

A Hybrid Rule-Based and Euclidean Distance Approach for Analyzing the Relationship Between Skincare Active Ingredients and Skin Types

Pendekatan Hibrida *Rule-Based* dan Jarak *Euclidean* dalam Menganalisis Hubungan Bahan Aktif *Skincare* dan Jenis Kulit

¹Indra Purnama, ^{2,*}Yeni Setiani, ³Habibah Mahmud, ⁴Witari Aryunani

¹²³⁴Prodi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma, Depok, Indonesia

*Korespondensi: jeni_setiani@staff.gunadarma.ac.id

<https://doi.org/10.59696/nexcore.v1i2.283>

Submitted: Apr 15, 2026 | **Accepted:** Mei 07, 2026 | **Published:** Mei 11, 2026

ABSTRACT

The growing demand for skincare products tailored to individual skin types has highlighted the need for data-driven approaches that effectively associate active ingredients with specific skin conditions. However, objective and systematic guidance for matching skincare ingredients to different skin types remains limited. This study aims to analyze the relationship between skincare active ingredients and recommended skin types using a quantitative algorithmic approach. Secondary datasets were collected from Kaggle, e-commerce platforms, and product review websites, focusing on nine major active ingredients and seven skin types. The proposed methodology integrates a rule-based scoring approach, developed based on dermatological literature, to evaluate ingredient suitability, with Euclidean distance measurements to enhance score validation through dermatological feature vectors. Statistical correlation analysis was performed using cross-tabulation and the Chi-square test. The results revealed a statistically significant association between active ingredients and skin types ($p < 0.05$), confirming, for example, the effectiveness of salicylic acid for acne-prone skin and ceramide for dry skin. These findings were intuitively presented through heatmaps, bubble charts, bar charts, and pie charts generated using Python. The study demonstrates that the proposed data-driven approach provides an objective and transparent analytical framework for evaluating the compatibility between skincare active ingredients and skin types. The findings can serve as scientific evidence to support consumer decision-making and assist skincare product developers in formulating products that are better aligned with specific skin characteristics.

Keywords: *Rule-Based Scoring; Euclidean Distance; Data Visualization; Chi-Square Test; Skincare Active Ingredients.*

ABSTRAK

Kebutuhan konsumen akan produk perawatan kulit (*skincare*) yang sesuai dengan jenis kulit terus meningkat, namun panduan berbasis data yang menjembatani kecocokan bahan aktif dan kebutuhan kulit spesifik masih terbatas. Penelitian ini bertujuan menganalisis korelasi antara bahan aktif produk perawatan kulit dan jenis kulit yang direkomendasikan melalui pendekatan kuantitatif algoritmik. Menggunakan *dataset* sekunder dari Kaggle, platform *e-commerce*, dan situs ulasan produk, studi ini memfokuskan ruang lingkup pada 9 bahan aktif utama dan 7 jenis kulit. Metode yang diterapkan meliputi *rule-based scoring* berdasarkan literatur dermatologi untuk menilai tingkat kecocokan bahan, serta pengukuran jarak menggunakan *Euclidean distance* untuk memperkuat validitas skor melalui vektor fitur dermatologis. Analisis korelasi statistik dilakukan menggunakan tabulasi silang dan uji *Chi-Square*. Hasil uji menunjukkan adanya hubungan yang signifikan antara bahan aktif dan jenis kulit ($p\text{-value} < 0,05$), seperti efektivitas *salicylic acid* untuk kulit berjerawat dan *ceramide* untuk kulit kering. Temuan ini divisualisasikan secara intuitif melalui *heatmap*, *bubble chart*, *bar chart*, dan *pie chart* menggunakan Python. Kesimpulannya, pendekatan berbasis data ini berhasil membangun sistem analisis yang objektif dan transparan, yang dapat menjadi acuan ilmiah bagi konsumen maupun pengembang dalam formulasi produk *skincare*.

Kata Kunci: *Rule-based scoring, Euclidean distance, Visualisasi Data, Python, Chi-Square.*

PENDAHULUAN

Perkembangan industri *skincare* dalam beberapa tahun terakhir menunjukkan peningkatan yang signifikan seiring meningkatnya kesadaran masyarakat terhadap kesehatan dan perawatan kulit. Beragam produk dengan kandungan bahan aktif seperti *niacinamide*, *salicylic acid*, *hyaluronic acid*, dan *ceramide* dipasarkan dengan klaim manfaat yang berbeda sesuai karakteristik kulit pengguna. Namun, banyak konsumen masih mengalami kesulitan dalam menentukan produk yang sesuai karena keterbatasan pemahaman mengenai hubungan antara bahan aktif dan jenis kulit. Oleh karena itu, identifikasi kesesuaian bahan aktif terhadap jenis kulit menjadi aspek penting dalam mendukung efektivitas penggunaan produk *skincare* (Ramrakhiani & Kalbande, 2024).

Pemilihan bahan aktif yang tepat merupakan faktor utama dalam keberhasilan perawatan kulit. Berbagai penelitian dermatologi menunjukkan bahwa *salicylic acid* efektif untuk mengatasi kulit berminyak dan berjerawat karena kemampuannya dalam mengurangi produksi sebum dan membersihkan pori-pori, sedangkan *niacinamide* berperan dalam memperbaiki fungsi *skin barrier*, mengurangi inflamasi, dan membantu mengatasi hiperpigmentasi pada berbagai jenis kulit (Shariff et al., 2022). Selain itu, bahan aktif seperti *hyaluronic acid* dan *ceramide* banyak direkomendasikan untuk kulit kering karena kemampuannya meningkatkan hidrasi dan mempertahankan kelembapan kulit (Zhang et al., 2022).

