *information retrieval*, khususnya untuk data berbentuk teks atau atribut diskrit, sehingga sesuai untuk membandingkan komposisi bahan pada produk skincare. Validitas metode ini dalam mengukur kemiripan himpunan data juga dibuktikan [8], yang berhasil menerapkannya untuk optimasi pencocokan profil kompetensi pada sistem rekrutmen karyawan. Selain itu, Desena dan Solichin [9] juga membuktikan efektivitas algoritma ini dalam sistem temu kembali informasi (*information retrieval*) untuk mendeteksi kemiripan dokumen teks pada abstrak tugas akhir mahasiswa.

2.5 Penelitian Terkait

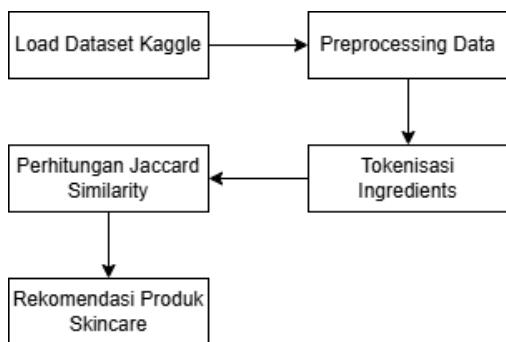
Beberapa penelitian sebelumnya menunjukkan bahwa metode analisis berbasis data dan fitur dapat digunakan secara efektif dalam berbagai domain sistem, salah satunya adalah sektor perbankan [10]. Selain itu, penelitian yang dilakukan oleh [11] menerapkan metode *Content-Based Filtering* menggunakan algoritma TF-IDF dan *Cosine Similarity* untuk merekomendasikan produk *skincare* berdasarkan deskripsi produk. Sementara itu,[12] mengembangkan sistem rekomendasi *skincare* menggunakan pendekatan *Hybrid Filtering* yang mengombinasikan *Content-Based Filtering* dan *Collaborative Filtering* untuk peningkatan akurasi rekomendasi berdasarkan preferensi pengguna lain. Pendekatan *content-based filtering* dengan metode pengukuran kemiripan, seperti *Jaccard Similarity*, mampu memberikan rekomendasi yang relevan berdasarkan karakteristik item. Namun, penerapan metode ini pada domain produk skincare dengan fokus pada kesamaan komposisi bahan masih terbatas, sehingga diperlukan penelitian lebih lanjut untuk mengembangkan sistem rekomendasi yang lebih spesifik dan terarah pada kebutuhan pengguna.

3. METODE PENELITIAN

Penelitian ini bertujuan untuk membangun sistem rekomendasi produk skincare berdasarkan kesamaan komposisi bahan menggunakan metode *Jaccard Similarity*. Metode penelitian disusun secara sistematis agar dapat direplikasi oleh peneliti lain. Proses penelitian meliputi perancangan sistem, pengumpulan data, *preprocessing* data, perhitungan tingkat kemiripan, serta penyajian rekomendasi produk *skincare*.

3.1 Rancangan Penelitian

Penelitian ini menggunakan pendekatan *eksperimental* dengan metode *content-based filtering*. Sistem rekomendasi dibangun dengan menganalisis kesamaan komposisi bahan antar produk skincare tanpa melibatkan data preferensi pengguna lain. Metode *Jaccard Similarity* digunakan sebagai teknik utama untuk mengukur tingkat kemiripan antar produk yang terlihat pada gambar 1.



Gambar 1. Diagram Alur Penelitian

3.2 Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari platform Kaggle. Dataset yang digunakan berjudul “Skincare Product Ingredients” yang disusun oleh dominoweir.

