

**SISTEM REKOMENDASI PRODUK SKINCARE BERDASARKAN
PROFIL JENIS KULIT WAJAH MENGGUNAKAN
PENDEKATAN *CONTENT-BASED FILTERING***

SKRIPSI

Oleh :
LENY LATIFATUROHMAH
NIM. 220605110154



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

**SISTEM REKOMENDASI PRODUK SKINCARE BERDASARKAN
PROFIL JENIS KULIT WAJAH MENGGUNAKAN
PENDEKATAN *CONTENT-BASED FILTERING***

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
LENY LATIFATUROHMAH
NIM. 220605110154

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

HALAMAN PERSETUJUAN

SISTEM REKOMENDASI PRODUK *SKINCARE* BERDASARKAN PROFIL JENIS KULIT WAJAH MENGGUNAKAN PENDEKATAN *CONTENT-BASED FILTERING*

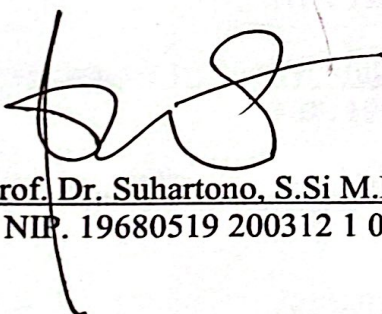
SKRIPSI

Oleh :

LENY LATIFATUROHMAH
NIM. 220605110154

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 2 Juni 2026

Pembimbing I,

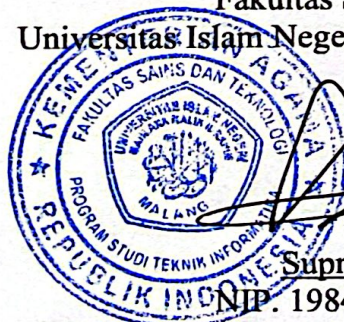

Prof. Dr. Suhartono, S.Si M.Kom
NIP. 19680519 200312 1 001

Pembimbing II,


Ahmad Fahmi Karami, M.Kom
NIP. 19870909 202012 1 001

Mengetahui,

Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M. Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

SISTEM REKOMENDASI PRODUK *SKINCARE* BERDASARKAN PROFIL JENIS KULIT WAJAH MENGGUNAKAN PENDEKATAN *CONTENT-BASED FILTERING*

SKRIPSI

Oleh :

LENY LATIFATUROHMAH
NIM. 220605110154

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 12 Juni 2026

Susunan Dewan Penguji

Ketua Penguji : Dr. Agung Teguh Wibowo Almais, M.T
NIP. 19860301 202321 1 016

Anggota Penguji I : Tri Mukti Lestari, M.Kom
NIP. 19911108 202012 2 005

Anggota Penguji II : Prof. Dr. Suhartono, S.Si M.Kom
NIP. 19680519 200312 1 001

Anggota Penguji III : Ahmad Fahmi Karami, M.Kom
NIP. 19870909 202012 1 001

(Handwritten signatures of the examiners)

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Leny Latifaturohmah

NIM : 220605110154

Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika

Judul Skripsi : Sistem Rekomendasi Produk *Skincare* Berdasarkan
Profil Jenis Kulit Wajah Menggunakan Pendekatan
Content-Based Filtering

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 20 Juni 2026

Yang membuat pernyataan,



Leny Latifaturohmah

NIM.220605110154

MOTTO

*“Allah SWT tidak mengatakan hidup ini mudah. Tetapi Allah SWT berjanji,
bahwa sesungguhnya bersama kesulitan ada kemudahan.”*

(Q.S Al-Insyirah, 94 : 5-6)

*“Sebutlah namaNya, tetap di jalanNya. kelak kau mengingat, kau akan teringat,
Terus Berenang Lanjutlah Mendaki.”*

(33x - Perunggu)

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur ke hadirat Allah SWT atas segala rahmat, kasih sayang, pertolongan, serta nikmat yang tiada henti diberikan kepada penulis, karya sederhana ini penulis persembahkan sebagai wujud rasa terima kasih dan penghormatan kepada orang-orang yang senantiasa hadir dalam setiap proses kehidupan penulis.

Persembahan pertama dan yang paling istimewa penulis tujukan kepada Ayah Ibu tercinta. Terima kasih atas setiap doa yang tidak pernah putus, kasih sayang yang tidak pernah berkurang, serta pengorbanan yang tidak pernah dapat terbalaskan. Terima kasih sudah menjadi sumber kekuatan dan alasan terbesar bagi penulis untuk terus berjuang serta semangat untuk terus melangkah. Di setiap langkah yang penulis tempuh, selalu ada doa, dukungan, dan kepercayaan dari Ayah dan Ibu yang menguatkan hingga akhirnya karya ini dapat terselesaikan. Semoga Allah SWT senantiasa melimpahkan kesehatan, kebahagiaan, keberkahan umur, serta membalas setiap pengorbanan Ayah dan Ibu dengan balasan yang terbaik.

Persembahan ini juga penulis tujukan kepada diri sendiri. Terima kasih telah memilih untuk tetap bertahan, bangkit setiap kali merasa lelah, dan tidak menyerah meskipun perjalanan ini dipenuhi tantangan dan air mata. Pencapaian ini menjadi bukti bahwa setiap doa dan usaha yang dilakukan dengan sungguh-sungguh tidak pernah sia-sia. Semoga karya sederhana ini tidak hanya menjadi akhir dari perjalanan menempuh pendidikan sarjana, tetapi juga menjadi awal untuk terus belajar, berkarya, dan memberikan manfaat bagi banyak orang.

KATA PENGANTAR

Alhamdulillahirabbil ‘alamin, segala puji hanya milik Allah Subhānahu wa Ta‘ālā atas segala nikmat dan kasih sayang-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Sistem Rekomendasi Produk *Skincare* Berdasarkan Profil Jenis Kulit Wajah Menggunakan Pendekatan *Content-Based Filtering*” dengan baik. Penulis menyadari bahwa dalam proses penyusunan skripsi ini terdapat berbagai keterbatasan dan tantangan, namun berkat bantuan, bimbingan, dukungan, serta motivasi dari berbagai pihak, skripsi ini dapat diselesaikan dengan baik. Ucapan ini penulis sampaikan kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., CAHRM., CRMP., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Prof. Dr. Suhartono, S.Si, M.Kom., selaku Dosen Pembimbing I, yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan arahan, bimbingan, serta masukan yang sangat berharga selama proses penyusunan skripsi ini.
5. Ahmad Fahmi Karami, M.Kom., selaku Dosen Pembimbing II, yang telah memberikan bimbingan, arahan, serta berbagai saran yang membangun dalam penyusunan skripsi ini.

6. Dr. Agung Teguh Wibowo Almais, M.T., selaku Ketua Penguji, yang telah memberikan saran, dan masukan yang konstruktif sehingga skripsi ini menjadi lebih baik.
7. Tri Mukti Lestari, M.Kom., selaku Anggota Penguji I, yang telah memberikan berbagai saran, masukan, dan evaluasi yang sangat bermanfaat dalam penyempurnaan skripsi ini.
8. Seluruh Dosen, Admin, dan Staf Program Studi Teknik Informatika, yang telah berbagi ilmu, tenaga, dan kebaikan selama masa studi ini, baik secara langsung maupun tidak langsung.
9. Cinta Pertamaku sekaligus panutan hidupku, Ayah Sunani dan pintu surgaku Ibu Istiqomah yang sangat penulis sayangi. Terimakasih atas cinta yang tak bertepi dan pengorbanan yang tak terhingga. Teruntuk Ayah, terimakasih selalu berjuang dalam mengupayakan yang terbaik untuk kehidupan penulis, berkorban keringat, tenaga dan pikiran sehingga penulis mampu meraih gelar sarjana. Teruntuk Ibu, terima kasih sudah menjadi ibu terbaik yang selalu mengusahakan apapun. Terimakasih atas cinta, kasih sayang, kesabaran, dan doa yang selalu dipanjatkan. Penulis percaya doa-doa dari Ayah dan Ibu yang selalu menyelamatkan dan menuntun penulis melewati masa-masa sulit. Semoga Allah SWT membalas kebaikan dan ketulusan dengan keberkahan dan kesehatan sepanjang usia. Semoga pencapaian ini menjadi salah satu bentuk bakti yang dapat penulis persembahkan kepada Ayah dan Ibu.

10. Keluarga tercinta, terutama Kakak penulis, Much. Aziz Hidayatulloh, beserta Kakak Ipar Ifa Nur Azizah, Tante Zuli, dan Adek Aini. Terima kasih atas doa, kasih sayang, perhatian, dan dukungan yang selalu diberikan kepada penulis. Terima kasih telah menjadi penguat dan tempat penulis kembali bangkit dalam setiap proses perjuangan hingga mampu menyelesaikan skripsi ini.
11. Sahabat seperjuangan penulis yang setia menemani awal perkuliahan hingga saat ini, Jenny, Annisa, dan Runa. Terima kasih atas kebersamaan, doa, dukungan, serta semangat yang telah diberikan kepada penulis selama menjalani perkuliahan hingga proses penyusunan skripsi ini. Terima kasih telah menjadi tempat berbagi cerita, berdiskusi, saling membantu, serta saling menguatkan dalam menghadapi setiap tantangan selama masa perkuliahan. Semoga Allah SWT membalas setiap kebaikan kalian dan semoga kita semua diberikan kemudahan dalam meraih cita-cita serta kesuksesan di masa depan.
12. Sahabat-sahabat penulis sejak kecil, Asna, Iis, Indana, dan Nisa. Terima kasih telah menjadi bagian dari perjalanan hidup penulis sejak masa kecil hingga saat ini. Terima kasih atas segala perhatian, doa, dukungan, dan kenangan yang telah tercipta bersama. Kehadiran kalian menjadi warna tersendiri dalam perjalanan hidup penulis dan menjadi tempat untuk kembali berbagi cerita di setiap keadaan.
13. Kepada sahabat SMA penulis yang tak kalah penting kehadirannya (Bocahpitu), Zihan, Elis, Fira, Maya, Citra, dan Alsa. Terima kasih atas

kebersamaan, canda tawa, dan semangat yang telah menemani masa-masa sekolah hingga saat ini. Terima kasih karena tetap hadir, saling mendukung, dan selalu memberikan semangat meskipun kini masing-masing telah menempuh jalan yang berbeda. Semoga segala impian dan cita-cita yang sedang diperjuangkan dapat terwujud.

14. Seluruh teman-teman Angkatan 2022 Teknik Informatika yang telah banyak membantu, mendukung, dan memotivasi dalam menyelesaikan penyusunan skripsi ini.

15. Last but not least, ya! Leny Latifaturohmah. Terima kasih telah bertahan, bangkit, dan terus melangkah di tengah berbagai ujian kehidupan. Terima kasih karena tidak menyerah, meskipun perjalanan ini tidak selalu mudah. Semoga setiap proses yang telah dilalui menjadi pelajaran berharga untuk menjadi pribadi yang lebih kuat, lebih ikhlas, dan lebih baik di masa yang akan datang.

Malang, 20 Juni 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN.....	v
MOTTO... ..	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xii
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL	xv
ABSTRAK	xvi
ABSTRACT	xvii
مستخلص البحث.....	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
BAB II STUDI PUSTAKA.....	6
2.1 Penelitian Terkait	6
2.2 <i>Skincare</i>	11
2.3 Jenis Kulit Wajah.....	13
2.4 Sistem Rekomendasi	16
2.5 <i>Content-Based Filtering</i>	17
2.6 <i>Term Frequency–Inverse Document Frequency (TF-IDF)</i>	19
2.7 <i>Cosine Similarity</i>	20
BAB III DESAIN DAN IMPLEMENTASI	21
3.1 Desain Penelitian	21
3.2 Desain Sistem.....	22
3.3 Pengumpulan Data	24
3.4 <i>Pre-processing Data</i>	28
3.4.1 <i>Data Cleaning</i>	28
3.4.2 <i>Data Integration</i>	30
3.4.3 <i>Text Normalization</i>	31
3.5 <i>Content-Based Filtering</i>	34

3.5.1 TF-IDF (<i>Term Frequency–Inverse Document Frequency</i>)	36
3.5.2 <i>Cosine similarity</i>	38
3.6 Evaluasi dan Pengujian Sistem	39
BAB IV HASIL DAN PEMBAHASAN	40
4.1 <i>Pre-processing Data</i>	40
4.1.1 <i>Data Cleaning</i>	40
4.1.2 <i>Data Integration</i>	41
4.1.3 <i>Text Normalization</i>	42
4.2 Pembobotan Kata (TF-IDF).....	47
4.3 Perhitungan <i>Cosine similarity</i>	51
4.4 Hasil Sistem Rekomendasi	54
4.5 Analisis Uji Coba.....	58
4.6 Pembahasan.....	69
4.7 Integrasi Islam.....	72
BAB V KESIMPULAN DAN SARAN	78
5.1 Kesimpulan	78
5.2 Saran... ..	79
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 2.1 Skema <i>Content-Based Filtering</i>	17
Gambar 3.1 Desain Penelitian.....	21
Gambar 3.2 Desain Sistem.....	23
Gambar 3.3 Tahapan <i>Preprocessing</i>	28
Gambar 3.4 Flowchart <i>Content-Based Filtering</i>	35
Gambar 4.1 Distribusi Jenis Kulit.....	41
Gambar 4.2 <i>Pseudocode Merge Dataset</i>	41
Gambar 4.3 Hasil Integrasi Data.....	42
Gambar 4.4 <i>Pseudocode Case Folding</i>	43
Gambar 4.5 <i>Pseudocode Remove Punctuation</i>	43
Gambar 4.6 <i>Pseudocode Tokenizing</i> dan Normalisasi Karakter Berulang	44
Gambar 4.7 <i>Pseudocode Stopword Removal</i>	45
Gambar 4.8 <i>Pseudocode Stemming</i>	45
Gambar 4.9 <i>Pseudocode Reconstruction</i>	46
Gambar 4.10 Hasil <i>Pre-processing</i> Data	47
Gambar 4.11 <i>Pseudocode TF-IDF</i>	47
Gambar 4.12 <i>Pseudocode Cosine Similarity</i>	51
Gambar 4.13 Halaman Input Pengguna	56
Gambar 4.14 Halaman Hasil Rekomendasi	57
Gambar 4.15 Grafik Rata-Rata <i>Precision</i> Sistem	68
Gambar 4.16 Grafik Distribusi Hasil Evaluasi Sistem	69

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	9
Tabel 3.1 Struktur Dataset InfoProduct.csv	24
Tabel 3.2 Struktur Dataset Review.csv	25
Tabel 3.3 Atribut yang Digunakan dalam Perhitungan.....	27
Tabel 3.4 Tabel Hasil Integrasi Atribut Input dan Output	31
Tabel 3.5 Contoh Dokumen pada Tahap <i>Case Folding</i>	32
Tabel 3.6 Contoh Dokumen pada Tahap <i>Remove Punctuation</i>	32
Tabel 3.7 Contoh Dokumen pada Tahap <i>Stopword Removal</i>	33
Tabel 3.8 Contoh Dokumen pada Tahap <i>Tokenizing</i>	33
Tabel 3.9 Contoh Dokumen pada Tahap <i>Stemming</i>	34
Tabel 3.10 Contoh Dokumen pada <i>Tahap Reconstruction</i>	34
Tabel 4.1 Hasil <i>Case Folding</i>	43
Tabel 4.2 Hasil <i>Punctuation Removal</i>	44
Tabel 4.3 Hasil <i>Tokenizing</i>	44
Tabel 4.4 Hasil <i>Stopword Removal</i>	45
Tabel 4.5 Hasil <i>Stemming</i>	46
Tabel 4.6 Hasil <i>Reconstruction</i>	46
Tabel 4.7 Sampel Hasil Perhitungan TF	48
Tabel 4.8 Sampel Hasil Perhitungan IDF	49
Tabel 4.9 Sampel Hasil Perhitungan TF-IDF	50
Tabel 4.10 Sampel Hasil Perhitungan <i>Cosine similarity</i>	53
Tabel 4.11 Hasil Skenario 1	60
Tabel 4.12 Hasil Skenario 2	60
Tabel 4.13 Hasil Skenario 3	61
Tabel 4.14 Hasil Skenario 4	62
Tabel 4.15 Hasil Skenario 5	63
Tabel 4.16 Hasil Skenario 6	64
Tabel 4.17 Hasil Skenario 7	65
Tabel 4.18 Hasil Skenario 8	65
Tabel 4.19 Hasil Skenario 9	66
Tabel 4.20 Hasil Skenario 10	67

ABSTRAK

Latifaturrohman, Leny. 2026. **Sistem Rekomendasi Produk Skincare Berdasarkan Profil Jenis Kulit Wajah Menggunakan Pendekatan Content-Based Filtering**. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof. Dr. Suhartono, S.Si M.Kom, (II) Ahmad Fahmi Karami, M.Kom.