Transformasi digital dan ketersediaan data dalam jumlah besar membuka peluang penerapan pendekatan berbasis data untuk mendukung pengambilan keputusan dalam pemilihan produk *skincare*. Berbagai penelitian terkini telah mengembangkan sistem rekomendasi kosmetik berbasis *machine learning*, *deep learning*, dan kecerdasan buatan untuk menghasilkan rekomendasi yang lebih personal sesuai kondisi kulit pengguna (Lee et al., 2024).

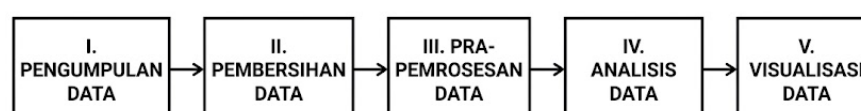
Metode statistik dan visualisasi data memiliki potensi untuk menghasilkan pemahaman yang lebih transparan mengenai keterkaitan antara bahan aktif dan karakteristik kulit. Analisis korelasi melalui tabulasi silang dan uji *Chi-Square* dapat digunakan untuk mengidentifikasi hubungan signifikan antarvariabel kategorikal, sedangkan visualisasi data mampu menyajikan pola hubungan secara lebih intuitif. Kombinasi pendekatan tersebut dengan metode *rule-based scoring* memungkinkan proses penilaian dilakukan berdasarkan pengetahuan dermatologis yang telah tervalidasi, sehingga hasil yang diperoleh tidak hanya bergantung pada frekuensi kemunculan data tetapi juga mempertimbangkan tingkat kecocokan klinis dari setiap bahan aktif terhadap jenis kulit tertentu (Ramrakhiani & Kalbande, 2024).

Selain pendekatan berbasis aturan, pengukuran kemiripan menggunakan *Euclidean Distance* dapat dimanfaatkan untuk mengevaluasi kesesuaian karakteristik bahan aktif dengan kebutuhan dermatologis masing-masing jenis kulit (Pangestu & Fitriani, 2022). Pendekatan ini memungkinkan representasi bahan aktif dan jenis kulit dalam bentuk vektor fitur sehingga hubungan keduanya dapat dianalisis secara kuantitatif. Integrasi antara *rule-based scoring* dan *Euclidean Distance* berpotensi menghasilkan proses evaluasi yang lebih objektif, terukur, dan mudah diinterpretasikan dibandingkan pendekatan rekomendasi yang hanya mengandalkan frekuensi data atau prediksi model semata (Lee et al., 2024).

Penelitian ini dilakukan untuk menganalisis korelasi antara bahan aktif *skincare* dan jenis kulit menggunakan pendekatan *rule-based scoring* dan *Euclidean Distance* yang didukung visualisasi data Python. Hasil penelitian diharapkan dapat menjadi dasar pengambilan keputusan yang lebih objektif bagi konsumen maupun pengembang produk dalam menentukan formulasi dan rekomendasi *skincare* yang sesuai dengan kebutuhan kulit pengguna.

METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif algoritmik menggunakan bahasa pemrograman Python untuk menganalisis korelasi antara bahan aktif *skincare* dan jenis kulit. Alur penelitian terdiri atas lima tahapan utama secara sistematis: pengumpulan data (*data collection*), pembersihan data (*data cleaning*), pra-pemrosesan data (*data preprocessing*), analisis data statistik (*data analysis*), dan visualisasi data (*data visualization*).



Gambar 1. Alur Penelitian

Pengumpulan dan Validasi Data

Penelitian ini menggunakan pendekatan kuantitatif dengan desain penelitian berbasis data (*data-driven research*) untuk menganalisis hubungan antara bahan aktif *skincare* dan jenis kulit yang direkomendasikan. Data penelitian diperoleh dari sumber sekunder yang berasal dari dataset Kaggle, platform *e-commerce*, dan situs ulasan produk. Dataset yang terkumpul memuat informasi mengenai nama produk, bahan aktif, jenis kulit yang direkomendasikan, serta kategori produk. Seluruh data kemudian diintegrasikan ke dalam satu basis data untuk mendukung proses analisis.

Konsolidasi Data: Data dari ketiga sumber digabungkan ke dalam satu berkas induk (*Skincare_Dataset.xlsx*). Dataset Kaggle bertindak sebagai basis data primer, sedangkan Tokopedia dan Female Daily digunakan untuk melengkapi informasi bahan aktif serta klaim kecocokan produk yang belum tertera.

Validasi Emid-Data: Konfirmasi keabsahan data dilakukan melalui pencocokan silang antar-sumber dengan menjadikan label resmi produk sebagai acuan utama.

Dimensi Dataset Akhir: Hasil validasi menghasilkan 1.019 entri data yang mencakup 873 produk unik dari 178 merek. Data tersebut dikelompokkan ke dalam 7 kategori produk dan ditargetkan untuk 7 jenis kulit (*oily, dry, sensitive, normal, acne-prone, dull*, dan kombinasi/semua jenis). Atribut utama yang diekstrak adalah *name, brand, ingredients, skintype*, dan *category*.