Dataset ini berisi informasi mengenai nama produk skincare, merek, kategori, serta daftar bahan (ingredients) yang terkandung dalam setiap produk. Data digunakan hanya untuk keperluan analisis dan penelitian akademik

3.3 Preprocessing Data

Tahap *preprocessing* data bertujuan untuk menyiapkan data agar dapat digunakan dalam proses analisis. Tahapan *preprocessing* yang dilakukan meliputi:

1. **Pembersihan Data**, yaitu menghapus data yang tidak lengkap atau tidak relevan.
2. **Normalisasi Teks**, dengan mengubah seluruh huruf menjadi huruf kecil.
3. **Tokenisasi Ingredients**, yaitu memisahkan komposisi bahan menjadi elemen bahan individual.
4. **Penghapusan Duplikasi**, sehingga setiap produk direpresentasikan oleh himpunan bahan yang unik.

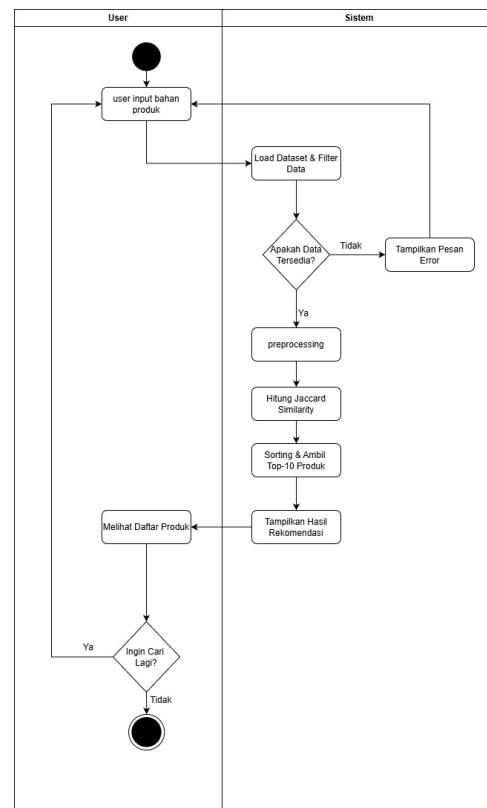
3.4 Metode Jaccard Similarity

Metode *Jaccard Similarity* digunakan untuk menghitung tingkat kemiripan antara dua produk *skincare* berdasarkan kesamaan komposisi bahan. Pendekatan perhitungan berbasis himpunan ini telah terbukti efektif dalam menangani perbandingan struktur data yang kompleks, seperti yang diterapkan Aqimuddin [13]. Perhitungan dilakukan dengan membandingkan jumlah irisan bahan terhadap

jumlah gabungan bahan dari dua produk yang dibandingkan.

3.5 Implementasi Sistem

Implementasi sistem pada penelitian ini dilakukan menggunakan *platform* Google Colaboratory (Google Colab) dengan bahasa pemrograman Python. Sistem yang dibangun bertujuan untuk memberikan rekomendasi produk skincare berdasarkan tingkat kemiripan bahan (*ingredients*) antara input pengguna dan data produk pada dataset menggunakan metode Jaccard Similarity. Alur implementasi sistem mengikuti *activity diagram* yang ditunjukkan pada Gambar 2.

Gambar 2. *Activity diagram* sistem rekomendasi produk *skincare* berbasis *Jaccard Similarity*

3.5.1 Input Pengguna

Proses diawali dengan pengguna memasukkan daftar bahan skincare yang diinginkan atau ingin dihindari dalam bentuk teks. Input ini berfungsi sebagai *query* yang akan dibandingkan dengan data bahan pada setiap produk skincare di dalam dataset.

3.5.2 Pemuatan Dataset dan Validasi Data

Sistem memuat dataset produk skincare yang diperoleh dari Kaggle. Dataset ini berisi informasi nama produk dan daftar bahan (*ingredients*). Setelah dataset dimuat, sistem melakukan pengecekan ketersediaan dan kelengkapan data. Apabila data tidak tersedia atau terjadi kesalahan saat pemuatan, sistem akan menampilkan pesan kesalahan (*error*). Jika data tersedia, proses dilanjutkan ke tahap berikutnya.