Kata Kunci: *Content-Based Filtering, Cosine Similarity, Skincare, Sistem Rekomendasi, TF-IDF.*

Perkembangan industri *skincare* yang semakin pesat menyebabkan pengguna dihadapkan pada banyak pilihan produk, namun kesulitan menentukan produk yang sesuai dengan jenis kulitnya. Penelitian ini bertujuan mengimplementasikan metode *Content-Based Filtering* (CBF) untuk membangun sistem rekomendasi produk *skincare* yang relevan dan personal berdasarkan kategori produk dan jenis kulit pengguna. Data yang digunakan merupakan dataset dari Kaggle yang berisi informasi produk dan ulasan pengguna. Tahapan penelitian meliputi *data cleaning, data integration, text normalization*, pembobotan teks menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), serta pengukuran kemiripan menggunakan *cosine similarity*. Sistem menghasilkan rekomendasi produk berdasarkan kemiripan konten ulasan pada kategori produk dan jenis kulit yang dipilih pengguna, kemudian menampilkan rekomendasi Top-K dengan nilai *similarity* tertinggi. Evaluasi sistem dilakukan menggunakan metrik *precision@3* dan *precision@5* pada seluruh kombinasi kategori produk dan jenis kulit dalam dataset yang menghasilkan 88 skenario pengujian. Hasil evaluasi menunjukkan nilai rata-rata *precision@3* sebesar 82,197% dan *precision@5* sebesar 76,591%, yang menunjukkan bahwa sebagian besar produk yang direkomendasikan termasuk dalam kategori relevan. Hasil penelitian membuktikan bahwa pendekatan *Content-Based Filtering* mampu memberikan rekomendasi produk *skincare* yang sesuai dengan karakteristik jenis kulit pengguna.

ABSTRACT

Latifaturohmah, Leny. 2026. **Skincare Product Recommendation System Based on Facial Skin Type Profile Using a Content-Based Filtering Approach**. Thesis. Informatics Engineering Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Advisors: (I) Prof. Dr. Suhartono, S.Si M.Kom, (II) Ahmad Fahmi Karami, M.Kom.

The rapid growth of the skincare industry has provided consumers with a wide variety of product choices; however, many users still face difficulties in selecting products that suit their skin type. This study aims to implement the Content-Based Filtering (CBF) method to develop a skincare recommendation system that provides relevant and personalized recommendations based on product categories and users' skin types. The dataset used in this study was obtained from Kaggle and contains product information and user reviews. The research stages included data cleaning, data integration, text normalization, feature weighting using Term Frequency–Inverse Document Frequency (TF-IDF), and similarity measurement using cosine similarity. The system generates product recommendations based on the similarity of review content within the selected product category and skin type, and then displays the Top-K recommendations with the highest similarity scores. System performance was evaluated using precision@3 and precision@5 metrics across all combinations of product categories and skin types, resulting in 88 evaluation scenarios. The evaluation results achieved an average precision@3 of 82.197% and precision@5 of 76.591%, indicating that most of the recommended products were relevant. The findings demonstrate that the Content-Based Filtering approach is capable of providing skincare product recommendations that match users' skin characteristics.

Keywords: Content-Based Filtering, Cosine Similarity, Recommendation System, Skincare, TF-IDF.

مستخلص البحث

لطيفة الرحمة، ليني. 2026. نظام توصية بمنتجات العناية بالبشرة اعتمادًا على نوع بشرة الوجه باستخدام منهجية الترشيح المعتمد على المحتوى (Content-Based Filtering). رسالة جامعية. قسم تقنية المعلومات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرفان: (١) الأستاذ الدكتور سوهارتونو S.Si، (٢) M.Kom أحمد فهمي كرامي، M.Kom

الكلمات المفتاحية: الترشيح المعتمد على المحتوى، نظام التوصية، العناية بالبشرة، TF-IDF، تشابه جيب التمام.

إن التطور السريع في صناعة منتجات العناية بالبشرة أدى إلى توفير عدد كبير من الخيارات للمستخدمين، إلا أنهم ما زالوا يواجهون صعوبة في اختيار المنتجات المناسبة لنوع بشرتهم. يهدف هذا البحث إلى تطبيق منهجية الترشيح المعتمد على المحتوى (Content-Based Filtering) لبناء نظام توصية بمنتجات العناية بالبشرة يقدم توصيات ذات صلة وشخصية اعتمادًا على فئة المنتج ونوع بشرة المستخدم. اعتمدت الدراسة على مجموعة بيانات من منصة Kaggle تحتوي على معلومات المنتجات ومراجعات المستخدمين. وشملت مراحل البحث تنظيف البيانات، ومجموعها، وتطبيع النصوص، واستخراج الأوزان باستخدام TF-IDF (Term Frequency-Inverse Document Frequency)، بالإضافة إلى قياس درجة التشابه باستخدام Cosine Similarity. يقوم النظام بإنتاج التوصيات اعتمادًا على تشابه محتوى المراجعات ضمن فئة المنتج ونوع البشرة اللذين يحددهما المستخدم، ثم يعرض أفضل Top-K من المنتجات ذات أعلى درجات التشابه. وتم تقييم أداء النظام باستخدام معياري Precision@3 و Precision@5 على جميع تراكيب فئات المنتجات وأنواع البشرة، مما أسفر عن 88 سيناريوًا للتقييم. وأظهرت النتائج متوسط Precision@3 بلغ 82.197٪ ومتوسط Precision@5 بلغ 76.591٪، مما يدل على أن معظم المنتجات الموصى بها كانت ذات صلة باحتياجات المستخدم. وتؤكد نتائج الدراسة أن منهجية الترشيح المعتمد على المحتوى قادرة على تقديم توصيات مناسبة لمنتجات العناية بالبشرة بما يتوافق مع خصائص نوع بشرة المستخدم.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan industri kecantikan, khususnya sektor *skincare*, menunjukkan pertumbuhan yang signifikan dalam satu dekade terakhir. Berdasarkan *Statista Market Insights*, nilai pasar global *skincare* diperkirakan mencapai USD 2,94 miliar pada tahun 2025 dan terus meningkat dengan *Compound Annual Growth Rate* (CAGR) sebesar 4,55% hingga 2030. Di Indonesia, diperkirakan nilai pasar industri kecantikan mencapai Rp146 triliun pada tahun 2024. Pertumbuhan ini didorong oleh meningkatnya kesadaran masyarakat terhadap kesehatan kulit, pengaruh media sosial, serta kemudahan akses melalui platform *e-commerce* (Geni, 2025).

Fenomena meningkatnya perhatian terhadap perawatan diri ini juga sejalan dengan nilai-nilai Islam yang menekankan pentingnya menjaga kesehatan dan kebersihan kulit wajah. Islam adalah agama yang sangat mengutamakan kebersihan. Itulah mengapa muslimah perlu menggunakan *skincare*. Hal ini diperkuat adanya hadis yang menyatakan, “Sesungguhnya Allah Maha Indah dan mencintai keindahan” (HR. Muslim). Begitu pula dalam firman Allah SWT dalam QS. Al-Baqarah ayat 222:

إِنَّ اللَّهَ يُحِبُّ التَّوَّابِينَ وَيُحِبُّ الْمُتَطَهِّرِينَ ﴿٢٢٢﴾

“Sesungguhnya Allah menyukai orang-orang yang bertaubat dan menyukai orang-orang yang mensucikan diri.” (QS. Al-Baqarah: 222)

Q.S. Al-Baqarah ayat 222 menegaskan pentingnya menjaga kebersihan dan kesehatan tubuh sebagai bagian dari ketaatan kepada Allah, baik secara fisik maupun spiritual. Ayat ini dipahami sebagai bentuk perhatian Islam terhadap kesehatan fisik, khususnya dengan menghindari aktivitas atau perlakuan yang berpotensi menimbulkan gangguan kesehatan pada kondisi tubuh tertentu, seperti saat haid, yang secara medis berkaitan dengan meningkatnya risiko iritasi dan infeksi (Aulia *et al.*, 2025). Penekanan pada pemahaman kondisi tubuh dan perlakuan yang tepat tersebut memiliki keterkaitan dengan perawatan kulit wajah, di mana setiap individu memiliki karakteristik kulit yang berbeda dan memerlukan produk yang sesuai agar kesehatan kulit tetap terjaga dan terhindar dari dampak negatif akibat penggunaan produk yang tidak tepat.

Sejalan dengan nilai tersebut, meskipun kesadaran akan pentingnya perawatan kulit terus meningkat, permasalahan utama yang dihadapi konsumen adalah kesulitan menentukan produk *skincare* yang sesuai karena ketidakmampuan menentukan jenis kulit, kondisi, dan permasalahan yang dialami, sehingga banyak yang mengalami kesalahan dalam pembelian produk dan penggunaannya (Kusumaningrum, 2021). Penelitian menunjukkan bahwa 44% pengguna kosmetik mengalami reaksi negatif akibat penggunaan produk *skincare* yang tidak tepat (Nayak *et al.*, 2023). Penggunaan produk *skincare* yang tidak tepat dapat menimbulkan permasalahan baru atau bahkan memperburuk kondisi kulit (Kusumaningrum, 2021). Hal ini menegaskan bahwa permasalahan ketidaksesuaian produk masih menjadi isu penting yang perlu diatasi dalam pemilihan produk *skincare*.

Meskipun demikian, sistem rekomendasi pada sebagian besar platform *e-commerce* saat ini masih berfokus pada popularitas dan rating umum pengguna tanpa mempertimbangkan faktor dermatologis seperti jenis kulit (Jannatin & Ratnasari, 2025). Hal ini menyebabkan rekomendasi yang dihasilkan sering kali bersifat umum dan tidak relevan secara personal. Akibatnya, pengguna berisiko memilih produk yang tidak cocok dan justru menimbulkan masalah kulit seperti iritasi, dehidrasi, atau jerawat (Ayuningrum *et al.*, 2024). Kondisi tersebut menunjukkan perlunya pengembangan sistem rekomendasi yang mampu mengintegrasikan aspek dermatologis dan preferensi pengguna untuk menghasilkan rekomendasi yang lebih akurat dan personal.

Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan pengembangan sistem rekomendasi produk *skincare* menggunakan pendekatan *Content-Based Filtering* (CBF). Pendekatan CBF dipilih karena mampu memberikan rekomendasi berdasarkan kesesuaian karakteristik produk dengan profil pengguna. Metode CBF menggunakan teknik *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk mengekstraksi fitur penting dari teks ulasan pengguna (review) dan *cosine similarity* untuk mengukur tingkat kesesuaian antara produk dengan profil jenis kulit pengguna. *Content-Based Filtering* terbukti efektif dan andal dalam sistem rekomendasi pemilihan *skincare* (Ardiana *et al.*, 2024).

Dengan menerapkan pendekatan *Content-Based Filtering*, penelitian ini diharapkan dapat menghasilkan sistem rekomendasi produk *skincare* yang lebih relevan dan personal berdasarkan jenis kulit wajah pengguna. Secara praktis, sistem ini dapat membantu pengguna dalam memilih produk *skincare* yang sesuai,

mengurangi risiko kesalahan pemilihan produk, serta meningkatkan kepuasan dalam penggunaan *skincare*. Bagi industri kosmetik, hasil penelitian ini dapat menjadi dasar dalam mengembangkan layanan rekomendasi berbasis personalisasi untuk meningkatkan loyalitas pelanggan dan keunggulan kompetitif di pasar digital.

1.2 Pernyataan Masalah

Berdasarkan uraian pada latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah bagaimana penerapan metode *Content-Based Filtering* (CBF) dalam membangun sistem rekomendasi produk *skincare* yang relevan dan personal berdasarkan profil jenis kulit wajah pengguna?

1.3 Batasan Masalah

Agar penelitian ini terfokus dan terarah, maka ditetapkan beberapa batasan masalah sebagai berikut:

1. Data yang digunakan merupakan data sekunder dari repositori publik Kaggle dengan judul “*Skincare Review Dataset*” yang dikumpulkan oleh Hafidah Musthaanah pada 18 September 2020 dan bersumber dari platform Female Daily Review.
2. Sistem rekomendasi yang dikembangkan menggunakan pendekatan *Content-Based Filtering* (CBF) dengan memanfaatkan data ulasan pengguna dan informasi jenis kulit wajah sebagai dasar pembentukan rekomendasi.
3. Evaluasi performa sistem dilakukan menggunakan metrik *precision@K* untuk menilai relevansi hasil rekomendasi yang diberikan oleh sistem.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan dari penelitian ini adalah mengimplementasikan metode *Content-Based Filtering* (CBF) dalam sistem rekomendasi produk *skincare* yang mampu menghasilkan rekomendasi yang relevan dan personal berdasarkan profil jenis kulit wajah pengguna.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Manfaat Teoritis

Penelitian ini diharapkan dapat memperkaya kajian akademik dalam bidang sistem rekomendasi melalui penerapan metode *Content-Based Filtering* dalam konteks rekomendasi produk *skincare*. Memberikan kontribusi metodologis bagi penelitian di bidang *personalized recommender systems*, terutama dalam konteks *health & beauty technology*.

2. Manfaat Praktis

- a. Bagi Pengguna: Membantu pengguna memilih produk *skincare* yang paling sesuai dengan jenis kulitnya, sehingga dapat mengurangi risiko ketidaksesuaian dan efek samping.
- b. Bagi Industri Kecantikan dan *E-commerce*: Menjadi solusi teknologi untuk meningkatkan kepuasan pelanggan, loyalitas merek, serta efektivitas promosi produk melalui sistem rekomendasi yang lebih personal.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Penelitian mengenai sistem rekomendasi produk *skincare* telah banyak dilakukan dengan berbagai pendekatan, khususnya menggunakan metode *Content-Based Filtering* (CBF) dan *Collaborative Filtering* (CF). Farah Ayu Fitriani (2022) melakukan penelitian sistem rekomendasi pemilihan produk *skincare* dengan pendekatan *Content-Based Filtering* menggunakan studi kasus review produk di Female Daily. Dataset berupa kumpulan review pengguna dimanfaatkan sebagai sumber utama rekomendasi. Proses rekomendasi dilakukan dengan menerapkan *preprocessing* teks, pembobotan TF-IDF, dan perhitungan *Cosine similarity* untuk menentukan kemiripan antar produk berdasarkan review. Hasil penelitian menunjukkan bahwa sistem menghasilkan 5 rekomendasi produk dengan nilai *Cosine similarity* tertinggi sebesar 0,875 berdasarkan satu produk acuan.. Namun, sistem ini masih bersifat *item-based*, sehingga rekomendasi yang dihasilkan sama untuk semua pengguna dan belum memperhatikan perbedaan karakteristik pengguna seperti jenis kulit.

Saskia Febrianti *et al.* (2024) mengembangkan model rekomendasi produk perawatan kulit wajah menggunakan metode *Content-Based Filtering* dengan mempertimbangkan atribut produk seperti komposisi bahan, tipe kulit, kategori, harga, dan merk. Dataset yang digunakan merupakan data internal dan preferensi pengguna yang diinput secara eksplisit. Sistem membangun profil pengguna dan

mencocokkannya dengan profil produk menggunakan pendekatan berbasis pengetahuan. Hasil eksperimen menunjukkan tingkat akurasi sebesar 88,89%. Meskipun penelitian ini sudah menerapkan personalisasi, sistem tidak memanfaatkan rekomendasi eksplisit dari pengguna lain sehingga belum menangkap pengalaman kolektif pengguna.

Penelitian oleh Viona Putri Ardiana *et al.* (2024) bertujuan membangun sistem rekomendasi pemilihan basic *skincare* menggunakan metode *Content-Based Filtering*. Dataset terdiri dari 35 produk basic *skincare* dengan enam atribut, dan TF-IDF digunakan untuk merepresentasikan atribut produk ke dalam bentuk vektor. *Cosine similarity* kemudian digunakan untuk menentukan kemiripan antar produk. Hasil penelitian menunjukkan bahwa sistem mampu merekomendasikan produk dengan nilai *Similarity* tertinggi sebesar 0,694. Namun, penelitian ini tidak melibatkan profil pengguna, sehingga rekomendasi masih bersifat umum dan tidak personal.

Putri *et al.* (2024) melakukan penelitian dengan tujuan mengembangkan sistem rekomendasi produk *skincare* yang lebih adaptif dengan mengombinasikan metode *Content-Based Filtering* dan *Collaborative Filtering* dalam pendekatan *hybrid filtering*. Dataset yang digunakan terdiri dari data produk *skincare*, ulasan pengguna, serta interaksi pengguna terhadap produk. Pada metode *Content-Based Filtering*, representasi konten produk dibangun menggunakan pembobotan TF-IDF, sedangkan pendekatan *Collaborative Filtering* diimplementasikan dengan memanfaatkan model *Convolutional Neural Network* (CNN) untuk mempelajari pola preferensi pengguna. Hasil pengujian menunjukkan bahwa sistem mampu

memberikan performa yang baik, khususnya pada skenario top-5 rekomendasi dengan nilai *precision* sebesar 0,776 serta tingkat kepuasan pengguna (*Customer Satisfaction Score/CSAT*) mencapai 80% dengan kategori *excellent*.

Ayuningrum *et al.* (2022) melakukan penelitian dengan tujuan membangun sistem rekomendasi produk *skincare* Korea berbasis web menggunakan metode *Collaborative Filtering*. Dataset yang digunakan terdiri dari 25 produk *skincare* dengan atribut kategori dan rating pengguna. Sistem bekerja dengan menghitung nilai kemiripan antar produk berdasarkan rating menggunakan *Similarity measure*, kemudian menghasilkan rekomendasi produk yang memiliki kemiripan tertinggi. Hasil evaluasi menggunakan *Mean Absolute Error* (MAE) menghasilkan nilai sebesar 0,25 yang menunjukkan performa sistem cukup baik. Namun, penelitian ini sangat bergantung pada data rating pengguna sehingga memiliki kelemahan dalam menangani permasalahan *cold-start* dan tidak mempertimbangkan karakteristik personal pengguna seperti jenis kulit.