Pembersihan Data (*Data Cleaning*)

Proses pembersihan data dijalankan secara otomatis menggunakan pustaka analisis data pada Python guna menjamin konsistensi, validitas, dan keandalan sebelum memasuki tahapan pemodelan. Langkah awal pada fase ini adalah melakukan deteksi terhadap keberadaan nilai kosong pada seluruh baris dan kolom. Meskipun hasil pemeriksaan awal menunjukkan bahwa dataset tidak memiliki nilai yang hilang, tindakan preventif tetap diaplikasikan dengan menerapkan mekanisme eliminasi otomatis terhadap setiap baris yang berpotensi tidak lengkap, khususnya pada variabel bahan aktif dan jenis kulit. Selanjutnya, penanganan terhadap redundansi data dilakukan dengan mengidentifikasi dan menghapus seluruh entri duplikat guna menghindari bias atau standardisasi yang berlebihan dalam perhitungan frekuensi produk. Langkah terakhir dalam pembersihan ini adalah standardisasi teks dan penyelarasan istilah. Seluruh karakter huruf pada kolom bahan aktif dan jenis kulit diubah menjadi huruf kecil secara seragam, diikuti dengan penghapusan spasi berlebih pada awal dan akhir teks untuk memastikan kesetaraan format data. Mengingat adanya keragaman istilah penulisan jenis kulit dari berbagai sumber data, sebuah kamus pemetaan khusus diterapkan untuk menyatukan variasi penulisan teks yang bermakna sama, seperti menyelaraskan istilah bahasa daerah atau variasi bahasa asing yang tidak baku menjadi satu label kategori yang konsisten dan terstandardisasi.

Pra-Pemrosesan Data

Tahap awal penelitian diawali dengan *data cleaning* untuk memastikan kualitas data. Proses ini meliputi penghapusan data duplikat, penanganan nilai kosong (*missing values*), serta standarisasi format penulisan pada atribut bahan aktif dan jenis kulit. Selanjutnya dilakukan *data preprocessing* melalui pemisahan data *multi-label* pada atribut jenis kulit sehingga setiap label direpresentasikan sebagai entitas tersendiri. Pada tahap ini juga diterapkan pembobotan proporsional untuk menjaga representasi distribusi data yang seimbang.

Penanganan Data Multi-Label

Banyak produk memiliki klaim multi-label pada atribut jenis kulit (misalnya "kering, sensitif"). Entri ini didekomposisi menjadi baris terpisah dengan menerapkan pembobotan proporsional (*Weight*) berbasis rumus:

$$Weight = \frac{1}{n}$$

Di mana n merupakan jumlah label jenis kulit dalam satu entri produk. Jika produk memiliki 3 label, setiap baris pecahan memperoleh bobot 0.333. Proses ini meningkatkan volume dataset dari 1.019 menjadi 1.437 baris fungsional.

Penyaringan Bahan Aktif (*Low-Frequency Filtering*)

Ekstraksi bahan aktif dari string dilakukan secara bertingkat berdasarkan berbagai karakter pemisah (koma, titik koma, garis miring, kata hubung). Menggunakan prinsip filter frekuensi rendah untuk menghindari *overfitting*, ditetapkan ambang batas kemunculan minimal ≥ 10 kali ($\approx 0.7\%$ dari total data). Proses ini menyaring 279 kandidat awal menjadi 49 bahan aktif. Melalui pendekatan *elbow method* informal berbasis selisih frekuensi konstan, dipilih 13 bahan aktif dengan dominansi tertinggi (merepresentasikan 74% kumulatif kemunculan) sebagai fokus sistem.

Implementasi Sistem Algoritmik

Selanjutnya, dilakukan penilaian tingkat kecocokan antara bahan aktif dan jenis kulit menggunakan pendekatan *rule-based scoring*. Metode ini mengombinasikan frekuensi kemunculan bahan aktif dengan bobot kecocokan dermatologis yang ditentukan berdasarkan literatur ilmiah dan referensi klinis. Frekuensi kemunculan terbobot dari bahan aktif (b_i) pada jenis kulit (s_j) dihitung berdasarkan akumulasi nilai *multi-label*. Matriks frekuensi ini kemudian dikalikan dengan matriks bobot dermatologis ordinal (w) yang disintesis dari literatur klinis (0 = tidak cocok, 1 = netral, 2 = sangat cocok). Rumus skor akhir diidentifikasi sebagai:

$$\text{Skor Akhir}(b_i, s_j) = f(b_i, s_j) \times w(b_i, s_j)$$

Hanya kombinasi dengan nilai Skor Akhir ≥ 5 yang dipertahankan untuk mengeliminasi *noise*.

Untuk memperkuat validitas penilaian, penelitian ini menerapkan pengukuran *Euclidean Distance*. Setiap bahan aktif dan jenis kulit direpresentasikan dalam bentuk vektor fitur dermatologis yang mencakup karakteristik hidrasi, antiinflamasi, eksfoliasi, pengendalian sebum, dan sensitivitas. Nilai jarak yang diperoleh kemudian dikonversi menjadi skor kemiripan (*similarity score*). Bahan aktif dan jenis kulit ditransformasikan menjadi vektor numerik berdimensi 5 berbasis fitur dermatologis: hidrasi, anti-inflamasi, eksfoliasi, kontrol sebum, dan sensitivitas (skala 0–1). Jarak Euclidean (d) antar-vektor dihitung menggunakan rumus:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Nilai jarak tersebut dikonversi menjadi indeks kemiripan (*similarity score*) melalui formulasi:

$$\text{Similarity} = \frac{1}{1 + d}$$

Skor hasil *rule-based scoring* dan skor kemiripan *Euclidean* selanjutnya digabungkan menggunakan pendekatan *hybrid scoring* dengan bobot yang telah ditentukan, sehingga menghasilkan nilai akhir yang lebih objektif dalam menggambarkan kecocokan bahan aktif terhadap jenis kulit. Nilai *Rule-Based Scoring* yang telah dinormalisasi menggunakan *MinMaxScaler* (bobot 60%) digabungkan dengan nilai *Euclidean Similarity* (bobot 40%) untuk membentuk *Hybrid Score*:

$$\text{Hybrid Score} = (0.6 \times \text{Skor Rule-Based Ternormalisasi}) + (0.4 \times \text{Similarity})$$

Berdasarkan kriteria *Top-N Recommendation Filtering*, diambil 3 bahan aktif teratas dengan *Hybrid Score* tertinggi untuk setiap jenis kulit. Melalui eliminasi bahan duplikat lintas kategori, diperoleh 9 bahan aktif final untuk analisis korelasi.