3.5.3 Preprocessing Data

ada tahap ini, sistem melakukan *preprocessing* terhadap input pengguna dan data bahan produk. Proses *preprocessing* meliputi:

1. Konversi teks ke huruf kecil (*lowercase*),
2. Penghapusan tanda baca dan karakter khusus,
3. Pemisahan bahan menjadi token (*tokenization*),
4. Penghapusan spasi berlebih dan duplikasi kata.

Tahap ini bertujuan untuk menyamakan format data agar perhitungan kemiripan dapat dilakukan secara optimal.

3.5.4 Perhitungan Jaccard Similarity

Setelah proses *preprocessing*, sistem menghitung nilai *Jaccard Similarity* antara bahan input pengguna dan bahan pada setiap produk skincare. Perhitungan dilakukan dengan menghitung irisan (*intersection*) dan gabungan (*union*) dari dua himpunan bahan menggunakan persamaan berikut:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

di mana A merupakan himpunan bahan input pengguna dan B merupakan himpunan bahan pada produk skincare.

Proses ini dilakukan secara berulang (*looping*) untuk seluruh produk yang terdapat pada dataset.

3.5.5 Penyaringan dan Pengurutan Produk

Nilai Jaccard Similarity yang dihasilkan kemudian digunakan untuk melakukan penyaringan (*filtering*) dan pengurutan (*sorting*) produk skincare. Produk dengan nilai

kemiripan tertinggi diprioritaskan sebagai rekomendasi. Sistem mengambil sejumlah produk dengan skor tertinggi (misalnya *Top-N*) untuk ditampilkan kepada pengguna.

3.5.6 Penyajian Hasil Rekomendasi

Hasil akhir berupa daftar produk skincare yang memiliki tingkat kemiripan bahan tertinggi dengan input pengguna. Informasi yang ditampilkan meliputi nama produk dan nilai skor Jaccard Similarity. Pengguna dapat melihat hasil rekomendasi dan memilih untuk melakukan pencarian ulang dengan input bahan yang berbeda atau mengakhiri proses rekomendasi.

4. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi sistem rekomendasi produk skincare berbasis **Jaccard Similarity** menggunakan dataset *Skincare Product Ingredients* dari Kaggle. Pembahasan difokuskan pada hasil pemrosesan data, perhitungan tingkat kemiripan bahan, serta analisis hasil rekomendasi yang dihasilkan oleh sistem.

4.1 Karakteristik Dataset

Dataset yang digunakan dalam penelitian ini diperoleh dari Kaggle dengan judul *Skincare Product Ingredients*. Dataset berhasil dimuat dan terdiri dari 1.472 data produk skincare. Setiap data memiliki beberapa atribut, yaitu label, merek (*brand*), nama produk (*name*), harga (*price*), peringkat (*rank*), serta daftar bahan (*ingredients*). Selain itu, dataset juga menyediakan informasi kecocokan produk terhadap beberapa jenis kulit.

Dalam penelitian ini, atribut *ingredients* digunakan sebagai fokus utama karena menjadi dasar dalam proses perhitungan tingkat kemiripan antar produk skincare yang terlihat pada Tabel 1.

Tabel 1. Struktur atribut dataset skincare

No	Atribut	Deskripsi
1	brand	Merek produk skincare
2	name	Nama produk

3	ingredients	Daftar bahan produk
4	price	Harga produk
5	rank	Peringkat produk
6	Combination	Kesesuaian kulit kombinasi
7	Dry	Kesesuaian kulit kering
8	Normal	Kesesuaian kulit normal
9	Oily	Kesesuaian kulit berminyak
10	Sensitive	Kesesuaian kulit sensitif

4.1.1 Identifikasi Atribut Dataset

Berdasarkan hasil pemrosesan awal dataset, sistem secara otomatis mengidentifikasi atribut-atribut utama yang digunakan dalam penelitian.

```
--- HASIL DETEKSI KOLOM ---
Kolom Ingredients pakai: 'ingredients'
Kolom Produk pakai : 'name'
Kolom Brand pakai : 'brand'
```

Gambar 3. Hasil identifikasi atribut dataset

Pada gambar 3 menunjukkan hasil identifikasi menunjukkan bahwa kolom ingredients digunakan sebagai sumber data bahan skincare, kolom name digunakan sebagai nama produk, dan kolom brand digunakan sebagai informasi merek produk.