Penelitian oleh Fajriansyah *et al.* (2021) meskipun berada pada domain film, memberikan kontribusi penting dalam pengembangan sistem rekomendasi berbasis *Content-Based Filtering* menggunakan TF-IDF dan *cosine similarity*. Penelitian ini menggunakan dataset besar sebanyak 4000 judul film dengan atribut sinopsis dan judul film. Sistem menghitung kemiripan antar item dan menghasilkan rekomendasi berdasarkan konten yang serupa. Evaluasi menggunakan MAP@K menunjukkan nilai akurasi yang tinggi, yaitu 0,82 untuk single query. Namun, pendekatan ini bersifat *item-to-item* dan tidak melibatkan karakteristik spesifik pengguna, sehingga personalisasi masih terbatas.

Tabel 2.1 Penelitian Terdahulu

No	Judul & Nama Peneliti (Tahun)	Metode	Hasil Penelitian	Perbedaan
1	Sistem Rekomendasi Pemilihan Produk <i>Skincare</i> dengan Pendekatan <i>Content-Based Filtering</i> Farah Ayu Fitriani (2022)	CBF + TF-IDF + <i>Cosine similarity</i>	Sistem menghasilkan 5 rekomendasi produk dengan nilai <i>Cosine similarity</i> tertinggi sebesar 0,875 berdasarkan satu produk acuan.	Penelitian ini menghasilkan rekomendasi berdasarkan kemiripan konten ulasan produk tanpa mempertimbangkan profil pengguna. Sementara itu, penelitian saya menambahkan personalisasi berdasarkan jenis kulit wajah pengguna.
2	Model Rekomendasi Produk Perawatan Kulit Wajah Menggunakan Metode <i>Content Based Filtering</i> (CBF) Saskia Febrianti et al. (2024)	CBF berbasis atribut produk	Model menunjukkan performa baik dengan tingkat akurasi sebesar 88,89%.	Penelitian ini menggunakan atribut produk sebagai dasar rekomendasi tanpa melibatkan karakteristik pengguna. Berbeda dengan itu, penelitian saya memanfaatkan jenis kulit pengguna sebagai faktor personalisasi.
3	Sistem Rekomendasi Pemilihan <i>Basic Skincare</i> Menggunakan Metode <i>Content Based Filtering</i> Viona Putri Ardiana et al. (2024)	CBF + TF-IDF + <i>Cosine similarity</i>	Produk dengan nilai kemiripan tertinggi memiliki <i>Similarity</i> sebesar 0,694.	Penelitian ini hanya merekomendasikan <i>skincare</i> dasar (<i>basic skincare</i>) dengan cakupan dataset yang terbatas pada 35 sampel produk. Sementara penelitian saya menggunakan dataset yang lebih besar dan memanfaatkan jenis kulit sebagai profil pengguna dalam proses personalisasi rekomendasi.
4	Sistem Rekomendasi <i>Skincare</i> Menggunakan Metode <i>Content Based Filtering</i> dan <i>Collaborative Filtering</i>	<i>Hybrid Filtering</i> (TF-IDF (CBF) + CNN (CF))	Hasil terbaik diperoleh pada top-5 rekomendasi dengan nilai <i>precision</i> 0,776 dan skor CSAT 80% (<i>Excellent</i>).	Fokus penelitian pada kombinasi metode CBF dan CF, bukan pada personalisasi berbasis jenis kulit pengguna secara spesifik. Sementara itu, penelitian saya menggunakan CBF

	Putri et al. (2024)			dengan fokus pada personalisasi jenis kulit.
5	Sistem Rekomendasi Produk <i>Skincare</i> Korea Berbasis Web Menggunakan Metode <i>Collaborative Filtering</i> Ayuningrum et al. (2022)	<i>Collaborative Filtering</i>	Sistem menghasilkan nilai MAE sebesar 0,25, namun bergantung pada rating pengguna.	Penelitian menggunakan pendekatan <i>Collaborative Filtering</i> dan bergantung pada rating pengguna. Sementara penelitian saya tidak bergantung pada rating, tetapi pada konten ulasan dan profil jenis kulit pengguna.
6	Sistem Rekomendasi Film Menggunakan <i>Content-Based Filtering</i> Fajriansyah et al. (2021)	CBF + TF-IDF + <i>Cosine similarity</i>	Sistem memperoleh nilai MAP@K sebesar 0,82, namun objek penelitian bukan <i>skincare</i> .	Penelitian ini menerapkan <i>Content-Based Filtering</i> pada domain film. Penelitian saya menerapkannya pada domain <i>skincare</i> dengan personalisasi jenis kulit.
Usulan Penelitian				
No	Judul	Metode	Hasil Penelitian	
7	Sistem Rekomendasi Produk <i>Skincare</i> Berdasarkan Profil Jenis Kulit Wajah Menggunakan Pendekatan <i>Content-Based Filtering</i>	CBF + TF-IDF + <i>Cosine similarity</i>	Diharapkan menghasilkan rekomendasi yang lebih relevan dengan mempertimbangkan kecocokan produk terhadap jenis kulit pengguna.	

Berdasarkan tabel 2.1, dapat diidentifikasi beberapa kesenjangan dalam pengembangan sistem rekomendasi produk *skincare* pada penelitian-penelitian terdahulu. Sebagian besar penelitian dengan pendekatan *Content-Based Filtering* masih berfokus pada kemiripan antar item atau kesamaan konten teks, sehingga rekomendasi yang dihasilkan cenderung bersifat umum dan belum sepenuhnya mempertimbangkan karakteristik pengguna. Penelitian yang telah menerapkan personalisasi umumnya hanya menggunakan preferensi eksplisit pengguna tanpa mengintegrasikannya secara langsung dengan karakteristik konten produk.

Akibatnya, sistem rekomendasi yang dihasilkan belum secara optimal merepresentasikan hubungan antara produk *skincare* dan kebutuhan kulit pengguna secara lebih spesifik.

Penelitian ini berupaya mengisi kesenjangan tersebut dengan mengembangkan sistem rekomendasi produk *skincare* menggunakan pendekatan *Content-Based Filtering* yang berfokus pada personalisasi berdasarkan jenis kulit wajah pengguna. Sistem rekomendasi dibangun dengan memanfaatkan informasi konten produk dan profil jenis kulit pengguna untuk menghasilkan rekomendasi yang lebih relevan dan sesuai dengan kebutuhan kulit. Dengan pendekatan ini, penelitian diharapkan dapat melengkapi penelitian-penelitian sebelumnya yang masih menghasilkan rekomendasi bersifat umum, serta memberikan solusi rekomendasi yang sederhana namun terarah pada domain *skincare*.

2.2 Skincare

Skincare adalah sebagai serangkaian produk perawatan kulit yang digunakan secara rutin untuk merawat dan melindungi kulit, terutama kulit wajah, dari kerusakan seperti paparan sinar UV, polusi, dan agen iritan lainnya (Noor *et al.*, 2018). Perawatan kulit bukan hanya aspek estetika, ahli dermatologi menyatakan bahwa kulit sehat berfungsi sebagai pelindung fisik dan imunologis, mengatur suhu tubuh, serta mencegah infeksi dan kerusakan akibat lingkungan eksternal seperti radikal bebas (Khaerunisa & Husain, 2025). Dari perspektif biologi kulit, lapisan epidermis dan dermis memerlukan perlindungan dan perawatan untuk mendukung regenerasi sel, yang berkontribusi pada tekstur dan

kekenyalan kulit. Kehilangan fungsi alami ini karena penuaan, stres, atau paparan lingkungan membuat penggunaan *skincare* menjadi penting.

Skincare merupakan serangkaian aktivitas merawat kulit wajah dengan menggunakan beberapa produk (Magfiroh, 2024). Dengan demikian ada lebih dari satu penggunaan produk dalam melakukan aktivitas penggunaan *skincare*. Ada beberapa kategori dasar pemakaian produk *skincare* (*basic skincare*) yang umum digunakan para pengguna *skincare* (Hariatiningsih, 2020), yaitu:

1. *Facial Wash*

Facial Wash adalah produk yang digunakan untuk membersihkan kotoran, minyak, dan sisa makeup dari permukaan kulit. Ada beberapa jenis *facial wash*, diantaranya yaitu *milk cleanser*, *balm cleanser*, *oil cleanser*, dan *micellar water*. *Facial Wash* harus memiliki pH yang sesuai agar tidak merusak lapisan pelindung kulit.

2. Toner

Toner berfungsi menyegarkan kulit, menyeimbangkan pH setelah mencuci wajah, serta membantu mengangkat sisa kotoran yang belum terangkat oleh *cleanser* sekaligus mempersiapkan kulit untuk produk selanjutnya. Toner tersedia dalam bentuk *hydrating toner* dan *exfoliating toner* (mengandung AHA/BHA), yang penggunaannya disesuaikan dengan jenis dan kondisi kulit.

3. *Moisturizer*

Moisturizer berperan menjaga kelembapan kulit, mengurangi kehilangan air (*Transepidermal Water Loss*), serta memperkuat *skin barrier* dari iritasi dan polutan. Pemilihan tekstur disesuaikan dengan jenis kulit, di mana *cream* atau

lotion cocok untuk kulit kering, sedangkan gel atau formula ringan lebih sesuai untuk kulit berminyak.

4. *Sunscreen*

Fungsi utama *sunscreen* adalah melindungi kulit dari sinar matahari yaitu UV-A dan UV-B untuk mencegah kerusakan kulit, penuaan dini, flek, dan risiko kanker kulit. Penggunaan SPF yang sesuai dengan intensitas paparan sinar matahari menjadi bagian yang sangat penting dalam rutinitas harian.

Kombinasi produk *basic skincare* ini telah menjadi “*skincare* minimalis” yang paling sering direkomendasikan oleh praktisi dan *influencer*. Misalnya dalam laporan MediaKom, jenis-jenis *skincare* seperti sabun muka, pelembab, toner, serum, dan *sunscreen* disebut sebagai jenis yang paling sering dibutuhkan oleh konsumen Indonesia berdasarkan kebutuhan kulit seperti kering, berminyak, atau berjerawat.

2.3 Jenis Kulit Wajah

Kulit wajah adalah lapisan tubuh yang sangat rentan terhadap perubahan lingkungan dan memiliki fungsi biologis penting, seperti regulasi suhu dan sensorik. Di sisi lain, kulit wajah juga memiliki karakteristik fisik dan fisiologis yang berbeda antar individu, seperti produksi sebum, kadar air, tekstur, pori, sensitivitas, dan respons terhadap iritan atau bahan aktif (Oliveira *et al.*, 2023). Oleh karena itu, pemilihan *skincare* yang tepat adalah yang sesuai dengan jenis kulitnya masing-masing. Secara umum, kulit wajah dapat dikategorikan menjadi beberapa tipe utama, yaitu:

1. Kulit Normal

Kulit normal biasanya memiliki keseimbangan antara kadar minyak dan kelembapan, dengan pori-pori yang tidak terlalu besar, tekstur halus, dan warna kulit merata. Kulit tipe ini umumnya lebih toleran terhadap berbagai bahan aktif dan jarang mengalami masalah serius, sehingga fleksibel dalam pemilihan produk *skincare* (Wulandari, 2019).

2. Kulit Berminyak (*Oily*)

Kulit berminyak ditandai oleh produksi sebum berlebih yang menyebabkan permukaan kulit tampak mengkilap, terutama di area dahi, hidung, dan dagu (T-zone). Pori-pori cenderung lebih besar dan kulit rentan terhadap pembentukan komedo maupun jerawat. Secara fisiologis, kadar minyak yang tinggi dapat memicu penyumbatan pori sehingga pemilihan produk yang bersifat ringan, non-komedogenik, dan oil-control sangat dianjurkan untuk tipe ini (Oliveira *et al.*, 2023).

3. Kulit Kering (*Dry*)

Kulit Kering memiliki karakteristik produksi sebum yang rendah sehingga kelembapan kulit tidak terjaga dengan baik. Kulit tampak kusam, terasa kencang, kadang disertai pengelupasan halus dan garis-garis halus lebih mudah terlihat. Secara fisiologis, kulit kering mengalami kehilangan air transepidermal lebih cepat, sehingga membutuhkan produk yang dapat menghidrasi sekaligus mengunci kelembapan (Oliveira *et al.*, 2023).

4. Kulit Kombinasi (*Combination*)

Kulit kombinasi menunjukkan ciri khas gabungan, di mana sebagian area wajah, terutama *T-zone*, cenderung berminyak, sementara area lain seperti pipi cenderung normal atau kering. Tipe ini menuntut perawatan yang lebih selektif karena setiap zona wajah memiliki kebutuhan berbeda. Misalnya, area berminyak membutuhkan produk pengontrol sebum, sementara area kering membutuhkan pelembap yang lebih kaya (Oliveira *et al.*, 2023). Hal ini membuat rekomendasi produk untuk kulit kombinasi harus bersifat adaptif dan seimbang.

Selain empat tipe kulit utama, terdapat kulit sensitif dan kulit mudah berjerawat yang umumnya dipandang sebagai kondisi tambahan karena karakteristik dan tingkat prevalensinya bervariasi pada setiap individu. Kulit sensitif ditandai dengan reaksi yang mudah muncul terhadap faktor lingkungan atau bahan tertentu, sedangkan kulit mudah berjerawat memiliki kecenderungan pori-pori tersumbat dan inflamasi. Pemilihan produk *skincare* yang sesuai dengan jenis kulit berpengaruh secara fisiologis terhadap kondisi kulit, seperti menjaga fungsi *skin barrier*, hidrasi, dan mencegah iritasi, sementara penggunaan produk yang tidak sesuai dapat menimbulkan gangguan kulit. Oleh karena itu, penelitian ini memfokuskan personalisasi rekomendasi pada empat tipe kulit dasar, yaitu normal, berminyak, kering, dan kombinasi, karena lebih umum dan memiliki ketersediaan data yang lebih representatif.

2.4 Sistem Rekomendasi

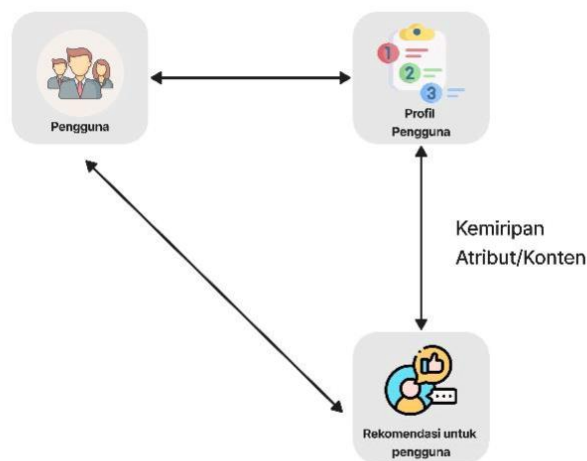
Sistem rekomendasi adalah teknologi yang dirancang untuk memberikan saran atau rekomendasi kepada pengguna dengan menganalisis pola preferensi, riwayat interaksi, serta karakteristik dari item atau produk yang tersedia (Triansyah Putra & Ronggo Alit, 2025). Tujuannya adalah membantu pengguna dalam membuat keputusan dengan menyajikan item-item yang mungkin bermanfaat. Sistem rekomendasi membantu mempersonalisasi pengalaman pengguna berdasarkan karakteristik individu (seperti preferensi sebelumnya, demografi, atau profil perilaku) dan mendukung pengambilan keputusan dengan menyediakan opsi-rekomendasi yang didasarkan pada data empiris (Lubis *et al.*, 2020). Misalnya, dalam platform *e-commerce*, rekomendasi dapat membantu pengguna menemukan barang yang mungkin tidak muncul di pencarian biasa tetapi sangat cocok berdasarkan preferensi mereka, sehingga meningkatkan kemungkinan konversi dan loyalitas pengguna.

Secara umum, sistem rekomendasi dapat diklasifikasikan ke dalam tiga teknik utama, yaitu *Content-Based Filtering*, *Collaborative Filtering*, dan *Knowledge-Based Filtering*. *Content-Based Filtering* bertujuan memberikan rekomendasi kepada pengguna berdasarkan karakteristik produk yang sesuai dengan preferensi atau minat pengguna tersebut. *Collaborative Filtering* bekerja dengan membangun basis data berupa matriks pengguna–item yang merepresentasikan preferensi pengguna terhadap suatu item, kemudian menghitung tingkat kemiripan antar pengguna untuk menghasilkan rekomendasi. Sementara itu, *Knowledge-Based Filtering* memberikan rekomendasi dengan memanfaatkan

pengetahuan domain tertentu yang menggambarkan hubungan antara fitur-fitur suatu item dan kebutuhan pengguna (Suhaillah & Hartatik, 2023).

2.5 Content-Based Filtering

Content-Based Filtering (CBF) merupakan salah satu metode utama dalam sistem rekomendasi yang bekerja dengan cara mencocokkan karakteristik suatu item dengan preferensi pengguna. Konsep dasarnya adalah bahwa pengguna akan menyukai item lain yang memiliki kesamaan fitur atau deskripsi dengan item yang sebelumnya mereka sukai. Sistem ini memanfaatkan data dari atribut atau deskripsi item seperti kategori, jenis, manfaat, dan kesesuaian dengan profil pengguna untuk membangun profil pengguna yang merepresentasikan minat dan kebutuhannya (Aggarwal, 2016).



Gambar 2.1 Skema *Content-Based Filtering*
(Sumber: Putri et al., 2020)

Gambar 2.1 menunjukkan skema kerja algoritma *Content-Based Filtering*. Sistem akan merekam profil pengguna dan kemudian akan memberikan rekomendasi sesuai dari jumlah kesamaan antara atribut konten dengan atribut yang

dimiliki oleh pengguna (B.Thorat *et al.*, 2015). Pada penelitian ini, sistem membentuk profil pengguna berdasarkan preferensi eksplisit yang diberikan, yaitu jenis kulit wajah dan kategori produk *skincare* yang diinginkan. Selanjutnya, sistem menghitung tingkat kesesuaian antara profil pengguna dengan karakteristik konten produk untuk menghasilkan rekomendasi. Hasil perhitungan tersebut digunakan untuk menentukan produk *skincare* yang memiliki tingkat kesesuaian tertinggi dengan profil pengguna.