Analisis Data Statistik

Tahap berikutnya adalah seleksi bahan aktif berdasarkan nilai *hybrid score* tertinggi untuk setiap jenis kulit. Bahan aktif yang secara konsisten menempati peringkat teratas dipilih sebagai kandidat utama dalam analisis statistik. Hubungan antara bahan aktif dan jenis kulit kemudian dianalisis menggunakan tabulasi silang (*crosstabulation*) dan uji Chi-Square. Pasangan data dari 9 bahan aktif final dan jenis kulit dipetakan ke dalam tabel kontingensi. Uji signifikansi inferensial dilakukan via uji Chi-Square (χ^2) untuk mengevaluasi dependensi antar-variabel kategorikal dengan rumus:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Uji Chi-Square digunakan untuk menguji signifikansi hubungan antara bahan aktif dan jenis kulit pada tingkat signifikansi $\alpha = 0,05$. Di mana O_i adalah frekuensi empiris yang diamati dan E_i adalah frekuensi teoritis yang diharapkan. Hipotesis nol (H_0) ditolak jika nilai *p-value* < 0.05 , mengindikasikan adanya korelasi yang signifikan secara statistik antara bahan aktif dan karakteristik kulit.

Visualisasi Data

Visualisasi data merupakan tahap akhir dalam penelitian yang bertujuan menyajikan hasil analisis dalam bentuk grafis sehingga pola hubungan antara bahan aktif *skincare* dan jenis kulit dapat diinterpretasikan secara lebih mudah, cepat, dan akurat. Pada penelitian ini, visualisasi dikembangkan menggunakan bahasa pemrograman Python dengan pustaka *Matplotlib* dan *Seaborn*. Visualisasi tidak hanya berfungsi sebagai media penyajian hasil, tetapi juga sebagai alat eksplorasi untuk mengidentifikasi pola distribusi, dominasi bahan aktif, serta kekuatan hubungan antarvariabel yang telah diuji menggunakan metode *Rule-Based Scoring*, *Euclidean Distance*, dan uji Chi-Square.

Heatmap Persentase Kecocokan Bahan Aktif dan Jenis Kulit

Heatmap digunakan untuk menampilkan tingkat kecocokan antara bahan aktif dan jenis kulit dalam bentuk matriks warna. Nilai yang divisualisasikan merupakan hasil normalisasi dari skor akhir (*final score*) sehingga seluruh nilai berada pada rentang 0–100%. Sebelum divisualisasikan, dilakukan normalisasi menggunakan metode *Min-Max Normalization*:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Keterangan

(X) = Nilai skor asli

(X_{min}) = Nilai skor minimum

(X_{max}) = Nilai skor maksimum

(X_{norm}) = Nilai hasil normalisasi

Selanjutnya nilai dinyatakan dalam bentuk persentase:

$$Persentase = X_{norm} \times 100$$

Interpretasi *heatmap* dilakukan berdasarkan intensitas warna. Semakin gelap warna yang ditampilkan, semakin tinggi tingkat kecocokan bahan aktif terhadap jenis kulit tertentu.

Bar Chart Frekuensi Bahan Aktif

Visualisasi *bar chart* digunakan untuk menggambarkan distribusi frekuensi kemunculan bahan aktif dalam dataset. Grafik ini membantu mengidentifikasi bahan aktif yang paling dominan digunakan dalam produk *skincare*. Frekuensi bahan aktif dihitung menggunakan:

$$f_i = \sum_{k=1}^n I(x_k = i)$$

Keterangan:

f_i Frekuensi bahan aktif ke- i

n Jumlah data

I Fungsi indikator

x_k Bahan aktif pada data ke- k

Tabel 1. Frekuensi Kemunculan Bahan Aktif Teratas

Bahan Aktif	Frekuensi
Niacinamide	147
Salicylic Acid	125
Centella Asiatica	118
Hyaluronic Acid	109
Ceramide	89
Glycerin	87
Tea Tree Oil	82
Vitamin C	78
Vitamin E	63

Heatmap Hybrid Score

Setelah diperoleh skor dari metode *Rule-Based Scoring* dan *Euclidean Similarity*, kedua skor digabungkan menjadi *Hybrid Score*.

$$HybridScore = (0.6 \times RuleScore) + (0.4 \times Similarity)$$

Keterangan:

Komponen Bobot

Rule-Based Score = 60%

Euclidean Similarity = 40%

Visualisasi *heatmap hybrid score* memungkinkan identifikasi bahan aktif terbaik untuk setiap jenis kulit secara simultan. Nilai tertinggi menunjukkan kombinasi yang paling direkomendasikan berdasarkan data empiris dan karakteristik dermatologis.