Identifikasi atribut ini memastikan bahwa data yang digunakan dalam proses selanjutnya telah sesuai dan konsisten, sehingga meminimalkan kesalahan dalam tahap preprocessing dan perhitungan kemiripan bahan.

4.2 Preprocessing Data

Tahap preprocessing data dilakukan untuk mempersiapkan data *ingredients* sebelum digunakan dalam proses perhitungan tingkat kemiripan produk skincare menggunakan metode Jaccard Similarity. Dataset yang digunakan masih berupa data mentah (*raw data*) dalam bentuk teks panjang yang memiliki

variasi penulisan huruf, tanda baca, serta pemisahan bahan yang belum seragam, sehingga tidak dapat langsung diproses oleh algoritma.

Pada tahap awal, data *ingredients* ditampilkan dalam bentuk teks mentah (*ingredients_raw*) untuk menunjukkan kondisi data sebelum dilakukan preprocessing. Selanjutnya, dilakukan proses pembersihan data dengan mengubah seluruh teks menjadi huruf kecil (*case folding*), memisahkan setiap bahan berdasarkan tanda koma, serta menghapus spasi berlebih pada setiap elemen. Proses ini bertujuan untuk menyeragamkan format penulisan bahan agar tidak terjadi perbedaan representasi pada bahan yang sama.

Hasil preprocessing kemudian direpresentasikan dalam bentuk himpunan (*set*) yang disimpan pada kolom *processed_set*. Representasi data dalam bentuk himpunan bertujuan untuk menghilangkan duplikasi bahan serta mempermudah proses perhitungan irisan dan gabungan data yang diperlukan dalam metode Jaccard Similarity. Dengan menggunakan himpunan bahan, setiap produk skincare memiliki daftar *ingredients* yang unik dan terstruktur.

Tabel 2. Contoh Data Ingredients Sebelum dan Sesudah Preprocessing

No	Nama Produk	Ingredients (Sebelum Preprocessing)	Ingredients (Sesudah Preprocessing)
1	Crème de la Mer	Algae (Seaweed) Extract, Mineral Oil, Petrolatum, Glycerin, Isohexadecane, ...	{algae extract, mineral oil, petrolatum, glycerin, iso hexadecane, ...}
2	Facial Treatment Essence	Galactomyces Ferment Filtrate (Pitera), Butylene Glycol, Pentyleneglycol, water, ...	{galactomyces ferment filtrate, butylene glycol, pentyleneglycol, water, ...}

		Glycol, Water, ...	
3	Protini™ Polypepti de Cream	Water, Dicaprylyl Carbonate, Glycerin, Cetearyl Alcohol, Cetearyl Olivate, ...	{water, dicaprylyl carbonate, glycerin, cetearyl alcohol, cetearyl olivate, ...}

Tabel 2 menampilkan contoh data *ingredients* sebelum dan sesudah preprocessing. Berdasarkan tabel tersebut, dapat dilihat bahwa sebelum preprocessing data masih berupa teks panjang yang tidak terstruktur, sedangkan setelah preprocessing data bahan menjadi lebih rapi, konsisten, dan siap digunakan dalam proses analisis kemiripan produk.

Hasil preprocessing ini memastikan bahwa data *ingredients* telah berada dalam format yang sesuai untuk tahap selanjutnya, yaitu perhitungan nilai Jaccard Similarity dalam sistem rekomendasi produk skincare.

4.3 Perhitungan Jaccard Similarity

Setelah data *ingredients* melalui tahap preprocessing, sistem melakukan perhitungan tingkat kemiripan antara bahan yang dimasukkan oleh pengguna dengan komposisi *ingredients* pada setiap produk skincare menggunakan metode Jaccard Similarity. Tahap ini bertujuan untuk mengukur tingkat kesamaan bahan antara preferensi pengguna dan produk yang tersedia dalam dataset, sehingga sistem dapat memberikan rekomendasi yang relevan.