Dalam konteks penelitian ini, metode CBF digunakan untuk memberikan rekomendasi produk *skincare* berdasarkan kesesuaian antara jenis kulit pengguna (normal, kering, berminyak, kombinasi) dengan karakteristik produk. Secara umum, metode CBF bekerja dalam beberapa tahapan penting, yaitu pembentukan profil pengguna, ekstraksi fitur produk, penghitungan kemiripan antar fitur, dan penyusunan daftar rekomendasi.

Metode *Content-Based Filtering* (CBF) memiliki sejumlah keunggulan, antara lain hanya memerlukan informasi konten item dan profil pengguna tanpa bergantung pada data dari pengguna lain. Pendekatan ini memungkinkan sistem untuk menganalisis kebutuhan pengguna secara lebih spesifik berdasarkan karakteristik yang relevan, seperti jenis kulit, kandungan bahan aktif, manfaat produk, dan deskripsi produk *skincare*. Pemanfaatan data deskriptif dan analisis konten dalam CBF terbukti efektif dalam menghasilkan rekomendasi produk *skincare* yang sesuai dengan kondisi dan kebutuhan individu. Selain itu, penggunaan atribut konten dan ulasan produk dapat meningkatkan ketepatan

rekomendasi, serta memberikan transparansi kepada pengguna mengenai alasan pemilihan produk yang direkomendasikan (Ardiana *et al.*, 2024).

2.6 Term Frequency–Inverse Document Frequency (TF-IDF)

TF-IDF (*Term Frequency–Inverse Document Frequency*) merupakan teknik pembobotan kata yang digunakan untuk merepresentasikan dokumen teks ke dalam bentuk vektor numerik berdasarkan tingkat kepentingan suatu kata di dalam dokumen dan keseluruhan koleksi dokumen. Konsep utama TF-IDF adalah bahwa kata yang sering muncul dalam satu dokumen namun jarang muncul di dokumen lain dianggap lebih informatif dalam merepresentasikan isi dokumen tersebut (Aggarwal, 2016).

TF-IDF memberikan kontribusi signifikan dalam proses pencarian dan pemeringkatan dokumen teks karena mampu menghasilkan representasi numerik yang efektif dari isi dokumen. Dengan memperhitungkan frekuensi kemunculan term dalam dokumen serta tingkat kemunculan term tersebut di seluruh koleksi dokumen, TF-IDF membantu membedakan kata-kata yang bersifat informatif dari kata-kata umum. Hal ini memungkinkan sistem untuk menyajikan hasil pencarian atau rekomendasi yang lebih relevan bagi pengguna. Selain meningkatkan relevansi hasil pencarian, TF-IDF juga memiliki keunggulan dalam hal skalabilitas dan fleksibilitas. Metode ini dapat diterapkan pada koleksi dokumen berukuran besar secara efisien dan digunakan dalam berbagai tugas pemrosesan bahasa alami, seperti ekstraksi fitur, klasifikasi teks, analisis sentimen, serta sistem rekomendasi berbasis konten untuk merepresentasikan karakteristik utama suatu dokumen teks (Septiani & Isabela, 2022).

2.7 Cosine Similarity

Cosine similarity merupakan metode yang digunakan untuk mengukur tingkat kemiripan antara dua dokumen teks dengan menghitung kosinus sudut antara dua vektor yang merepresentasikan dokumen tersebut. Setiap dokumen dipandang sebagai sebuah vektor dalam ruang berdimensi banyak, di mana setiap dimensi merepresentasikan fitur atau *term* yang terdapat dalam dokumen. Semakin kecil sudut yang terbentuk antara dua vektor, maka semakin besar nilai *cosine similarity* yang dihasilkan, yang menunjukkan bahwa kedua dokumen memiliki tingkat kemiripan yang tinggi. Sebaliknya, nilai *cosine similarity* yang mendekati nol menunjukkan bahwa kedua dokumen memiliki karakteristik yang berbeda (Fajriansyah *et al.*, 2021).

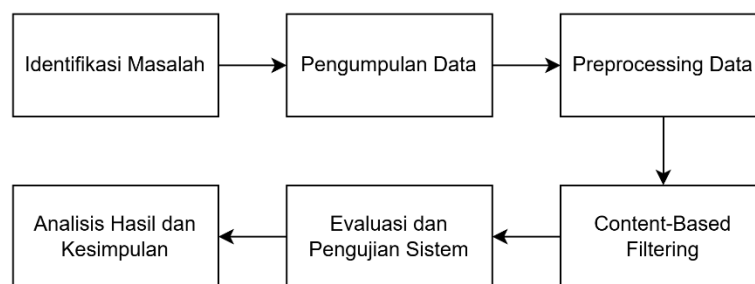
Dalam *text mining*, *cosine similarity* digunakan untuk mengukur kesamaan dokumen atau kalimat berdasarkan arah vektor fitur tanpa dipengaruhi oleh panjang dokumen, sehingga efektif untuk membandingkan teks dengan ukuran yang berbeda. Metode ini tidak mempertimbangkan makna semantik kata secara langsung, melainkan mengandalkan distribusi fitur dalam dokumen. Oleh karena itu, *cosine similarity* umumnya dikombinasikan dengan metode pembobotan TF-IDF, yang memberikan bobot numerik pada setiap *term* berdasarkan tingkat kepentingannya dalam dokumen dan keseluruhan koleksi dokumen. Kombinasi TF-IDF dan *cosine similarity* menghasilkan representasi vektor yang lebih informatif dan pengukuran kemiripan yang lebih akurat, khususnya dalam sistem rekomendasi berbasis konten (Azmi *et al.*, 2024).

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental yang berfokus pada pengembangan dan pengujian model sistem rekomendasi menggunakan data numerik yang dapat diukur secara objektif. Pendekatan ini bertujuan menghasilkan model rekomendasi berdasarkan pola dan hubungan antar variabel, serta memungkinkan evaluasi kinerja sistem melalui teknik komputasi, statistik, dan metrik tertentu agar hasil yang diperoleh dapat diinterpretasikan secara ilmiah. Desain penelitian mengacu pada model *Research and Development* (R&D) dalam pengembangan sistem berbasis data, yang dilaksanakan secara sistematis melalui tahapan analisis kebutuhan, perancangan model, implementasi, pengujian, dan evaluasi, di mana setiap tahap disusun secara logis dan saling berkaitan untuk mencapai tujuan penelitian.



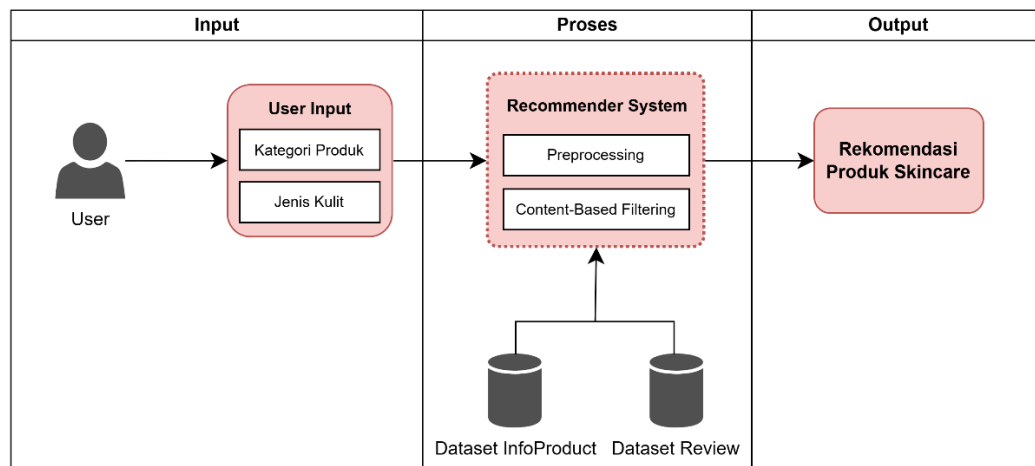
Gambar 3.1 Desain Penelitian

Pada gambar 3.1 menunjukkan tahapan penelitian diawali dengan identifikasi masalah, yang bertujuan untuk merumuskan permasalahan penelitian serta menentukan fokus kajian terkait kebutuhan sistem rekomendasi produk

skincare yang sesuai dengan jenis kulit pengguna. Langkah berikutnya adalah pengumpulan data, yaitu proses memperoleh dataset yang akan digunakan sebagai bahan analisis dan pemodelan sistem. Dataset yang digunakan harus memuat atribut produk dan informasi pengguna yang berkaitan dengan personalisasi rekomendasi, terutama variabel jenis kulit wajah sebagai faktor utama dalam sistem rekomendasi. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari repositori publik Kaggle. Selanjutnya, dilakukan *preprocessing* data yang mencakup proses *cleaning*, transformasi, integrasi, serta normalisasi agar data siap digunakan dalam pemodelan. Kemudian dilakukan perancangan sistem rekomendasi menggunakan metode *Content-Based Filtering* dan diimplementasikan menggunakan algoritma yang sesuai dan dilakukan pengujian untuk mengevaluasi kinerja sistem rekomendasi. Tahap terakhir adalah analisis hasil dan penarikan kesimpulan, yang bertujuan untuk menginterpretasikan hasil evaluasi sistem rekomendasi serta menilai ketercapaian tujuan penelitian berdasarkan kinerja model yang telah dibangun.

3.2 Desain Sistem

Desain sistem pada penelitian ini menggambarkan alur kerja sistem rekomendasi produk *skincare* yang diusulkan, mulai dari proses input pengguna hingga menghasilkan rekomendasi produk yang sesuai. Sistem dirancang dengan pendekatan *Content-Based Filtering* yang berfokus pada pemanfaatan karakteristik konten produk dan profil pengguna untuk menentukan kesesuaian rekomendasi.



Gambar 3.2 Desain Sistem

Gambar 3.2 menunjukkan alur kerja sistem rekomendasi produk *skincare* yang diusulkan dalam penelitian ini. Proses diawali dengan tahap input pengguna, di mana pengguna memasukkan preferensi berupa kategori produk *skincare* dan jenis kulit wajah. Informasi tersebut digunakan sebagai parameter awal untuk menentukan ruang pencarian rekomendasi yang sesuai dengan kebutuhan pengguna. Selanjutnya, sistem memanfaatkan dataset *skincare* sebagai sumber data utama yang berisi informasi produk dan ulasan pengguna. Dataset tersebut diproses melalui tahap *preprocessing* yang mencakup *data cleaning*, *data integration*, dan *text normalization* untuk menghasilkan data yang bersih dan siap digunakan dalam proses pemodelan. Tahap *preprocessing* bertujuan untuk memastikan kualitas dan konsistensi data sehingga proses rekomendasi dapat berjalan secara optimal. Pada tahap berikutnya, sistem menerapkan metode *Content-Based Filtering* dengan terlebih dahulu melakukan penyaringan produk berdasarkan kategori *skincare* dan jenis kulit yang dipilih pengguna. Dari hasil penyaringan tersebut, sistem memilih satu produk acuan secara acak yang digunakan sebagai dasar dalam proses

perhitungan kemiripan. Selanjutnya, sistem menghitung tingkat kemiripan antara produk acuan dengan seluruh produk lainnya menggunakan representasi TF-IDF dan metode *cosine similarity*. Hasil perhitungan tersebut digunakan untuk mengurutkan produk berdasarkan nilai *similarity* tertinggi, kemudian menampilkan rekomendasi Top-K yang paling sesuai dengan preferensi pengguna.

3.3 Pengumpulan Data

Pengumpulan data dalam penelitian ini menggunakan data sekunder yang diperoleh dari repositori publik Kaggle melalui dataset berjudul “*Skincare Review*” (Mustha’annah, 2020). Dataset ini dikumpulkan oleh Hafidah Musthaanah pada 18 September 2020, yang bersumber dari ulasan pengguna di platform Female Daily Review, salah satu komunitas daring terbesar di Indonesia untuk produk kecantikan dan perawatan kulit. Dataset ini menggunakan bahasa Indonesia sebagai bahasa utama pada teks ulasan, sehingga sangat relevan dengan konteks penelitian lokal.

Data yang digunakan tersedia dalam format *Comma Separated Values* (CSV) dan terdiri atas dua berkas utama, yaitu *InfoProduct.csv* dan *Review.csv*. Struktur atribut dari masing-masing berkas dataset ditunjukkan pada Tabel 3.1 dan Tabel 3.2.

Tabel 3.1 Struktur Dataset *InfoProduct.csv*

Atribut	Keterangan	Tipe Data
Category	Jenis kategori produk <i>skincare</i>	String
Merk	Nama merek atau produsen produk	String
Product	Nama produk <i>skincare</i>	String
Price	Harga produk	String
OverallRating	Nilai rata-rata rating produk dari seluruh pengguna	Float
Reviewer	Jumlah pengguna yang memberikan ulasan pada produk	Integer

Tabel 3.1 menunjukkan struktur atribut dari berkas InfoProduct.csv yang berisi informasi deskriptif mengenai produk *skincare* yang digunakan dalam penelitian ini. Dataset ini memuat 861 entri produk yang berasal dari 215 merek dan dikelompokkan ke dalam 22 kategori *skincare*. Atribut yang tersedia pada berkas ini mencakup kategori produk, nama merek, nama produk, harga, nilai rating rata-rata dari pengguna, serta jumlah pengguna yang memberikan ulasan.

Tabel 3.2 Struktur Dataset Review.csv

Atribut	Keterangan	Tipe Data
Product	Nama produk yang diulas	String
UserName	Nama pengguna yang memberikan ulasan	String
SkinCond_Age	Informasi gabungan antara jenis kulit dan rentang usia pengguna	String
Recommend	Indikasi apakah pengguna merekomendasikan produk tersebut	String
PostDate	Tanggal ulasan dipublikasikan	Date
Review	Isi teks ulasan pengguna terhadap produk	Text
Rating	Nilai rating yang diberikan pengguna (1–5)	Integer

Tabel 3.2 menunjukkan struktur atribut dari berkas Review.csv yang berisi data ulasan pengguna terhadap produk *skincare* yang terdapat dalam dataset. Dataset ini mencakup 8.662 data ulasan terhadap 818 produk *skincare*. Atribut yang tersedia meliputi nama produk yang diulas, nama pengguna, informasi jenis kulit yang digabung dengan rentang usia, indikator rekomendasi produk, tanggal ulasan, teks ulasan, serta nilai rating numerik.

Meskipun dataset InfoProduct.csv terdiri atas 6 atribut dan Review.csv terdiri atas 7 atribut, tidak seluruh atribut digunakan dalam proses penerapan metode *Content-Based Filtering*. Pemilihan atribut dilakukan berdasarkan relevansinya terhadap tujuan penelitian, yaitu membangun sistem rekomendasi produk *skincare* berbasis kemiripan konten ulasan. Oleh karena itu, hanya atribut

yang berperan dalam proses penyaringan data, pembentukan representasi produk, dan evaluasi sistem yang dipertahankan, sedangkan atribut lain yang tidak berkontribusi secara langsung terhadap proses rekomendasi tidak digunakan dalam perhitungan.

Adapun beberapa atribut pada dataset tidak digunakan dalam proses perhitungan *Content-Based Filtering* karena tidak berpengaruh terhadap pembentukan representasi konten maupun proses evaluasi sistem. Pada berkas InfoProduct.csv, atribut Merk, Price, OverallRating, dan Reviewer tidak digunakan dalam proses perhitungan TF-IDF maupun *cosine similarity*. Atribut Merk, Price, dan OverallRating hanya dimanfaatkan sebagai informasi tambahan yang ditampilkan pada hasil rekomendasi agar pengguna memperoleh informasi produk yang lebih lengkap, sedangkan atribut Reviewer tidak digunakan karena jumlah ulasan tidak memengaruhi tingkat kemiripan antarproduk. Pada berkas Review.csv, atribut UserName, PostDate, dan Rating tidak digunakan karena identitas pengguna, waktu publikasi ulasan, dan nilai rating tidak berkontribusi secara langsung terhadap pembentukan representasi teks maupun perhitungan kemiripan dalam pendekatan *Content-Based Filtering*. Ringkasan atribut yang digunakan dalam penelitian ini ditunjukkan pada Tabel 3.3.