Bubble Chart Tingkat Kecocokan

Bubble chart digunakan untuk memperlihatkan tiga dimensi informasi secara bersamaan, yaitu:

1. Jenis kulit (sumbu X)
2. Bahan aktif (sumbu Y)
3. Nilai *Hybrid Score* (ukuran gelembung)

Ukuran gelembung dihitung menggunakan:

menunjukkan frekuensi tertinggi pada kulit berjerawat sebesar 89,5, sedangkan Centella Asiatica memiliki frekuensi tertinggi pada kulit sensitif sebesar 57,5.

✓ Matriks frekuensi (dengan pembobotan multi-label):

	Berminyak	Kering	Sensitif	Normal	Berjerawat	Kusam	Semua Jenis
niacinamide	20.0	8.0	8.5	44.5	18.5	76.0	41.5
hyaluronic acid	4.5	37.0	19.5	17.0	1.0	12.5	30.5
salicylic acid	33.0	0.0	2.0	0.0	89.5	0.5	2.0
glycerin	11.0	22.0	37.0	10.0	3.0	8.0	21.0
vitamin c	2.0	2.0	1.0	27.5	2.0	54.5	14.0
centella asiatica	6.0	8.5	58.5	1.0	20.0	0.0	6.0
tea tree oil	17.0	0.0	2.5	0.0	55.5	0.0	1.0
aloe vera	4.5	6.0	24.0	7.5	7.0	2.0	26.0
ceramide	2.0	20.0	22.5	2.0	2.0	2.5	2.0
allantoin	6.0	1.5	11.0	1.5	5.0	2.5	4.5
vitamin e	0.5	12.0	4.5	5.5	0.5	5.0	10.0
panthenol	2.5	7.5	20.5	1.0	0.5	0.0	3.0
collagen	0.5	7.5	1.0	7.5	0.5	5.0	3.0

Gambar 3. Hasil Program Pembuatan Matriks Frekuensi

Tahap berikutnya dilakukan perhitungan *rule-based scoring* dengan menggabungkan frekuensi kemunculan dan bobot kecocokan dermatologis. Hasil perhitungan menunjukkan bahwa Salicylic Acid memperoleh skor tertinggi pada kulit berjerawat sebesar 179,0. Niacinamide memperoleh skor tertinggi pada kulit kusam sebesar 152,0, sedangkan Centella Asiatica memperoleh skor tertinggi pada kulit sensitif sebesar 117,0.

✓ Skor akhir (Frekuensi × Bobot):

	Berminyak	Kering	Sensitif	Normal	Berjerawat	Kusam	Semua Jenis
niacinamide	40.0	8.0	17.0	89.0	37.0	152.0	83.0
hyaluronic acid	4.5	74.0	39.0	34.0	1.0	12.5	61.0
salicylic acid	66.0	0.0	0.0	0.0	179.0	0.5	0.0
glycerin	11.0	44.0	74.0	20.0	3.0	8.0	42.0
vitamin c	2.0	2.0	0.0	55.0	2.0	109.0	0.0
centella asiatica	12.0	8.5	117.0	2.0	40.0	0.0	12.0
tea tree oil	34.0	0.0	0.0	0.0	111.0	0.0	0.0
aloe vera	4.5	6.0	24.0	7.5	7.0	2.0	26.0
ceramide	2.0	40.0	45.0	4.0	2.0	2.5	4.0
allantoin	6.0	1.5	11.0	1.5	5.0	2.5	4.5
vitamin e	0.5	24.0	4.5	11.0	0.5	5.0	20.0
panthenol	5.0	15.0	41.0	2.0	1.0	0.0	6.0
collagen	0.5	7.5	1.0	7.5	0.5	5.0	3.0

Gambar 4. Hasil Perhitungan Skor Akhir

Selanjutnya dilakukan pengukuran kemiripan menggunakan metode Euclidean Distance yang dikonversi menjadi nilai *similarity*. Nilai *similarity* kemudian digabungkan dengan skor *rule-based* menggunakan metode *hybrid scoring* dengan komposisi bobot 60% untuk skor *rule-based* dan 40% untuk skor Euclidean Similarity. Hasil *hybrid scoring* menunjukkan bahwa bahan aktif dengan skor tertinggi untuk masing-masing jenis kulit adalah sebagai berikut:

1. Kulit berjerawat: Salicylic Acid (0,786)
2. Kulit sensitif: Centella Asiatica (0,789)
3. Kulit kering: Hyaluronic Acid (0,784)
4. Kulit normal: Niacinamide (0,838)
5. Kulit kusam: Niacinamide (0,834)
6. Semua jenis kulit: Niacinamide (0,831)

Berdasarkan hasil seleksi *Top-3 Hybrid Recommendation*, diperoleh 9 bahan aktif yang secara konsisten berada pada peringkat tiga besar untuk satu atau lebih jenis kulit. Bahan aktif tersebut digunakan pada tahap analisis statistik selanjutnya.