Pada pengujian yang dilakukan, pengguna memasukkan bahan Niacinamide sebagai input. Sistem kemudian menghitung nilai Jaccard Similarity antara himpunan bahan input pengguna dan himpunan *ingredients* dari seluruh produk skincare dalam dataset. Produk dengan nilai kemiripan tertinggi dipilih sebagai hasil rekomendasi utama.

Tabel 3 menyajikan contoh lima produk skincare dengan nilai Jaccard Similarity tertinggi berdasarkan input pengguna.

Tabel 3. Contoh Hasil Rekomendasi Produk Berdasarkan Jaccard Similarity

No	Nama Produk	Brand	Jaccard Similarity (%)	Bahan Cocok
1	Brightening Derm Revival Mask	SK-II	6,7	Niacinamide
2	20% Glycolic Solutions Jumbo Peel Swab	Peter Thomas Roth	6,7	Niacinamide
3	Essential Power Skin Toner	Laneige	5,9	Niacinamide
4	Lift & Plump Sheet Mask	Lancaster	5,0	Niacinamide
5	Beauty Shroom™ Exfoliating Acid Potion	Moon Juice	5,0	Niacinamide

Berdasarkan hasil perhitungan tersebut, sistem berhasil menghasilkan lima produk teratas dengan nilai Jaccard Similarity berkisar antara 5,0% hingga 6,7%. Produk dengan nilai tertinggi menunjukkan tingkat kesesuaian bahan yang paling mendekati input pengguna, yaitu Niacinamide, sehingga dapat direkomendasikan sebagai pilihan utama.

4.4 Pembahasan

Hasil perhitungan Jaccard Similarity pada tahap sebelumnya menunjukkan bahwa sistem rekomendasi skincare mampu mengidentifikasi produk-produk yang memiliki kesesuaian bahan dengan preferensi pengguna. Pada pengujian dengan input bahan Niacinamide, sistem berhasil menghasilkan lima produk teratas dengan nilai Jaccard Similarity berkisar

antara 5,0% hingga 6,7%. Nilai ini mencerminkan tingkat kemiripan antara himpunan bahan yang dimasukkan pengguna dengan komposisi *ingredients* pada masing-masing produk.

Nilai *Jaccard Similarity* yang relatif kecil disebabkan oleh karakteristik data *ingredients skincare* yang umumnya memiliki jumlah bahan cukup banyak. Dalam perhitungan *Jaccard Similarity*, nilai kemiripan dipengaruhi oleh perbandingan antara jumlah irisan (*intersection*) dan jumlah gabungan (*union*) dari dua himpunan. Meskipun suatu produk mengandung bahan yang sesuai dengan input pengguna, keberadaan banyak bahan lain dalam produk menyebabkan nilai gabungan meningkat sehingga persentase kemiripan menjadi lebih kecil. Namun demikian, nilai tersebut tetap relevan karena menunjukkan keberadaan bahan utama yang dicari oleh pengguna.

Pemilihan metode *Jaccard Similarity* dalam penelitian ini didasarkan pada karakteristik data komposisi bahan yang bersifat himpunan biner (*binary set*), di mana keberadaan suatu bahan lebih penting daripada frekuensi kemunculannya. Hal ini berbeda dengan pendekatan yang dilakukan [14] dalam penelitiannya mengenai sistem rekomendasi lagu Indonesia. Dalam penelitian tersebut, metode *Cosine Similarity* digunakan bersama dengan pembobotan TF-IDF karena data lirik lagu memiliki karakteristik frekuensi kata yang memengaruhi kemiripan antar dokumen. Sementara itu, dalam konteks *skincare*, penggunaan *Cosine Similarity* kurang relevan jika dataset tidak menyertakan informasi konsentrasi (%) bahan, sehingga metode *Jaccard* yang menghitung rasio irisan himpunan dinilai lebih representatif untuk mendeteksi kesamaan profil bahan kimia secara langsung dan transparan.