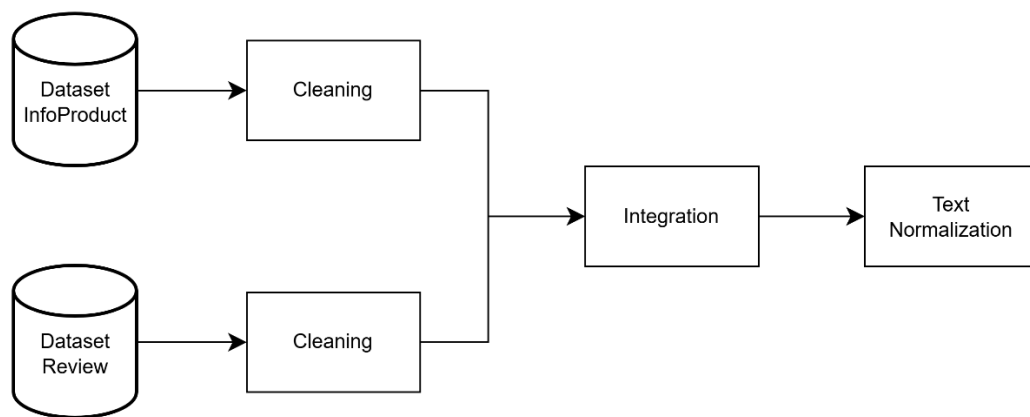
Tabel 3.3 Atribut yang Digunakan dalam Perhitungan

No	Atribut	Kegunaan dalam Penelitian (CBF)	Tipe Data	Sumber Dataset
1	Product	Identitas unik produk untuk merepresentasikan satu dokumen dalam pemodelan CBF.	String	InfoProduct dan Review
2	Category	Digunakan sebagai filter awal sesuai kebutuhan pengguna	String	InfoProduct
3	SkinType (hasil ekstraksi dari SkinCond_Age)	Digunakan sebagai dasar personalisasi rekomendasi berdasarkan jenis kulit wajah pengguna.	String	Review
4	Review	Digunakan sebagai sumber utama ekstraksi fitur menggunakan TF-IDF untuk membentuk vektor konten produk dan menghitung kesamaan antarproduk dengan <i>cosine similarity</i> .	Text	Review
5	Recommend	Digunakan sebagai dasar penentuan relevansi (user-defined ground truth) dalam evaluasi sistem menggunakan metrik <i>precision@K</i> .	String	Review

Tabel 3.3 menunjukkan atribut yang digunakan dalam penerapan metode *Content-Based Filtering* pada penelitian ini. Atribut Product digunakan sebagai identitas unik yang merepresentasikan setiap produk sebagai dokumen dalam proses CBF. Atribut Category digunakan sebagai mekanisme penyaringan awal (*pre-filtering*) untuk membatasi ruang pencarian rekomendasi agar tetap relevan dengan kebutuhan pengguna. Atribut Review merupakan fitur utama yang digunakan dalam proses perhitungan CBF melalui ekstraksi fitur TF-IDF dan pengukuran kemiripan menggunakan *cosine similarity*. Atribut SkinType berperan sebagai atribut personalisasi yang digunakan untuk menyesuaikan rekomendasi berdasarkan jenis kulit pengguna sebelum proses perhitungan kemiripan dilakukan. Selain itu, atribut Recommend digunakan sebagai dasar penentuan relevansi (*user-defined ground truth*) dalam proses evaluasi sistem menggunakan metrik *precision@K*.

3.4 Pre-processing Data

Pada tahap pengolahan data dilakukan serangkaian langkah-langkah *preprocessing* untuk mempersiapkan data mentah menjadi data yang siap diolah. Pada Gambar 3.3 adalah rincian mengenai tahapan *preprocessing*. Dataset InfoProduct dan Review dibersihkan terlebih dahulu, kemudian digabungkan melalui proses integrasi, dan dilanjutkan dengan normalisasi teks untuk menyiapkan data sebelum tahap pemodelan.



Gambar 3.3 Tahapan *Preprocessing*

3.4.1 Data Cleaning

Tahap pembersihan data (*data cleaning*) merupakan proses awal dalam pengolahan dataset agar data yang digunakan pada tahap pemodelan sistem rekomendasi memiliki kualitas yang baik, bebas dari duplikasi, dan memiliki format yang konsisten. Dalam penelitian ini, proses pembersihan dilakukan pada kedua berkas dataset, yaitu InfoProduct.csv dan Review.csv, untuk menyesuaikan struktur dan format data dengan kebutuhan analisis pada metode *Content-Based Filtering* (CBF).

1. Menghapus Data Duplikat

Pemeriksaan data duplikat dilakukan pada kolom Product dan Review untuk mengidentifikasi adanya entri ulasan yang identik terhadap produk yang sama. Baris data yang terdeteksi sebagai duplikat dihapus agar tidak memengaruhi representasi konten produk dalam proses ekstraksi fitur dan perhitungan kesesuaian pada metode *Content-Based Filtering*.

2. Menangani Nilai Kosong (*Missing Values*)

Dataset diperiksa untuk memastikan tidak terdapat nilai kosong pada kolom-kolom penting, seperti Product, Review, dan SkinCond_Age. Baris data yang memiliki nilai kosong pada kolom tersebut dihapus karena atribut-atribut tersebut menjadi dasar utama dalam analisis konten produk dan pembentukan profil pengguna. Nilai kosong pada kolom sekunder seperti Recommend dapat diisi dengan nilai default “*Unknown*” untuk menjaga konsistensi format.

3. Konversi Format Data

Atribut OverallRating pada dataset InfoProduct.csv dikonversi ke tipe data float agar memiliki format numerik yang konsisten. Proses konversi ini bertujuan untuk memastikan data memiliki struktur yang seragam dan dapat ditampilkan sebagai informasi tambahan pada hasil rekomendasi produk.

4. Pemisahan Kolom Gabungan (SkinCond_Age)

Kolom SkinCond_Age yang berisi gabungan antara jenis kulit dan rentang usia (misalnya “*Combination, 20–24*”) dipisahkan menjadi dua kolom, yaitu SkinType dan AgeRange. Kolom SkinType digunakan sebagai fitur utama

dalam personalisasi rekomendasi, sedangkan AgeRange tidak dilibatkan dalam proses perhitungan kesesuaian produk.

5. Validasi Format dan Rentang Nilai Data

Validasi dilakukan pada kolom SkinType untuk memastikan hanya memuat kategori jenis kulit yang valid, yaitu *oily*, *dry*, *combination*, dan *normal*. Tahap ini bertujuan untuk menjaga konsistensi data dalam proses pembentukan profil pengguna dan personalisasi rekomendasi.

3.4.2 Data Integration

Tahap integrasi data (*data integration*) merupakan proses penyatuan dua sumber data utama, yaitu InfoProduct.csv dan Review.csv, menjadi satu dataset terpadu yang siap digunakan dalam analisis dan pemodelan sistem rekomendasi. Proses integrasi dilakukan dengan menggabungkan kedua dataset berdasarkan atribut Product sebagai kunci utama (*primary key*). Teknik *inner join* digunakan sehingga hanya data yang memiliki kesesuaian nilai product pada kedua berkas yang dipertahankan. Dengan pendekatan ini, setiap ulasan pengguna dapat dihubungkan secara langsung dengan atribut produk yang bersangkutan. Produk yang tidak memiliki ulasan atau ulasan yang tidak memiliki referensi produk dikeluarkan dari dataset karena tidak relevan untuk analisis kesamaan konten pada metode *Content-Based Filtering*. Selain itu, dilakukan pengecekan kesesuaian penulisan nama produk dan format data untuk menghindari duplikasi serta ketidakkonsistenan antarentri.

Hasil dari tahap integrasi data ini ditunjukkan pada tabel 3.4 dan 3.5, yaitu berupa satu dataset yang memuat atribut input dan output.

Tabel 3.4 Tabel Hasil Integrasi Atribut Input dan Output

No	Nama Atribut	Tipe Data	Peran Atribut	Kegunaan dalam Sistem Rekomendasi
1	Category	String	Input	Filter kategori produk sesuai kebutuhan pengguna
2	SkinType	String	Input	Dasar pembentukan profil pengguna untuk personalisasi rekomendasi
3	Review	Text	Input (Proses)	Representasi konten produk yang digunakan dalam ekstraksi fitur TF-IDF
4	Recommend	String	Input (Proses)	Indikator ulasan positif sebagai informasi pendukung relevansi produk
5	Product	String	Input (Proses)	Kunci utama (<i>primary key</i>) untuk menghubungkan Informasi produk dan ulasan pengguna.
6	Merk	String	Output	Informasi pendukung untuk identifikasi produk
7	OverallRating	Float	Output	Informasi tambahan kualitas produk pada hasil rekomendasi
8	Price	String	Output	Informasi tambahan harga produk pada hasil rekomendasi

3.4.3 Text Normalization

Proses *text normalization* dilakukan untuk menstandarkan bentuk teks ulasan agar dapat diproses secara efektif pada tahap analisis *Content-Based Filtering (CBF)* menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*). Teks mentah yang berasal dari ulasan pengguna sering kali mengandung variasi penulisan, tanda baca, dan kata-kata tidak penting yang dapat mengganggu proses ekstraksi fitur (Aggarwal, 2016). Oleh karena itu, diperlukan tahapan normalisasi agar setiap kata memiliki bentuk dasar dan makna yang seragam.

a. *Case Folding*

Pada tahap *case folding* dilakukan perubahan seluruh huruf kapital menjadi huruf kecil (*lowercase*) (Triansyah Putra & Ronggo Alit, 2025). Tujuan dari tahap ini adalah untuk menyeragamkan karakter agar sistem tidak membedakan huruf besar dan kecil.

Tabel 3.5 Contoh Dokumen pada Tahap *Case Folding*

Sebelum	Sesudah
Untuk kulit kering kayak aku sih cocok banget pake ini dan pas dipake ga terlalu greasy gt. Efek melembabkannya pas bgt, jadi abis pake night cream ini, keesokan paginya aku ngerasa kulitku langsung kenyel2 gitu.	untuk kulit kering kayak aku sih cocok banget pake ini dan pas dipake ga terlalu greasy gt. efek melembabkannya pas bgt, jadi abis pake night cream ini, keesokan paginya aku ngerasa kulitku langsung kenyel2 gitu.

b. *Punctuation Removal*

Tahap ini menghapus tanda baca seperti titik, koma, tanda seru, tanda petik, dan simbol lain karena tidak memiliki makna semantic (Cheryl Ayu Melyani, 2022). Penghapusan karakter ini juga membantu mengurangi noise pada data agar tidak mengganggu perhitungan bobot kata pada tahap TF-IDF.

Tabel 3.6 Contoh Dokumen pada Tahap *Remove Punctuation*

Sebelum	Sesudah
untuk kulit kering kayak aku sih cocok banget pake ini dan pas dipake ga terlalu greasy gt. efek melembabkannya pas bgt, jadi abis pake night cream ini, keesokan paginya aku ngerasa kulitku langsung kenyel2 gitu.	untuk kulit kering kayak aku sih cocok banget pake ini dan pas dipake ga terlalu greasy gt efek melembabkannya pas bgt jadi abis pake night cream ini keesokan paginya aku ngerasa kulitku langsung kenyel gitu

c. *Stopword Removal*

Tahap ini bertujuan untuk menghapus kata-kata umum yang tidak memiliki makna penting dalam konteks analisis seperti “di”, “dan”, “aja”, “udah”, “yang”, “itu”, serta kata ganti orang seperti “aku”. Stopwords bertujuan untuk

menyederhanakan data teks tanpa mengurangi informasi yang diperlukan (Triansyah Putra & Ronggo Alit, 2025).

Tabel 3.7 Contoh Dokumen pada Tahap *Stopword Removal*

Sebelum	Sesudah
untuk kulit kering kayak aku sih cocok banget pake ini dan pas dipake ga terlalu greasy gt efek melembabkannya pas bgt jadi abis pake night cream ini keesokan paginya aku ngerasa kulitku langsung kenyel gitu	kulit kering cocok banget pake dipake greasy efek melembabkannya bgt abis pake night cream keesokan paginya ngerasa kulit langsung kenyel

d. *Tokenizing*

Pada tahap ini dilakukan proses pemisahan teks menjadi potongan-potongan kata yang disebut token (Yanisa Putri *et al.*, 2024). Setiap kata diperlakukan sebagai satu unit analisis yang akan dihitung frekuensinya. Tokenisasi dilakukan menggunakan metode pemisahan berbasis spasi.

Tabel 3.8 Contoh Dokumen pada Tahap *Tokenizing*

Sebelum	Sesudah
kulit kering cocok banget pake dipake greasy efek melembabkannya bgt abis pake night cream keesokan paginya ngerasa kulit langsung kenyel	["kulit","kering","cocok","banget","pake","dipake","greasy","efek","melembabkannya","bgt","abis","pake","night","cream","keesokan","paginya","ngerasa","kulit","langsung","kenyel"]

e. *Stemming*

Tahap *stemming* dilakukan untuk mengembalikan setiap kata ke bentuk dasarnya dengan menggunakan algoritma *Sastrawi Stemmer* (Yanisa Putri *et al.*, 2024). Tujuannya agar variasi kata yang memiliki akar sama seperti “melembabkan”, “lembabkan”, dan “kelembaban” diperlakukan sebagai kata yang sama, yaitu “lembab”.

Tabel 3.9 Contoh Dokumen pada Tahap *Stemming*

Sebelum	Sesudah
["kulit", "kering", "cocok", "banget", "pake", "dipake", "greasy", "efek", "melembabkannya", "bgt", "abis", "pake", "night", "cream", "keesokan", "paginya", "ngerasa", "kulit", "langsung", "kenyal"]	["kulit", "kering", "cocok", "banget", "pakai", "greasy", "efek", "lembab", "bgt", "abis", "pakai", "night", "cream", "pagi", "rasa", "kulit", "langsung", "kenyal"]

f. *Reconstruction*

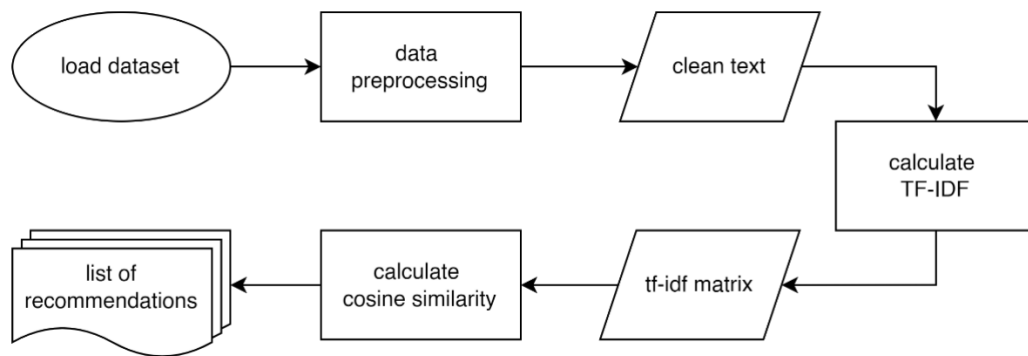
Setelah seluruh tahapan selesai, token hasil stemming disatukan kembali menjadi satu kalimat bersih. Teks hasil normalisasi ini kemudian digunakan sebagai masukan untuk proses ekstraksi fitur TF-IDF dalam metode *Content-Based Filtering*.

Tabel 3.10 Contoh Dokumen pada *Tahap Reconstruction*

Sebelum	Sesudah
Untuk kulit kering kayak aku sih cocok banget pake ini dan pas dipake ga terlalu greasy gt. Efek melembabkannya pas bgt, jadi abis pake night cream ini, keesokan paginya aku ngerasa kulitku langsung kenyal2 gitu.	kulit kering cocok banget pakai greasy efek lembab pagi rasa kulit langsung kenyal

3.5 *Content-Based Filtering*

Metode *Content-Based Filtering* (CBF) digunakan dalam penelitian ini untuk menghasilkan rekomendasi produk *skincare* berdasarkan kesesuaian antara karakteristik konten produk dan preferensi eksplisit pengguna. Pendekatan ini bekerja dengan menganalisis konten produk, khususnya teks ulasan pengguna yang telah melalui tahap *pre-processing*. Sehingga mampu merepresentasikan karakteristik, manfaat, dan pengalaman penggunaan produk secara lebih mendalam.



Gambar 3.4 Flowchart *Content-Based Filtering*
(Sumber: Fitria, 2024)

Gambar 3.4 menunjukkan proses *Content-Based Filtering* dalam penelitian ini yang diawali dengan pemuatan dataset *skincare* yang berisi informasi produk dan ulasan pengguna. Dataset tersebut terlebih dahulu melalui tahap *preprocessing* dan *text normalization* untuk menghasilkan teks ulasan yang bersih dan siap dianalisis. Selanjutnya, sistem membentuk profil pengguna berdasarkan kategori *skincare* dan jenis kulit wajah yang dipilih pengguna. Berdasarkan kedua informasi tersebut, sistem melakukan penyaringan dataset sehingga hanya produk yang memiliki kategori dan jenis kulit yang sesuai dengan input pengguna yang diproses lebih lanjut. Dari hasil penyaringan tersebut, sistem memilih satu produk acuan secara acak sebagai dasar dalam proses perhitungan kemiripan antarproduk. Setelah proses penyaringan dilakukan, teks ulasan dari seluruh produk yang terpilih dihitung bobot katanya menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Hasil perhitungan TF-IDF berupa matriks numerik yang merepresentasikan setiap produk *skincare* sebagai vektor berdasarkan konten ulasannya. Selanjutnya, sistem menghitung tingkat kemiripan antara produk acuan dengan seluruh produk lainnya menggunakan metode *cosine similarity*. Nilai *similarity* yang dihasilkan kemudian digunakan untuk mengurutkan produk

berdasarkan tingkat kemiripan tertinggi. Produk dengan nilai *similarity* tertinggi selanjutnya ditampilkan sebagai rekomendasi Top-K yang paling sesuai dengan kategori produk dan jenis kulit pengguna.

3.5.1 TF-IDF (*Term Frequency–Inverse Document Frequency*)

Metode *Term Frequency–Inverse Document Frequency* (TF-IDF) digunakan dalam penelitian ini untuk merepresentasikan teks ulasan produk *skincare* ke dalam bentuk vektor numerik. TF-IDF merupakan teknik pembobotan kata yang bertujuan untuk mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap seluruh dokumen dalam dataset. Metode ini efektif untuk menonjolkan kata-kata yang bersifat informatif dan membedakan karakteristik antar dokumen teks.

Pada tahap ini, teks ulasan produk yang telah melalui proses text normalization digunakan sebagai masukan dalam perhitungan TF-IDF. Sebelum dilakukan perhitungan, sistem terlebih dahulu menerapkan filter kategori produk dan jenis kulit (*SkinType*) sesuai dengan preferensi pengguna. Dengan demikian, hanya ulasan dari produk pada kategori dan jenis kulit yang dipilih pengguna yang diproses lebih lanjut, sehingga representasi konten produk yang dihasilkan tetap relevan dengan kebutuhan pengguna.