Jenis Kulit: BERMINYAK			
Salicylic Acid	Rule: 1.000	Sim: 0.455	Hybrid: 0.782
Niacinamide	Rule: 0.603	Sim: 0.479	Hybrid: 0.554
Tea Tree Oil	Rule: 0.511	Sim: 0.455	Hybrid: 0.489
Jenis Kulit: KERING			
Hyaluronic Acid	Rule: 1.000	Sim: 0.459	Hybrid: 0.784
Glycerin	Rule: 0.595	Sim: 0.442	Hybrid: 0.534
Ceramide	Rule: 0.541	Sim: 0.481	Hybrid: 0.517
Jenis Kulit: SENSITIF			
Centella Asiatica	Rule: 1.000	Sim: 0.473	Hybrid: 0.789
Glycerin	Rule: 0.632	Sim: 0.471	Hybrid: 0.568
Ceramide	Rule: 0.385	Sim: 0.512	Hybrid: 0.435
Jenis Kulit: NORMAL			
Niacinamide	Rule: 1.000	Sim: 0.596	Hybrid: 0.838
Vitamin C	Rule: 0.618	Sim: 0.581	Hybrid: 0.603
Hyaluronic Acid	Rule: 0.382	Sim: 0.538	Hybrid: 0.444
Jenis Kulit: BERJERAWAT			
Salicylic Acid	Rule: 1.000	Sim: 0.464	Hybrid: 0.786
Tea Tree Oil	Rule: 0.619	Sim: 0.447	Hybrid: 0.550
Centella Asiatica	Rule: 0.221	Sim: 0.480	Hybrid: 0.325
Jenis Kulit: KUSAM			
Niacinamide	Rule: 1.000	Sim: 0.586	Hybrid: 0.834
Vitamin C	Rule: 0.717	Sim: 0.528	Hybrid: 0.641
Vitamin E	Rule: 0.033	Sim: 0.622	Hybrid: 0.268
Jenis Kulit: SEMUA JENIS			
Niacinamide	Rule: 1.000	Sim: 0.579	Hybrid: 0.831
Hyaluronic Acid	Rule: 0.735	Sim: 0.499	Hybrid: 0.640
Glycerin	Rule: 0.506	Sim: 0.533	Hybrid: 0.517

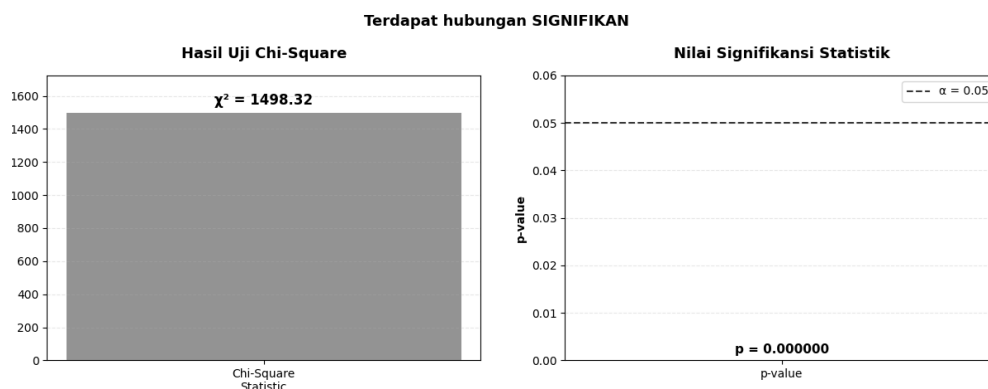
Gambar 5. Hasil Perhitungan *Hybrid Score*

Analisis hubungan antara bahan aktif dan jenis kulit dilakukan menggunakan tabulasi silang (*crosstab*) dan uji Chi-Square. Tabel kontingensi menunjukkan distribusi frekuensi yang berbeda antar kombinasi bahan aktif dan jenis kulit. Salicylic Acid memiliki frekuensi tertinggi pada kulit berjerawat sebanyak 119 kemunculan. Vitamin C memiliki frekuensi tertinggi pada kulit kusam sebanyak 79 kemunculan, sedangkan Centella Asiatica memiliki frekuensi tertinggi pada kulit sensitif sebanyak 80 kemunculan.

Tabel 2. Output Tabel *Crosstab*

<i>Ingredient</i>	Berjerawat	Berminyak	Kering	Kusam	Normal	Semua Jenis	Sensitif
<i>Centella Asiatica</i>	32	12	15	0	2	6	80
<i>Ceramide</i>	3	4	35	4	4	2	37
<i>Glycerin</i>	5	15	37	15	17	26	51
<i>Hyaluronic Acid</i>	2	6	64	20	30	31	36
<i>Niacinamide</i>	31	33	16	111	83	42	15
<i>Salicylic Acid</i>	119	61	0	1	0	2	4
<i>Tea Tree Oil</i>	72	34	0	0	0	1	4
<i>Vitamin C</i>	3	3	4	79	50	17	2
<i>Vitamin E</i>	1	1	16	6	11	10	6

Hasil uji Chi-Square menghasilkan nilai χ^2 sebesar 1498,321 dengan derajat kebebasan (df) sebesar 48 dan nilai signifikansi (p-value) sebesar 0,000.

**Gambar 6.** Hasil Visualisasi Hasil Uji Chi-Square

Nilai p-value yang lebih kecil dari 0,05 menunjukkan adanya hubungan yang signifikan antara bahan aktif skincare dan jenis kulit yang direkomendasikan.