Metode *Jaccard Similarity* terbukti efektif dalam konteks sistem rekomendasi berbasis konten CBF (*Content-Based Filtering*), khususnya untuk data *ingredients* yang direpresentasikan dalam bentuk himpunan. Sejalan dengan penelitian terbaru oleh Ni Putu [15], yang mengembangkan sistem rekomendasi skincare berbasis ekstraksi kata kunci (*keyword extraction*) dan berhasil mencapai tingkat kepuasan pengguna sebesar 91,9%. Pendekatan ini memungkinkan sistem

untuk memberikan rekomendasi tanpa bergantung pada riwayat penilaian atau preferensi pengguna lain, sehingga cocok diterapkan pada kondisi *cold-start* atau pengguna baru. Selain itu, penggunaan himpunan *ingredients* hasil *preprocessing* membantu mengurangi pengaruh duplikasi dan perbedaan penulisan bahan yang sama.

Jika dibandingkan dengan pendekatan berbasis rating atau kolaboratif, metode *Jaccard Similarity* memiliki keunggulan dari sisi transparansi hasil. Sistem dapat menampilkan bahan apa saja yang cocok antara input pengguna dan produk, sehingga meningkatkan kepercayaan pengguna terhadap rekomendasi yang diberikan. Namun, metode ini juga memiliki keterbatasan, yaitu belum mempertimbangkan bobot atau tingkat kepentingan masing-masing bahan dalam suatu produk. Secara keseluruhan, hasil penelitian ini mengindikasikan bahwa metode *Jaccard Similarity* dapat digunakan sebagai pendekatan dasar yang efektif dalam pengembangan sistem rekomendasi produk *skincare*.

5. KESIMPULAN

Penelitian ini berhasil merancang dan mengimplementasikan sistem rekomendasi produk *skincare* berbasis *Content-Based Filtering* menggunakan metode *Jaccard Similarity*. Berdasarkan hasil pengujian, sistem mampu mengidentifikasi produk yang relevan dengan input bahan pengguna, menghasilkan nilai similaritas berkisar antara 5,0% hingga 6,7%. Meskipun nilai kemiripan cenderung rendah akibat banyaknya variasi komposisi bahan yang memperbesar nilai gabungan himpunan, sistem ini memiliki keunggulan utama dalam mengatasi masalah *cold-start* bagi pengguna baru serta menawarkan transparansi hasil dengan menampilkan irisan bahan yang cocok secara eksplisit. Namun, metode ini memiliki keterbatasan karena menggunakan pendekatan biner yang belum mempertimbangkan bobot prioritas atau konsentrasi bahan dalam suatu produk. Oleh karena itu, pengembangan selanjutnya disarankan untuk mengintegrasikan teknik pembobotan fitur (seperti TF-IDF) atau menggabungkannya dengan metode kolaboratif (*Hybrid Recommender System*) guna meningkatkan akurasi dan kualitas rekomendasi yang dihasilkan.

UCAPAN TERIMA KASIH

Penulis mengucapkan puji syukur kepada Tuhan Yang Maha Esa atas rahmat-Nya sehingga penelitian ini dapat diselesaikan. Penulis menyampaikan terima kasih kepada Program Studi Teknik Informatika, Universitas Esa Unggul yang telah memberikan dukungan fasilitas dan akademik. Ucapan terima kasih secara khusus ditujukan kepada Ibu Vitri Tundjungsari atas bimbingan dan arahannya dalam penyusunan naskah ini. Selain itu, penulis mengapresiasi kontributor dataset pada platform Kaggle yang telah menyediakan data publik sehingga memungkinkan terlaksananya penelitian ini.