Secara umum, TF-IDF terdiri dari dua komponen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). *Term Frequency* (TF) menunjukkan seberapa sering suatu kata muncul dalam dokumen ke- j . Semakin tinggi nilai TF, semakin penting kata tersebut dalam konteks ulasan produk tersebut. *Term Frequency* (TF) dinyatakan dalam persamaan 3.1 (Roziqiin, 2024):

$$tf_{ij} = f_{ij} \quad (3.1)$$

Keterangan:

tf_{ij} = nilai *Term Frequency* dari term ke- i dalam dokumen ke- j

f_{ij} = jumlah kemunculan term ke- i pada dokumen ke- j

Setelah melakukan perhitungan TF, langkah berikutnya adalah menghitung *Inverse Document Frequency* (IDF), yaitu ukuran yang menunjukkan seberapa penting suatu kata dalam seluruh koleksi dokumen. Kata yang sering muncul di banyak dokumen akan memiliki nilai IDF rendah karena dianggap kurang informatif. Rumus perhitungan IDF ditunjukkan pada persamaan 3.2 (Lubis *et al.*, 2020):

$$idf_i = \ln \left(\frac{N}{df_i} \right) \quad (3.2)$$

Keterangan:

idf_i = nilai *Inverse Document Frequency* untuk term t

N = jumlah total dokumen ulasan

df_i = jumlah dokumen yang mengandung term t

Perhitungan IDF menggunakan logaritma natural ($\ln$), sebagaimana diterapkan dalam pustaka *Scikit-learn* versi *TF-IDF Vectorizer*. Nilai IDF yang tinggi menunjukkan bahwa kata tersebut jarang muncul dan dianggap lebih penting dalam membedakan antar dokumen. Tahap selanjutnya adalah menghitung nilai TF-IDF, yaitu kombinasi antara frekuensi kemunculan kata dalam dokumen (TF) dan tingkat kepentingan kata dalam seluruh dokumen (IDF). Rumus perhitungan TF-IDF ditunjukkan pada persamaan 3.3 (Lubis *et al.*, 2020):

$$w_{ij} = tf_{ij} \times idf_i \quad (3.3)$$

Keterangan:

w_{ij} = bobot TF-IDF dari term ke- i pada dokumen ke- j

tf_{ij} = nilai *Term Frequency* dari term ke- i pada dokumen ke- j

idf_i = nilai *Inverse Document Frequency* dari term t

Hasil dari proses ini adalah matriks TF-IDF, di mana setiap baris merepresentasikan satu produk *skincare* dan setiap kolom merepresentasikan kata unik yang muncul dalam kumpulan ulasan. Nilai pada matriks tersebut menunjukkan bobot kepentingan masing-masing kata terhadap produk tertentu. Matriks TF-IDF ini selanjutnya digunakan sebagai dasar dalam perhitungan tingkat kemiripan antarproduk pada tahap *cosine similarity*.

3.5.2 *Cosine similarity*

Metode *cosine similarity* digunakan dalam penelitian ini untuk mengukur tingkat kesamaan antar produk *skincare* berdasarkan representasi vektor TF-IDF dari teks ulasan pada kategori dan jenis kulit yang dipilih pengguna. Setelah direpresentasikan dalam bentuk vektor TF-IDF, tahap berikutnya adalah menghitung tingkat kemiripan antarproduk menggunakan *cosine similarity*, sebagaimana ditunjukkan pada persamaan 3.4 (Cheryl Ayu Melyani, 2022):

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (3.4)$$

Keterangan:

A dan B = dua vektor produk yang dibandingkan

A_i dan B_i = nilai komponen ke- i dari vektor A dan B

$A \cdot B$ = hasil perkalian dot Product antara dua vektor

$|A|$ dan $|B|$ = panjang (norma) masing-masing vektor

Nilai *Cosine similarity* berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan tingkat kemiripan yang tinggi antar produk, sedangkan nilai mendekati 0 menunjukkan tingkat kemiripan yang rendah. Dalam konteks sistem rekomendasi ini, semakin tinggi nilai *cosine similarity* antara produk,

semakin besar kemungkinan produk tersebut memiliki karakteristik konten yang serupa berdasarkan ulasan pengguna.

Hasil perhitungan *cosine similarity* selanjutnya digunakan untuk mengurutkan produk *skincare* berdasarkan tingkat kemiripan tertinggi terhadap produk acuan. Produk dengan nilai *similarity* tertinggi dipilih sebagai kandidat rekomendasi dan ditampilkan kepada pengguna sebagai hasil akhir sistem rekomendasi.

3.6 Evaluasi dan Pengujian Sistem

Tahap evaluasi dan pengujian sistem bertujuan untuk menilai sejauh mana sistem rekomendasi yang dibangun mampu menghasilkan rekomendasi produk *skincare* yang relevan sesuai dengan preferensi pengguna, khususnya berdasarkan jenis kulit wajah. Evaluasi dilakukan dengan mengukur tingkat kesesuaian antara produk yang direkomendasikan oleh sistem dengan kebutuhan pengguna yang direpresentasikan melalui data ulasan.

Pada penelitian ini, metrik evaluasi yang digunakan adalah *precision@K*, dengan nilai $K = 3$ dan $K = 5$. *Precision@K* digunakan untuk mengukur proporsi produk yang relevan dari sejumlah K produk teratas yang direkomendasikan oleh sistem. Metrik ini dipilih karena mampu memberikan gambaran sederhana mengenai kualitas rekomendasi pada posisi teratas, yang merupakan bagian paling penting bagi pengguna dalam sistem rekomendasi.

Precision merupakan metrik yang digunakan untuk mengukur proporsi item relevan dalam hasil rekomendasi yang diberikan kepada pengguna (Aggarwal, 2016). Selain itu, *precision@K* dinilai efektif dalam mengevaluasi sistem

rekomendasi berbasis Top-K karena berfokus pada kualitas hasil rekomendasi yang ditampilkan kepada pengguna (Bałchanowski & Boryczka, 2023). Oleh karena itu, penggunaan *precision@K* dalam penelitian ini dianggap sesuai untuk mengukur performa sistem rekomendasi produk *skincare*.

Dalam penelitian ini, penentuan relevansi produk dilakukan berdasarkan data ulasan pengguna. Suatu produk dianggap relevan untuk jenis kulit tertentu apabila mayoritas ulasan pengguna dengan jenis kulit yang sama memberikan indikator Recommend = Yes terhadap produk tersebut. Apabila jumlah rekomendasi positif dan negatif seimbang, maka produk tersebut dianggap tidak relevan untuk jenis kulit tersebut dan tidak dimasukkan dalam perhitungan evaluasi. *Precision@K* menghitung seberapa banyak produk relevan muncul di antara *K* rekomendasi teratas, yang dinyatakan dalam rumus 3.5 (Aggarwal, 2016):

$$Precision@K = \frac{\text{Jumlah item relevan dalam Top-}K}{K} \quad (3.5)$$

Keterangan:

K = jumlah item teratas yang direkomendasikan

Dalam proses pengujian, dilakukan 88 skenario yang merupakan kombinasi dari 22 kategori *skincare* dan 4 jenis kulit, yaitu *oily*, *dry*, *normal*, dan *combination*. Pada setiap skenario, sistem memilih satu produk acuan secara acak dari seluruh produk yang memenuhi kategori dan jenis kulit yang diuji, kemudian menghasilkan rekomendasi Top-K berdasarkan nilai *cosine similarity* tertinggi. Selanjutnya, dihitung jumlah produk yang termasuk dalam kategori relevan pada hasil rekomendasi untuk memperoleh nilai *precision@3* dan *precision@5*.

Nilai *precision@3* dan *precision@5* dihitung pada setiap skenario pengujian untuk mengukur kualitas hasil rekomendasi yang dihasilkan sistem. Nilai

rata-rata *precision* digunakan sebagai indikator kinerja sistem dalam menghasilkan rekomendasi yang relevan. Semakin tinggi nilai *precision* yang diperoleh, maka semakin baik kinerja sistem rekomendasi dalam memberikan hasil yang sesuai dengan kebutuhan pengguna.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Pre-processing Data

Pada penelitian ini, preprocessing data dilakukan melalui beberapa tahapan, yaitu *data cleaning*, *data integration*, dan *text normalization*. Setiap tahapan memiliki peran penting dalam memastikan kualitas data yang digunakan dalam sistem rekomendasi.

4.1.1 Data Cleaning

Pada dataset InfoProduct, proses pembersihan dilakukan dengan mengidentifikasi duplikasi berdasarkan atribut Product. Ditemukan sebanyak 102 data duplikat yang berasal dari 38 produk yang sama, sehingga dilakukan penghapusan duplikasi agar setiap produk hanya direpresentasikan satu kali. Setelah proses tersebut, jumlah data berkurang dari 861 menjadi 797 data. Selain itu, atribut OverallRating telah dipastikan dalam format numerik (float), serta atribut Price dinormalisasi dengan mengganti nilai yang tidak valid seperti nilai kosong atau tidak realistis menjadi “Unknown”. *Missing value* pada atribut Reviewer juga telah ditangani dengan mengganti nilai kosong menjadi 0.

Pada dataset Review, proses pembersihan diawali dengan penghapusan data yang tidak memiliki teks ulasan (Review) dan data tanpa informasi SkinCond_Age, sehingga jumlah data berkurang dari 8662 menjadi 8566 data. Selanjutnya, dilakukan identifikasi duplikasi berdasarkan kombinasi atribut Product dan Review, di mana ditemukan 283 data duplikat yang membentuk 83 pasangan ulasan

identik. Setelah dilakukan penghapusan duplikasi, jumlah data menjadi 8366. Atribut Recommend kemudian dinormalisasi menjadi tiga kategori, yaitu Yes, No, dan Unknown untuk memastikan konsistensi data.

SkinType	
combination	count
combination	3758
oily	2194
dry	1097
normal	846

Gambar 4.1 Distribusi Jenis Kulit

Gambar 4.1 menunjukkan ekstraksi jenis kulit (SkinType) dari atribut SkinCond_Age, sehingga diperoleh empat kategori utama, yaitu *combination*, *oily*, *dry*, dan *normal*. Data yang tidak termasuk dalam kategori tersebut dieliminasi, sehingga jumlah data akhir menjadi 7895. Distribusi jenis kulit menunjukkan bahwa kategori *combination* memiliki jumlah data terbanyak sebanyak 3758 data, diikuti oleh *oily* sebanyak 2194 data, *dry* sebanyak 1097 data, dan *normal* sebanyak 846 data.

4.1.2 Data Integration

Pada tahap ini dilakukan penggabungan dataset *Review* dan *InfoProduct* untuk membentuk satu dataset yang terintegrasi.

```
merged_df = pd.merge(
    review_df,
    info_df,
    on='Product',
    how='inner'
```

Gambar 4.2 Pseudocode Merge Dataset

Gambar 4.2 adalah proses penggabungan dilakukan menggunakan atribut Product sebagai *primary key* dengan metode *inner join*, sehingga hanya data yang memiliki pasangan pada kedua dataset yang akan digunakan. Sebelum proses integrasi, terdapat 49 data pada dataset Review yang tidak memiliki pasangan pada dataset InfoProduct, sehingga data tersebut tidak ikut tergabung dalam proses integrasi.

	Product	Merk	Category	SkinType	Review	Recommend	OverallRating	Price
0	Perfect 3D Gel	Hada Labo	Night Cream	normal	ga cocok bgttt bikin muka bruntusan dan lgsg m...	No	3.8	Rp. 91.000
1	Perfect 3D Gel	Hada Labo	Night Cream	combination	[SOLD]baru pertama kali nyoba dan beli karna l...	No	3.8	Rp. 91.000
2	Perfect 3D Gel	Hada Labo	Night Cream	dry	enak sih di mukaku, yg tadinya kering jd lemb...	No	3.8	Rp. 91.000
3	Perfect 3D Gel	Hada Labo	Night Cream	oily	SOLD OUT [PRELOVED] Halo aku preloved ini ya, ...	Yes	3.8	Rp. 91.000
4	Perfect 3D Gel	Hada Labo	Night Cream	oily	awalnya sih pas udah beli pembersih Hada Labo ...	Yes	3.8	Rp. 91.000

Gambar 4.3 Hasil Integrasi Data

Gambar 4.3 merupakan hasil dari proses integrasi menghasilkan dataset merged_df dengan jumlah 7.846 baris data dan 13 atribut. Dataset ini telah menggabungkan informasi ulasan pengguna dengan informasi produk, seperti Merk, Category, Price, dan OverallRating. Selanjutnya, dilakukan seleksi atribut untuk menyesuaikan kebutuhan penelitian. Atribut yang digunakan meliputi Product, Merk, Category, SkinType, Review, Recommend, OverallRating, dan Price. Hasil seleksi ini menghasilkan dataset dengan jumlah 7.846 baris data dan 8 atribut utama.

4.1.3 Text Normalization

Pada tahap ini dilakukan normalisasi teks pada kolom *Review* untuk menghasilkan teks yang bersih dan siap digunakan dalam proses pembobotan kata. Proses normalisasi dilakukan melalui beberapa tahapan sebagai berikut:

a. *Case Folding*

Pada penelitian ini, proses case folding dilakukan dengan menggunakan fungsi `.lower()` untuk mengubah seluruh teks menjadi huruf kecil.

```
sample_df['Case_Folding'] = sample_df['Review'].str.lower()
```

Gambar 4.4 Pseudocode Case Folding

Hasil dari proses ini adalah seluruh teks pada kolom Review menjadi seragam dalam huruf kecil seperti pada tabel 4.1, sehingga tidak terjadi perbedaan representasi kata akibat penggunaan huruf kapital.

Tabel 4.1 Hasil *Case Folding*

Sebelum	Sesudah
Dengan harga sekian, kualitas oke. Bisa dipake buat melembabkan kulit atau make up lock. Aku pake ini make up jadi awet meski udah wudhu. Sayangnya kalau pake kebanyakan bikin kulit wajah aku lengket Harumnya menyegarkan	dengan harga sekian, kualitas oke. bisa dipake buat melembabkan kulit atau make up lock. aku pake ini make up jadi awet meski udah wudhu. sayangnya kalau pake kebanyakan bikin kulit wajah aku lengket harumnya menyegarkan

b. *Punctuation Removal*

Proses penghapusan URL, tanda baca, dan angka dilakukan menggunakan fungsi `re.sub()` dari *library regex*.

```
def remove_punctuation(text):
    text = re.sub(r'http\S+|www\S+', '', text) # hapus URL
    text = re.sub(r'^a-z\s', ' ', text)       # hapus
    simbol & angka
    return text
sample_df['Remove_Punctuation'] =
sample_df['Case_Folding'].apply(remove_punctuation)
```

Gambar 4.5 Pseudocode Remove Punctuation

Tabel 4.2 merupakan hasil dari proses ini yaitu teks yang hanya terdiri dari huruf dan spasi, sehingga elemen yang tidak relevan seperti simbol dan angka telah dihilangkan.

Tabel 4.2 Hasil *Punctuation Removal*

Sebelum	Sesudah
dengan harga sekian, kualitas oke. bisa dipake buat melembabkan kulit atau make up lock. aku pake ini make up jadi awet meski udah wudhu . sayangnya kalau pake kebanyakan bikin kulit wajah aku lengket harumnya menyegarkan	dengan harga sekian kualitas oke bisa dipake buat melembabkan kulit atau make up lock aku pake ini make up jadi awet meski udah wudhu sayangnya kalau pake kebanyakan bikin kulit wajah aku lengket harumnya

c. *Tokenizing*

Proses *tokenizing* dilakukan dengan memecah teks menjadi daftar kata menggunakan fungsi `.split()`. Normalisasi karakter berulang dilakukan menggunakan fungsi `re.sub()` untuk menghilangkan pengulangan huruf berlebih dalam kata.

```
sample_df['Tokenizing'] =
sample_df['Remove_Punctuation'].apply(lambda x: x.split())
def normalize_repeated_chars(word):
    return re.sub(r'(\w)\1{2,}', r'\1', word)
sample_df['Normalized'] = sample_df['Tokenizing'].apply(
    lambda tokens: [normalize_repeated_chars(t) for t in
tokens]
)
```

Gambar 4.6 *Pseudocode Tokenizing* dan Normalisasi Karakter Berulang

Tabel 4.3 merupakan hasil dari proses ini yaitu teks yang semula berupa kalimat diubah menjadi kumpulan kata (token) yang siap diproses pada tahap berikutnya. Kata-kata yang memiliki pengulangan karakter berlebih menjadi lebih sederhana dan konsisten.

Tabel 4.3 Hasil *Tokenizing*

Sebelum	Sesudah
dengan harga sekian kualitas oke bisa dipake buat melembabkan kulit atau make up lock aku pake ini make up jadi awet meski udah wudhu sayangnya kalau pake kebanyakan bikin kulit wajah aku lengket harumnya menyegarkan	[dengan,harga,sekian,kualitas,oke,bisa,dipake,buat,melembabkan,kulit,atau,make,up,lock,aku,pake,ini,make,up,jadi,awet,meski,udah,wudhu,sayangnya,kalau,pake,kebanyakan,bikin,kulit,wajah,aku,lengket,harumnya,menyegarkan]

d. *Stopword Removal*

Penghapusan *stopword* dilakukan menggunakan daftar *stopword* Bahasa Indonesia dari library NLTK.

```
sample_df['Stopword_Removal'] =
sample_df['Normalized'].apply(
    lambda tokens: [t for t in tokens if t not in
stop_words]
)
```

Gambar 4.7 *Pseudocode Stopword Removal*

Hasil dari proses ini tertera pada tabel 4.4 yaitu hanya kata-kata penting yang dipertahankan, sedangkan kata-kata umum yang tidak memiliki makna signifikan dihilangkan.