Pembahasan

Hasil penelitian menunjukkan bahwa terdapat hubungan yang signifikan antara bahan aktif skincare dan jenis kulit yang direkomendasikan. Temuan ini mengindikasikan bahwa distribusi bahan aktif pada produk skincare tidak terjadi secara acak, melainkan mengikuti pola tertentu yang berkaitan dengan kebutuhan karakteristik masing-masing jenis kulit. Hubungan signifikan yang diperoleh dari uji Chi-Square menguatkan bahwa bahan aktif tertentu cenderung diasosiasikan dengan jenis kulit tertentu secara konsisten. Temuan mengenai dominasi Salicylic Acid pada kulit berjerawat menunjukkan bahwa bahan aktif tersebut memiliki tingkat relevansi yang tinggi terhadap kebutuhan kulit yang rentan terhadap penyumbatan pori dan produksi sebum berlebih. Hasil ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa Salicylic Acid efektif digunakan sebagai agen eksfoliasi dan pengontrol jerawat. Tingginya skor *hybrid* yang diperoleh menunjukkan bahwa baik dari sisi frekuensi penggunaan maupun kesesuaian karakteristik dermatologis, bahan ini memiliki kecocokan yang kuat terhadap kulit berjerawat. Pada kategori kulit sensitif, Centella Asiatica memperoleh skor tertinggi dibandingkan bahan aktif lainnya. Hasil ini mendukung temuan penelitian terdahulu yang menjelaskan bahwa Centella Asiatica memiliki sifat antiinflamasi dan kemampuan mempercepat regenerasi kulit sehingga sering digunakan pada formulasi untuk kulit sensitif. Kesesuaian tersebut juga tercermin pada nilai *similarity* yang relatif tinggi dalam analisis Euclidean Distance. Hasil penelitian juga menunjukkan bahwa Hyaluronic Acid menjadi bahan aktif dengan skor tertinggi pada kulit kering. Temuan ini konsisten dengan teori dermatologi yang menyatakan bahwa Hyaluronic Acid berfungsi sebagai humektan yang mampu meningkatkan retensi air pada lapisan kulit. Tingginya nilai *hybrid score* menunjukkan bahwa kebutuhan hidrasi pada kulit kering memiliki kesesuaian yang kuat dengan karakteristik bahan tersebut. Niacinamide muncul sebagai bahan aktif yang paling dominan pada beberapa kategori sekaligus, yaitu kulit normal, kulit kusam, dan semua jenis kulit. Temuan ini menunjukkan bahwa Niacinamide memiliki sifat multifungsi dan fleksibilitas penggunaan yang lebih luas dibandingkan bahan aktif lainnya. Hasil ini memperluas temuan penelitian sebelumnya yang menyebutkan bahwa Niacinamide mampu meningkatkan fungsi *skin barrier*, membantu mencerahkan kulit, serta mengontrol produksi minyak secara seimbang. Dominasi Niacinamide pada beberapa kategori mengindikasikan bahwa bahan tersebut memiliki tingkat adaptabilitas yang tinggi terhadap berbagai kebutuhan kulit.

Pendekatan *hybrid scoring* yang menggabungkan *rule-based scoring* dan Euclidean Distance memberikan kontribusi metodologis dalam proses rekomendasi bahan aktif skincare. Pendekatan ini tidak hanya mempertimbangkan frekuensi kemunculan bahan aktif dalam dataset, tetapi juga mempertimbangkan kemiripan karakteristik dermatologis antara bahan aktif dan kebutuhan jenis kulit. Dengan demikian, hasil rekomendasi menjadi lebih objektif dibandingkan pendekatan yang hanya mengandalkan frekuensi data atau pengetahuan pakar secara terpisah. Temuan ini menunjukkan bahwa integrasi metode berbasis aturan dan pengukuran kemiripan dapat meningkatkan kualitas proses identifikasi kecocokan bahan aktif terhadap jenis kulit. Hasil visualisasi melalui *heatmap*, *bar chart*, dan grafik distribusi menunjukkan pola hubungan yang jelas antara bahan aktif dan jenis kulit. Visualisasi tersebut mempermudah identifikasi bahan aktif dominan pada setiap kategori kulit dan membantu menjelaskan pola yang tidak mudah diamati melalui tabel numerik. Dengan demikian, visualisasi data berperan sebagai alat pendukung interpretasi hasil analisis yang efektif. Meskipun penelitian ini berhasil menunjukkan hubungan signifikan antara bahan aktif dan jenis kulit, terdapat beberapa keterbatasan yang perlu diperhatikan. Pertama, penelitian menggunakan data sekunder yang diperoleh dari dataset publik, platform *e-commerce*, dan situs ulasan produk sehingga kualitas informasi sangat bergantung pada kelengkapan data sumber. Kedua, bobot kecocokan dermatologis dalam *rule-based scoring* masih disusun berdasarkan literatur dan belum divalidasi langsung oleh pakar dermatologi. Ketiga, representasi fitur dalam Euclidean Distance menggunakan lima karakteristik utama sehingga belum sepenuhnya merepresentasikan kompleksitas kondisi kulit dan karakteristik bahan aktif. Oleh karena itu, penelitian selanjutnya dapat memperluas jumlah fitur dermatologis, melibatkan validasi pakar, serta menguji metode rekomendasi lain berbasis *machine learning* untuk meningkatkan akurasi hasil rekomendasi.

KESIMPULAN

Berdasarkan hasil penelitian mengenai analisis korelasi antara bahan aktif *skincare* dan jenis kulit menggunakan pendekatan *rule-based scoring*, pengukuran *Euclidean Distance*, serta visualisasi data berbasis Python, dapat disimpulkan bahwa terdapat hubungan yang signifikan antara bahan aktif dengan jenis kulit yang direkomendasikan. Penelitian ini berhasil membangun pendekatan berbasis