DAFTAR PUSTAKA

- [1] E. Ayuningrum, Y. Azhar, and G. I. Marthasari, "Sistem Rekomendasi Produk Skincare Korea Berbasis Web Menggunakan Metode Collaborative Filtering," *J. Repos.*, vol. 4, no. 4, pp. 497–506, 2022.
- [2] V. No, J. Hal, N. Azizah, and A. Fauzan, "Sistem Rekomendasi Produk Somethinc Menggunakan Metode Content-based Filtering," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 3, pp. 461–468, 2024.
- [3] D. Ayu *et al.*, "Sistem Rekomendasi Skincare Menggunakan Metode Content-Based Filtering dan Algoritma Apriori," *Semin. Nas. Inov. Teknol.*, pp. 242–248, 2021.
- [4] V. Tundjungsari, *Dasar Machine Learning*. Yogyakarta: Deepublish, 2024.
- [5] C. Koloman, R. Maulana, R. Dwi, Z. Putri, and W. Abadi, "Sistem Rekomendasi Pekerjaan di bidang IT Menggunakan Algoritma Content-Based Filtering," *J. Creat. Student Res.*, vol. 1, no. 6, 2023.
- [6] M. Fajriansyah, P. P. Adikara, and A. W. Widodo, "Sistem Rekomendasi Film Menggunakan Content Based Filtering," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 6, 2021.
- [7] A. A. Firdausi, D. Hartanti, and A. A. Sari, "MENGGUNAKAN METODE CONTENT BASED FILTERING PADA PRODUK," *J. Sist. Inf. DAN Tek. Komput.*, vol. 10, no. 2, pp. 232–237, 2025.
- [8] I. M. Siregar, D. Pratama, and C. Himawan, "Penggunaan Jaccard Similarity Coefficient dalam Optimasi Proses Rekrutmen Karyawan Berbasis Profil dan Kompetensi," *Sci. Inf. Technol.*, vol. 7, no. 2, pp. 101–111, 2024.
- [9] W. Desena and A. Solichin, "Pencarian Abstrak Tugas Akhir Mahasiswa Berdasarkan Tingkat Kemiripan Menggunakan Algoritma Winnowing dan Jaccard Similarity pada Universitas Budi Luhur," *J. Inform.*, vol. 4221, pp. 112–122, 2021.
- [10] V. Tundjungsari *et al.*, "SENTIMEN ANALISIS PADA APLIKASI E-BRANCH BCA MENGGUNAKAN METODE NAÏVE BAYES," *IKRA-ITH Teknol. J. Sains dan Teknol.*, vol. 8, no. 1, pp. 78–87, 2024.
- [11] A. Sulami *et al.*, "PENERAPAN METODE CONTENT BASED FILTERING DALAM SISTEM REKOMENDASI PEMILIHAN PRODUK SKINCARE," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 9, no. 2, 2024.
- [12] K. Sri, Y. Putri, I. M. Agus, D. Suarjaya, and W. O. Vihikan, "Sistem Rekomendasi Skincare Menggunakan Metode Content Based Filtering dan Collaborative Filtering," *J. Pendidik. Teknol. Inf.*, vol. 4, no. 3, pp. 764–774, 2024.
- [13] I. M. Aqimuddin, S. M. Pamungkas, C. Gunawan, and M. A. Yaqin, "Analisis Kemiripan Model Proses Bisnis PMBoK dan Scrum menggunakan Metode Jaccard Coefficient Similarity dan Semantic Similarity," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 5, no. 2, pp. 53–64, 2023.
- [14] A. Y. Timur and A. N. Rohman, "SISTEM REKOMENDASI LAGU INDONESIA MENGGUNAKAN METODE CONTENT-BASED FILTERING DAN COSINE SIMILARITY," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 13, no. 1, pp. 1415–1423, 2025.
- [15] N. Putu, A. Puspita, D. Purnami, S. Putri, and I. N. P. Trisna, "Fine-Tuned Transformer Models for Keyword Extraction in Skincare Recommendation Systems," *J. Appl. Informatics Comput.*, vol. 9, no. 3, pp. 966–972, 2025.