Tabel 4.4 Hasil *Stopword Removal*

Sebelum	Sesudah
[dengan,harga,sekian,kualitas,oke,bisa,dipake,buat,melembabkan,kulit,atau,make,up,lock,aku,pake,ini,make,up,jadi,awet,meski,udah,wudhu,sayangnya,kalau,pake,kebanyakan,bikin,kulit,wajah,aku,lengket,harumnya,menyegarkan]	[harga,sekian,kualitas,oke,dipake,melembabkan,kulit,make,up,lock,pake,make,up,awet,udah,wudhu,sayangnya,pake,kebanyakan,bikin,kulit,wajah,lengket,harumnya,menyegarkan]

e. *Stemming*

Proses *stemming* dilakukan menggunakan *library* Sastrawi untuk mengubah kata menjadi bentuk dasarnya.

```
sample_df['Stemming'] = sample_df['Stopword_Removal'].apply(
    lambda tokens: [stemmer.stem(t) for t in tokens]
)
```

Gambar 4.8 *Pseudocode Stemming*

Hasil dari proses ini tertera pada tabel 4.5 yaitu kata-kata dalam teks telah disederhanakan ke bentuk dasar sehingga variasi kata dapat direpresentasikan secara seragam.

Tabel 4.5 Hasil *Stemming*

Sebelum	Sesudah
[harga,sekian,kualitas,oke,dipake,melembabkan,kulit,make,up,lock,pake,make,up,awet,udah,wudhu,sayangnya,pake,kebanyakan,bikin,kulit,wajah,lengket,harumnya,menyegarkan]	[harga,sekian,kualitas,oke,dipake,melembabkan,kulit,make,up,lock,pake,make,up,awet,udah,wudhu,sayang,pake,banyak,bikin,kulit,wajah,lengket,harum,segar]

f. *Reconstruction*

Pada tahap ini, token yang telah diproses digabung kembali menjadi teks menggunakan fungsi *.join()*.

```
sample_df['Reconstruction'] = sample_df['Stemming'].apply(
    lambda tokens: ' '.join(tokens)
)
```

Gambar 4.9 *Pseudocode Reconstruction*

Hasil dari proses ini tertera pada tabel 4.6 yaitu teks yang telah dinormalisasi secara lengkap dan disimpan dalam kolom baru *Clean_Review*.

Tabel 4.6 Hasil *Reconstruction*

Sebelum	Sesudah
[harga,sekian,kualitas,oke,dipake,melembabkan,kulit,make,up,lock,pake,make,up,awet,udah,wudhu,sayang,pake,banyak,bikin,kulit,wajah,lengket,harum,segar]	harga sekian kualitas oke dipake melembabkan kulit make up lock pake make up awet udah wudhu sayang pake banyak bikin kulit wajah lengket harum segar

Seluruh tahapan normalisasi diterapkan pada dataset menggunakan fungsi *preprocess_text()*. Setelah normalisasi, dilakukan penggabungan seluruh *Clean_Review* berdasarkan produk dan jenis kulit menggunakan fungsi *groupby()*. Hasil dari proses ini adalah setiap produk berdasarkan jenis kulit direpresentasikan dalam satu dokumen gabungan, sehingga jumlah data berkurang menjadi 2.590 dan sistem dapat membangun representasi konten yang lebih kaya dan kontekstual terhadap pengalaman pengguna dengan jenis kulit tertentu. Proses ini juga

menghasilkan kolom `Clean_Review` yang berisi teks ulasan yang telah dibersihkan, dinormalisasi, dan siap digunakan pada tahap pembobotan TF-IDF.

	Product	Merk	Category	SkinType	OverallRating	Price	Clean_Review
0	100% Organic Cold-Pressed Rose Hip Seed Oil	The Ordinary	Face Oil	combination	4.1	Unknown	face oil turut bagus ya cocok tipe kulit acne ...
1	100% Organic Cold-Pressed Rose Hip Seed Oil	The Ordinary	Face Oil	dry	4.1	Unknown	temen nyaranin beli produk to km ga macem ing...
2	100% Organic Cold-Pressed Rose Hip Seed Oil	The Ordinary	Face Oil	normal	4.1	Unknown	sebenarnya produk bagus banget banget banget k...
3	100% Organic Cold-Pressed Rose Hip Seed Oil	The Ordinary	Face Oil	oily	4.1	Unknown	baca ttg manfaat rosehip oil bagus banget putu...
4	2.5	Benzolac	Acne Treatment	combination	3.4	Rp. 15.000	pakai produk inget bgt udh yg pakai tekstur ge...
5	2.5	Benzolac	Acne Treatment	dry	3.4	Rp. 15.000	suka banget bikin jerawat cpt kering panas pag...
6	2.5	Benzolac	Acne Treatment	normal	3.4	Rp. 15.000	tau produk internet coba emang bagus ga cepet ...

Gambar 4.10 Hasil *Pre-processing* Data

4.2 Pembobotan Kata (TF-IDF)

Pada tahap ini dilakukan pembobotan kata menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk merepresentasikan setiap dokumen produk dalam bentuk vektor numerik.

```
# Inisialisasi TF-IDF
tfidf = TfidfVectorizer(
    max_df=0.9,          # abaikan kata terlalu umum
    min_df=2,            # abaikan kata sangat jarang
    ngram_range=(1,2)    # unigram + bigram
)
# Transform teks review menjadi matriks TF-IDF
tfidf_matrix = tfidf.fit_transform(filtered_df['Clean_Review'])
```

Gambar 4.11 *Pseudocode* TF-IDF

Proses pembobotan dilakukan menggunakan kelas *TfidfVectorizer* dari *library scikit-learn* dengan parameter *max_df=0.9* untuk mengabaikan kata yang terlalu sering muncul, *min_df=2* untuk mengabaikan kata yang sangat jarang muncul, serta *ngram_range=(1,2)* untuk mempertimbangkan *unigram* dan *bigram*. Selanjutnya, fungsi *fit_transform()* digunakan untuk mengubah teks pada kolom

Clean_Review menjadi matriks TF-IDF yang merepresentasikan bobot setiap kata pada masing-masing dokumen. Hasil dari proses ini berupa matriks dengan ukuran tertentu yang menunjukkan jumlah dokumen dan jumlah fitur (kata unik) yang dihasilkan.

Sebagai contoh, dilakukan perhitungan TF-IDF menggunakan data hasil filtering berdasarkan kategori Cream & Lotion dan jenis kulit normal sebanyak 22 dokumen. Kata yang digunakan sebagai sampel adalah “cerah”, “jerawat”, “kulit”, “bersih”, dan “kering”.

Tabel 4.7 Sampel Hasil Perhitungan TF

Index	Dokumen	cerah	jerawat	kulit	bersih	kering
0	1	0	0	0	0	0
1	2	0	0	1	0	0
2	3	0	0	0	1	0
3	4	0	0	2	1	0
4	5	0	0	1	0	1
5	6	1	0	2	2	0
6	7	0	1	1	1	0
7	8	0	0	3	0	1
8	9	0	0	0	0	0
9	10	0	0	0	0	0
10	11	1	0	1	1	2
11	12	0	0	0	0	0
12	13	1	0	1	2	0
13	14	0	0	2	0	1
14	15	0	4	2	0	2
15	16	0	0	0	0	0
16	17	0	0	1	4	1
17	18	0	3	1	0	1
18	19	0	0	1	0	0
19	20	0	0	2	0	1
20	21	0	0	0	0	0
21	22	0	0	2	0	0

Berdasarkan tabel 4.7, nilai yang ditampilkan menunjukkan jumlah kemunculan setiap kata pada masing-masing dokumen. Sebagai contoh, pada dokumen ke-6 kata “cerah” muncul sebanyak 1 kali, sedangkan pada dokumen ke-15 kata “jerawat” muncul sebanyak 4 kali. Hal ini menunjukkan bahwa semakin

sering suatu kata muncul dalam dokumen, maka nilai *Term Frequency* (TF) akan semakin tinggi.

Selanjutnya dilakukan perhitungan *Inverse Document Frequency* (IDF) untuk mengetahui tingkat kepentingan suatu kata dalam seluruh dokumen. Jumlah dokumen (N) pada sampel ini adalah 22. Sebagai contoh, kata “jerawat” muncul pada 3 dokumen sehingga nilai IDF dihitung menggunakan persamaan 3.2.

$$idf_{jerawat} = \ln\left(\frac{N}{df_{jerawat}}\right) = \ln\left(\frac{22}{3}\right) = 1.992$$

Tabel 4.8 Sampel Hasil Perhitungan IDF

Index	Term	TF	IDF
0	cerah	4	1.705
1	jerawat	3	1.992
2	kulit	15	0.383
3	bersih	11	0.693
4	kering	8	1.012

Berdasarkan tabel 4.8, nilai yang ditampilkan untuk mengetahui tingkat kepentingan suatu kata dalam seluruh dokumen. Sebagai contoh, kata “jerawat” muncul pada 3 dokumen sehingga nilai IDF dihitung menggunakan persamaan 3.2 dan menghasilkan nilai sebesar 1.992. Sementara itu, kata “kulit” muncul pada 15 dokumen sehingga memiliki nilai IDF yang lebih rendah sebesar 0.383. Nilai IDF yang lebih kecil menunjukkan bahwa kata tersebut sering muncul di banyak dokumen sehingga kurang spesifik dalam membedakan karakteristik antar dokumen.

Tahap berikutnya adalah perhitungan TF-IDF yang diperoleh dari hasil perkalian antara nilai TF dan IDF. Sebagai contoh, perhitungan manual untuk kata “jerawat” pada dokumen ke-15 adalah sebagai berikut:

$$w_{jerawat,15} = tf_{jerawat,15} \times idf_{jerawat} = 4 \times 1.992 = 7.968 \approx 7.970$$

Tabel 4.9 Sampel Hasil Perhitungan TF-IDF

Index	Dokumen	cerah	jerawat	kulit	bersih	kering
0	1	0.000	0.000	0.000	0.000	0.000
1	2	0.000	0.000	0.383	0.000	0.000
2	3	0.000	0.000	0.000	0.693	0.000
3	4	0.000	0.000	0.766	0.693	0.000
4	5	0.000	0.000	0.383	0.000	1.012
5	6	1.705	0.000	0.766	1.386	0.000
6	7	0.000	1.992	0.383	0.693	0.000
7	8	0.000	0.000	1.149	0.000	1.012
8	9	0.000	0.000	0.000	0.000	0.000
9	10	0.000	0.000	0.000	0.000	0.000
10	11	1.705	0.000	0.383	0.693	2.023
11	12	0.000	0.000	0.000	0.000	0.000
12	13	1.705	0.000	0.383	1.386	0.000
13	14	0.000	0.000	0.766	0.000	1.012
14	15	0.000	7.970	0.766	0.000	2.023
15	16	0.000	0.000	0.000	0.000	0.000
16	17	0.000	0.000	0.383	2.773	1.012
17	18	0.000	5.977	0.383	0.000	1.012
18	19	0.000	0.000	0.383	0.000	0.000
19	20	0.000	0.000	0.766	0.000	1.012
20	21	0.000	0.000	0.000	0.000	0.000
21	22	0.000	0.000	0.766	0.000	0.000

Sebagai contoh pada tabel 4.9, pada dokumen ke-15 untuk kata “jerawat” dengan nilai TF sebesar 4 dan IDF sebesar 1.992 menghasilkan nilai TF-IDF sebesar 7.970. Contoh lain pada dokumen ke-6 untuk kata “cerah” diperoleh nilai TF-IDF adalah 1.705. Selain itu, pada dokumen ke-4 untuk kata “kulit” dengan nilai TF sebesar 2 dan IDF sebesar 0.383 diperoleh nilai TF-IDF sebesar 0.766. Hasil ini menunjukkan bahwa nilai TF-IDF akan semakin tinggi apabila suatu kata sering muncul dalam dokumen tertentu namun jarang muncul pada dokumen lain.

Berdasarkan hasil perhitungan tersebut, dapat disimpulkan bahwa kata “jerawat” memiliki bobot tertinggi karena muncul cukup sering pada dokumen tertentu namun relatif jarang pada dokumen lain, sehingga memiliki nilai IDF yang tinggi. Sebaliknya, kata “kulit” memiliki bobot yang rendah karena muncul di banyak dokumen sehingga kurang mampu membedakan karakteristik antar produk. Dengan demikian, metode TF-IDF mampu menonjolkan kata-kata yang bersifat informatif dan relevan dalam merepresentasikan konten ulasan produk *skincare*, sehingga dapat digunakan sebagai dasar dalam perhitungan kemiripan antarproduk pada tahap selanjutnya.

4.3 Perhitungan *Cosine similarity*

Pada tahap ini dilakukan perhitungan tingkat kemiripan antar produk menggunakan metode *cosine similarity* berdasarkan matriks TF-IDF yang telah diperoleh pada tahap sebelumnya. Perhitungan dilakukan setelah sistem melakukan penyaringan data berdasarkan kategori produk dan jenis kulit yang dipilih pengguna sehingga hanya produk yang relevan yang dibandingkan.

```
idx = filtered_df[
    filtered_df['Product'] == input_product
].index[0]

similarity_scores = cosine_similarity(
    tfidf_matrix[idx],
    tfidf_matrix
).flatten()
```

Gambar 4.12 *Pseudocode Cosine Similarity*

Pseudocode pada gambar 4.12 menunjukkan proses perhitungan *cosine similarity* yang digunakan dalam sistem rekomendasi. Tahap pertama adalah

membentuk representasi vektor TF-IDF dari teks ulasan produk yang telah melalui proses preprocessing. Selanjutnya, sistem mengambil indeks produk acuan yang dipilih secara acak dari seluruh produk yang berada pada kategori dan jenis kulit yang dipilih pengguna. Produk acuan tersebut kemudian digunakan sebagai dasar dalam proses perhitungan tingkat kemiripan dengan produk lainnya. Setelah itu, fungsi *cosine_similarity()* digunakan untuk menghitung tingkat kemiripan antara vektor TF-IDF produk acuan dengan seluruh vektor produk lainnya dalam dataset yang telah difilter. Hasil perhitungan berupa nilai *similarity* yang kemudian digunakan untuk mengurutkan produk dan menghasilkan rekomendasi dengan tingkat kemiripan tertinggi.

Menghitung nilai *cosine similarity* dilakukan untuk membandingkan kemiripan antar dokumen produk berdasarkan hasil pembobotan TF-IDF. Nilai *cosine similarity* berada pada rentang 0 hingga 1. Jika nilai *cosine similarity* semakin mendekati 1, maka semakin tinggi tingkat kemiripan antara dua produk. Sebaliknya, jika nilai mendekati 0, maka semakin rendah tingkat kemiripan antar produk.

Sebagai contoh, dilakukan perhitungan *cosine similarity* antara produk dalam kategori *facial wash* dengan jenis kulit *oily* yaitu "Oily Skin Cleanser" dan "Tea Tree & Lemon Facial Cleanser" menggunakan persamaan 3.4 sebagai berikut:

$$\begin{aligned} \text{Cosine Similarity}(A, B) &= \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \\ &= \frac{(0.216 \times 0.324) + (0.089 \times 0.268) + (0.170 \times 0.127) + \dots}{\sqrt{(0.216)^2 + (0.089)^2 + (0.170)^2 + \dots} \times \sqrt{(0.324)^2 + (0.268)^2 + (0.127)^2 + \dots}} \end{aligned}$$

$$\begin{aligned}
 &= \frac{0.315}{0.999 \times 0.999} \\
 &= \frac{0.315}{1 \times 1} \\
 &= 0.315
 \end{aligned}$$

Perhitungan manual ditampilkan untuk menjelaskan konsep, sedangkan pada sistem menggunakan *library Scikit-learn*. Pembobotan TF-IDF secara default telah menerapkan normalisasi L2, sehingga panjang vektor setiap dokumen bernilai 1. Dengan demikian, perhitungan *cosine similarity* pada sistem menjadi setara dengan *dot product* antar vektor TF-IDF.

Tabel 4.10 Sampel Hasil Perhitungan *Cosine similarity*

Product	Oily Skin Cleanser	Tea Tree & Lemon Facial Cleanser	Facial Treatment Gentle Cleanser	White Secret Facial Wash	Lightening Whip Facial Foam
Oily Skin Cleanser	1.000	0.315	0.308	0.238	0.236
Tea Tree & Lemon Facial Cleanser	0.315	1.000	0.127	0.072	0.113
Facial Treatment Gentle Cleanser	0.308	0.127	1.000	0.228	0.227
White Secret Facial Wash	0.238	0.072	0.228	1.000	0.185
Lightening Whip Facial Foam	0.236	0.113	0.227	0.185	1.000

Pada tabel 4.10 ditampilkan sampel hasil perhitungan *cosine similarity* antar produk pada kategori *Facial Wash* dengan jenis kulit *oily*. Nilai pada diagonal utama matriks bernilai 1 karena setiap produk dibandingkan dengan dirinya sendiri sehingga memiliki tingkat kemiripan sempurna. Sebagai contoh, produk "Oily Skin

Cleanser" memiliki nilai *similarity* sebesar 1 terhadap dirinya sendiri. Sementara itu, nilai *cosine similarity* antara "Oily Skin Cleanser" dan "Tea Tree & Lemon Facial Cleanser" sebesar 0.315. Nilai tersebut menunjukkan bahwa kedua produk memiliki tingkat kemiripan berdasarkan representasi TF-IDF dari teks ulasan pengguna. Selain itu, nilai *similarity* antara "Oily Skin Cleanser" dan "Facial Treatment Gentle Cleanser" sebesar 0.308, sedangkan dengan "White Secret Facial Wash" sebesar 0.238.