data untuk mengidentifikasi pola keterkaitan antara bahan aktif dan kebutuhan kulit melalui proses pengolahan data, pembobotan kecocokan, serta analisis statistik. Hasil seleksi data menunjukkan bahwa dari 279 bahan aktif yang terdapat dalam dataset awal, dilakukan penyaringan berdasarkan frekuensi kemunculan sehingga diperoleh bahan aktif yang lebih representatif untuk dianalisis. Melalui kombinasi metode *rule-based scoring* dan *Euclidean Distance*, sistem mampu menentukan tingkat kesesuaian bahan aktif terhadap karakteristik jenis kulit berdasarkan aspek frekuensi penggunaan dan kemiripan fitur dermatologis. Hasil analisis menunjukkan beberapa pola dominan, seperti *Salicylic Acid* yang memiliki tingkat kecocokan tinggi terhadap kulit berjerawat, *Hyaluronic Acid* terhadap kulit kering, serta *Centella Asiatica* terhadap kulit sensitif. Pengujian korelasi menggunakan metode *crosstab* dan uji Chi-Square menghasilkan nilai signifikansi $p\text{-value} < 0,05$, yang menunjukkan bahwa hubungan antara bahan aktif dan jenis kulit bersifat signifikan secara statistik. Hal tersebut menunjukkan bahwa pemilihan bahan aktif tidak terjadi secara acak, melainkan memiliki pola keterkaitan tertentu berdasarkan karakteristik kulit. Selain itu, penggunaan visualisasi data berupa *Heatmap*, *Bar Chart*, dan *Pie Chart* menggunakan Python mampu menyajikan pola hubungan secara lebih informatif sehingga hasil analisis lebih mudah dipahami dan diinterpretasikan. Dengan demikian, penelitian ini membuktikan bahwa pendekatan analisis berbasis data dapat digunakan untuk mendukung proses rekomendasi bahan aktif *skincare* secara lebih objektif, terstruktur, dan transparan. Sistem yang dikembangkan dapat menjadi referensi bagi konsumen maupun pengembang produk dalam memahami kecocokan bahan aktif terhadap jenis kulit tertentu.

REFERENSI

- Ramrakhiani, N., & Kalbande, D. (2024). A comprehensive review of AI-powered skincare product recommendation systems: From data collection to user experience. *Digital Health*, 10, 1–23. <https://doi.org/10.1177/20427530241304073>
- Shariff, R., Du, Y., Dutta, M., Kumar, S., Thimmaiah, S., Doraiswamy, C., ... Damodaran, A. (2022). Superior even skin tone and anti-ageing benefit of a combination of 4-hexylresorcinol and niacinamide. *International Journal of Cosmetic Science*, 44(1), 103–117. <https://doi.org/10.1111/ics.12759>
- Pangestu, M. S., & Fitriani, M. A. (2022). Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means. *Sainteks*, 19(2), 141–155. <https://doi.org/10.30595/sainteks.v19i2.14495>
- Zhang, L., Shao, X., Chen, Y., Wang, J., Ariyawati, A., Zhang, Y., ... Chen, J. (2022). 30% supramolecular salicylic acid peels effectively treats acne vulgaris and reduces facial sebum. *Journal of Cosmetic Dermatology*, 21(8), 3398–3405. <https://doi.org/10.1111/jocd.14799>
- Lee, J., Yoon, H., Kim, S., Lee, C., Lee, J., & Yoo, S. (2024). Deep learning-based skin care product recommendation: A focus on cosmetic ingredient analysis and facial skin conditions. *Journal of Cosmetic Dermatology*, 23(6), 2066–2077. <https://doi.org/10.1111/jocd.16218>
- Draelos, Z. D. (2021). The science behind skin care: Moisturizers and barrier repair. *Dermatologic Clinics*, 39(3), 433–440. <https://doi.org/10.1016/j.det.2021.02.003>
- Boo, Y. C. (2023). Mechanistic understanding and clinical applications of salicylic acid in dermatology. *International Journal of Molecular Sciences*, 24(4), 3568. <https://doi.org/10.3390/ijms24043568>
- Nguyen, H. L., & Soulika, A. M. (2022). The dynamics of the skin's immune system. *International Journal of Molecular Sciences*, 23(11), 6177. <https://doi.org/10.3390/ijms23116177>
- McHugh, M. L. (2021). The Chi-square test of independence. *Biochemia Medica*, 23(2), 143–149. <https://doi.org/10.11613/BM.2013.018>
- Rana, M. M., & Singha, M. (2024). Application of Chi-square test in categorical data analysis: A review. *International Journal of Statistics and Applied Mathematics*, 9(1), 45–52. <https://doi.org/10.22271/math.2024.v9.i1a.1560>
- Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, 9(1), 59. <https://doi.org/10.1186/s40537-022-00592-5>
- Raziperchikolaei, R., & Chung, Y. J. (2022). One-class recommendation systems with the hinge pairwise distance loss and orthogonal representations. <https://doi.org/10.48550/arXiv.2208.14594>
- Adebanjo, S., & Banchani, E. (2023). Application of different Python libraries for visualisation of data. *International Journal of Data Science*, 4(2), 67–83. <https://doi.org/10.18517/ijods.4.2.67-83.2023>
- Wornow, M., Ross, E. G., Callahan, A., & Shah, N. H. (2023). APLUS: A Python library for usefulness simulations of machine learning models in healthcare. *Journal of Biomedical Informatics*, 139,

104319. <https://doi.org/10.1016/j.jbi.2023.104319>
- Kim, H. Y. (2022). Statistical notes for clinical researchers: Chi-square test and Fisher's exact test. *Restorative Dentistry & Endodontics*, 47(1), e1. <https://doi.org/10.5395/rde.2022.47.e1>
- Pannadhitthana Candra, A. (2025). Analisis Data Menggunakan Python: Memperkenalkan Pandas dan NumPy. *Journal of Information System and Education Development*, 3(1), 11–16. <https://doi.org/10.62386/jised.v3i1.118>