Nilai-nilai tersebut menunjukkan tingkat kedekatan karakteristik antar produk, di mana semakin mendekati 1 maka semakin tinggi tingkat kemiripannya. Matriks *cosine similarity* memberikan gambaran hubungan kemiripan antar produk dalam kategori dan jenis kulit yang sama. Namun, dalam implementasi sistem rekomendasi, perhitungan kemiripan difokuskan pada satu produk acuan terhadap seluruh produk lainnya. Produk kemudian diurutkan berdasarkan nilai *similarity* tertinggi untuk menghasilkan rekomendasi Top-K yang ditampilkan kepada pengguna.

4.4 Hasil Sistem Rekomendasi

Pada penelitian ini, sistem rekomendasi dibangun menggunakan pendekatan *Content-Based Filtering* (CBF), di mana rekomendasi diberikan berdasarkan kemiripan karakteristik antar produk. Karakteristik yang digunakan dalam penelitian ini diperoleh dari teks ulasan (*review*) pengguna yang telah melalui tahap preprocessing.

Implementasi metode *Content-Based Filtering* direalisasikan dalam fungsi *skincare_recommender()*. Tahap awal dalam fungsi ini adalah melakukan validasi

terhadap input pengguna, yaitu memastikan bahwa kategori *skincare* dan jenis kulit yang dimasukkan tersedia dalam dataset. Selanjutnya, sistem melakukan penyaringan data berdasarkan kategori produk dan jenis kulit yang dipilih pengguna. Dari hasil penyaringan tersebut, sistem memilih satu produk acuan secara acak dari seluruh produk yang memenuhi kriteria kategori dan jenis kulit yang dipilih. Produk acuan ini digunakan sebagai dasar dalam proses pencarian produk yang memiliki karakteristik serupa. Dataset yang telah difilter kemudian digunakan sebagai ruang pencarian rekomendasi sehingga hanya produk yang relevan dengan kebutuhan pengguna yang dibandingkan.

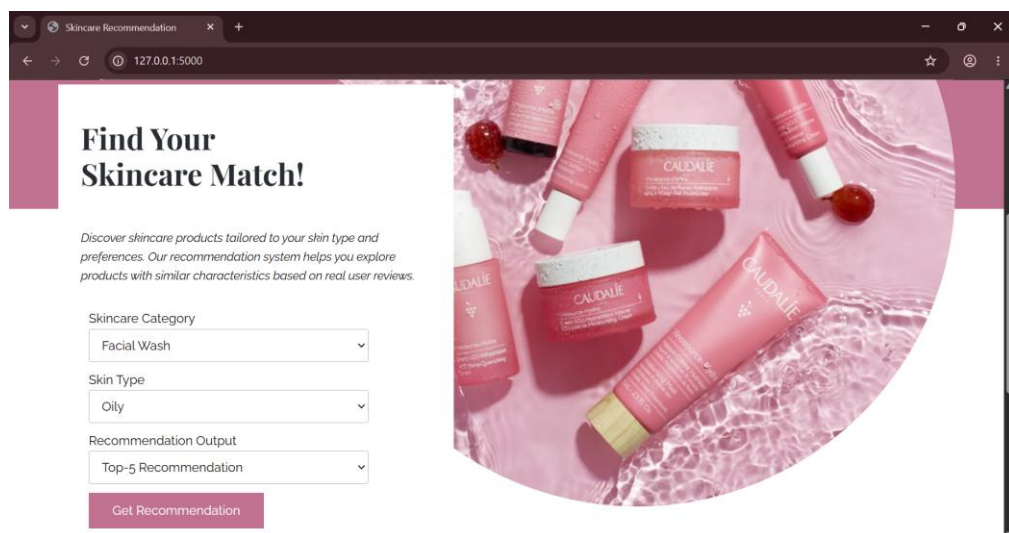
Tahap inti dari metode *Content-Based Filtering* terletak pada representasi fitur dan perhitungan kemiripan. Representasi fitur dilakukan menggunakan metode TF-IDF melalui *TfidfVectorizer*. Pada tahap ini, teks ulasan yang telah melalui proses preprocessing diubah menjadi vektor numerik yang merepresentasikan bobot kata pada setiap produk. Setelah representasi fitur terbentuk, sistem menghitung tingkat kemiripan antarproduk menggunakan metode *cosine similarity*. Perhitungan dilakukan dengan membandingkan vektor TF-IDF dari produk acuan dengan seluruh produk lain yang berada pada kategori dan jenis kulit yang sama. Hasil perhitungan berupa nilai *similarity* yang menunjukkan tingkat kedekatan antarproduk berdasarkan konten ulasan pengguna. Produk dengan nilai *similarity* tertinggi kemudian diurutkan dan ditampilkan sebagai rekomendasi kepada pengguna.

Nilai *similarity* kemudian disimpan dalam bentuk *dataframe* yang berisi pasangan produk dan nilai kemiripannya. Jika terdapat beberapa data untuk produk

yang sama, dilakukan proses *groupby* untuk menghitung rata-rata nilai *similarity*. Selanjutnya, data diurutkan secara menurun berdasarkan nilai *similarity* sehingga produk dengan kemiripan tertinggi berada di posisi teratas. Sebagai bagian dari metode CBF, produk acuan dihapus dari hasil rekomendasi karena sistem hanya menampilkan produk lain yang memiliki karakteristik serupa. Sistem kemudian mengambil sejumlah Top-K produk, yaitu Top-3 dan Top-5, sebagai hasil rekomendasi utama.

Tahap akhir adalah penggabungan informasi tambahan seperti merek, rating, dan harga untuk memperkaya hasil rekomendasi. Dengan demikian, sistem tidak hanya memberikan rekomendasi berdasarkan kemiripan konten, tetapi juga menyediakan informasi pendukung untuk membantu pengguna dalam pengambilan keputusan. Berikut adalah tampilan antarmuka (*user interface*) hasil dari sistem rekomendasi produk *skincare* yang telah diimplementasikan.

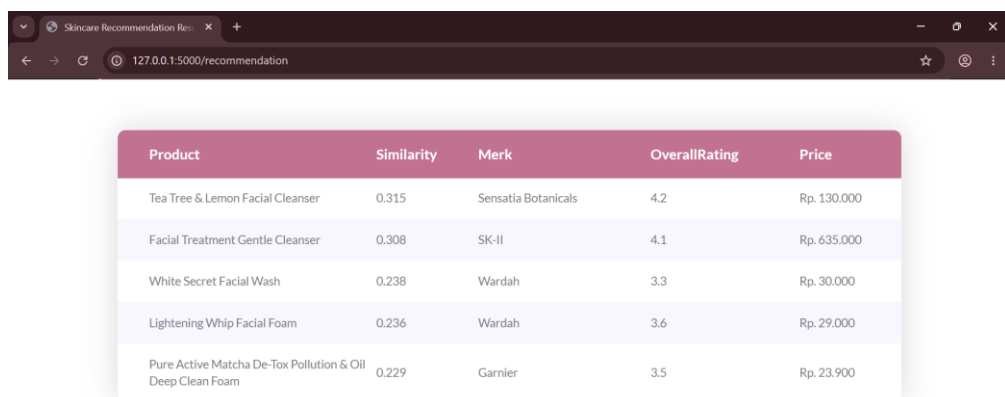
1. Halaman Input Pengguna



Gambar 4.13 Halaman Input Pengguna

Gambar 4.13 adalah halaman input yang digunakan untuk menerima masukan dari pengguna berupa kategori skincare dan jenis kulit. Data kategori ditampilkan dalam bentuk *dropdown* yang diambil dari dataset. Pengguna dapat memilih kategori *skincare* dan jenis kulit, kemudian menentukan jumlah rekomendasi yang ingin ditampilkan melalui pilihan output Top-K, yaitu Top-3 atau Top-5 rekomendasi. Setelah itu, pengguna dapat menekan tombol *Get Recommendation* untuk mendapatkan hasil rekomendasi. Pada contoh yang ditampilkan, pengguna memilih kategori *Facial Wash* dengan jenis kulit *oily* dan output rekomendasi Top-5.

2. Halaman Hasil Rekomendasi



Product	Similarity	Merk	OverallRating	Price
Tea Tree & Lemon Facial Cleanser	0.315	Sensatia Botanicals	4.2	Rp. 130.000
Facial Treatment Gentle Cleanser	0.308	SK-II	4.1	Rp. 635.000
White Secret Facial Wash	0.238	Wardah	3.3	Rp. 30.000
Lightening Whip Facial Foam	0.236	Wardah	3.6	Rp. 29.000
Pure Active Matcha De-Tox Pollution & Oil Deep Clean Foam	0.229	Garnier	3.5	Rp. 23.900

Gambar 4.14 Halaman Hasil Rekomendasi

Setelah proses perhitungan dilakukan, sistem menampilkan hasil rekomendasi dalam bentuk tabel. Tabel tersebut berisi atribut Product, Similarity, Merk, OverallRating, dan Price. Berdasarkan hasil yang ditampilkan pada gambar 4.14, produk dengan nilai *similarity* tertinggi berada pada urutan pertama, yaitu *Tea Tree & Lemon Facial Cleanser* dengan nilai *similarity*

sebesar 0.315. Nilai tersebut menunjukkan bahwa produk memiliki tingkat kemiripan paling tinggi dengan produk acuan berdasarkan analisis teks ulasan pengguna. Selanjutnya, produk *Facial Treatment Gentle Cleanser* memiliki nilai *similarity* sebesar 0.308, diikuti oleh *White Secret Facial Wash* dengan nilai *similarity* sebesar 0.238. Produk lain seperti *Lightening Whip Facial Foam* dan *Pure Active Matcha De-Tox Pollution & Oil Deep Clean Foam* masing-masing memiliki nilai *similarity* sebesar 0.236 dan 0.229.

Perbedaan nilai *similarity* tersebut mencerminkan tingkat kedekatan karakteristik antar produk terhadap produk acuan. Nilai *similarity* yang dihasilkan merupakan implementasi langsung dari metode *Content-Based Filtering*, di mana rekomendasi diberikan berdasarkan kemiripan konten ulasan (*review*). Semakin tinggi nilai *similarity*, maka semakin besar tingkat kemiripan karakteristik produk sehingga produk tersebut dianggap lebih relevan untuk direkomendasikan kepada pengguna.

4.5 Analisis Uji Coba

Pengujian dilakukan untuk mengetahui tingkat relevansi hasil rekomendasi yang dihasilkan oleh sistem menggunakan metode *Content-Based Filtering*. Evaluasi dilakukan pada seluruh kombinasi kategori *skincare* dan jenis kulit pengguna yaitu terdapat 22 kategori dan 4 jenis kulit yang berarti ada 88 skenario. Pada setiap skenario, sistem memilih satu produk acuan secara acak dari seluruh produk yang memenuhi kategori dan jenis kulit yang dipilih. Produk acuan tersebut digunakan untuk menghitung tingkat kemiripan dengan produk lain menggunakan

metode *Cosine Similarity*. Produk kemudian diurutkan berdasarkan nilai *similarity* tertinggi dan menghasilkan rekomendasi Top-3 dan Top-5.

Selanjutnya dilakukan analisis relevansi menggunakan data *ground truth* dari atribut Recommend yang bersifat *user-defined* berdasarkan pengalaman pengguna. Dalam hal ini, nilai relevansi ditentukan dari opini pengguna (*review*), bukan berdasarkan penilaian ahli (*expert judgement*). Pendekatan ini dipilih karena dataset yang digunakan berbasis ulasan pengguna, sehingga relevansi ditentukan berdasarkan persepsi kolektif pengguna terhadap produk. Nilai 1 diberikan jika produk relevan (Recommend = Yes) dan 0 jika tidak relevan (Recommend = No). Kinerja sistem diukur menggunakan metrik *precision@3* dan *precision@5* yang nilainya dihitung dengan membandingkan jumlah produk relevan terhadap total rekomendasi yang diberikan sistem. Dari seluruh 88 skenario yang diuji, ditampilkan 10 skenario terbaik sebagai sampel hasil rekomendasi dan analisis sistem.

a. Skenario 1

Pada skenario pengujian 1, digunakan kategori *skincare Face Mist* dengan jenis kulit *dry*. Sistem secara acak memilih produk acuan pada kategori dan jenis kulit tersebut yaitu *DermaSafe Booster Mist*. Selanjutnya, sistem menghasilkan rekomendasi produk berdasarkan nilai *similarity* tertinggi terhadap produk acuan tersebut.

Tabel 4.11 Hasil Skenario 1

Index	Product	Nilai <i>Similarity</i>	Relevansi
0	Aloe Vera Soothing Mist	0.228	1
1	Organic Rose Water	0.183	1
2	Aqua Beauty Protecting Mist	0.182	1
3	Innisfree Aloe vera	0.179	1
4	Baby Face & Body Spray	0.173	1
Precision@3			100%
Precision@5			100%

Berdasarkan tabel 4.11, produk *Aloe Vera Soothing Mist* memiliki nilai *similarity* tertinggi sebesar 0.228, diikuti oleh *Organic Rose Water*, *Aqua Beauty Protecting Mist*, *Innisfree Aloe Vera*, dan *Baby Face & Body Spray*. Seluruh produk yang direkomendasikan memiliki nilai relevansi 1, sehingga diperoleh *precision@3* dan *precision@5* sebesar 100%. Hasil ini menunjukkan bahwa seluruh rekomendasi yang dihasilkan sistem pada skenario ini termasuk ke dalam kategori relevan.

b. Skenario 2

Pada skenario pengujian 2, digunakan kategori *skincare Face Mist* dengan jenis kulit *oily*. Sistem secara acak memilih produk acuan pada kategori dan jenis kulit tersebut yaitu *Mist Toner*. Selanjutnya, sistem menghasilkan rekomendasi produk berdasarkan nilai *similarity* tertinggi terhadap produk acuan tersebut.

Tabel 4.12 Hasil Skenario 2

Index	Product	Nilai <i>Similarity</i>	Relevansi
0	Facial Spray With Aloe Herbs And Rosewater	0.239	1
1	Beauty Water	0.118	1
2	Innisfree Aloe vera	0.105	1
3	Pure Swiss Thermal Spring Water	0.103	1
4	Hydrating Mist	0.078	1
Precision@3			100%
Precision@5			100%

Berdasarkan tabel 4.12, produk *Facial Spray With Aloe Herbs And Rosewater* memiliki nilai *similarity* tertinggi sebesar 0.239, diikuti oleh *Beauty Water*, *Innisfree Aloe Vera*, *Pure Swiss Thermal Spring Water*, dan *Hydrating Mist*. Seluruh produk yang direkomendasikan memiliki nilai relevansi 1, sehingga diperoleh *precision@3* dan *precision@5* sebesar 100%. Hasil ini menunjukkan bahwa seluruh rekomendasi yang dihasilkan sistem pada skenario ini termasuk ke dalam kategori relevan.

c. Skenario 3

Pada skenario pengujian 3, digunakan kategori *skincare Day Cream* dengan jenis kulit *combination*. Sistem secara acak memilih produk acuan pada kategori dan jenis kulit tersebut yaitu *Pearly White UV Essence Cream*. Selanjutnya, sistem menghasilkan rekomendasi produk berdasarkan nilai *similarity* tertinggi terhadap produk acuan tersebut.

Tabel 4.13 Hasil Skenario 3

Index	Product	Nilai <i>Similarity</i>	Relevansi
0	Instabright Pink Tone Up Cream	0.240	1
1	Ms. Pimple Moisturizing Gel	0.213	1
2	Light Complete White Speed Yuzu Whitening Seru...	0.209	1
3	Madagascar Centella-Asiatica Cream	0.204	1
4	Light Complete Bright Up Tone Up Cream	0.202	1
<i>Precision@3</i>			100%
<i>Precision@5</i>			100%

Berdasarkan tabel 4.13, produk *Instabright Pink Tone Up Cream* memiliki nilai *similarity* tertinggi sebesar 0.240, diikuti oleh *Ms. Pimple Moisturizing Gel*, *Light Complete White Speed Yuzu Whitening Serum*, *Madagascar Centella-Asiatica Cream*, dan *Light Complete Bright Up Tone Up Cream*.

Seluruh produk yang direkomendasikan memiliki nilai relevansi 1, sehingga diperoleh *precision@3* dan *precision@5* sebesar 100%. Hasil ini menunjukkan bahwa seluruh rekomendasi yang dihasilkan sistem pada skenario ini termasuk ke dalam kategori relevan.

d. Skenario 4

Pada skenario pengujian 4, digunakan kategori *skincare Sun Protection* dengan jenis kulit *combination*. Sistem secara acak memilih produk acuan pada kategori dan jenis kulit tersebut yaitu *The Beet Shield/Beet The Sun*. Selanjutnya, sistem menghasilkan rekomendasi produk berdasarkan nilai *similarity* tertinggi terhadap produk acuan tersebut.

Tabel 4.14 Hasil Skenario 4

Index	Product	Nilai <i>Similarity</i>	Relevansi
0	UV Perfect Instant White SPF50	0.239	1
1	Skin Aqua UV Moisture Gel	0.228	1
2	UV Aqua Rich Watery Essence SPF 50+ PA++++	0.225	1
3	Aloe Soothing Sun Cream	0.219	1
4	Sun Protect & White Oil Control Serum SPF50 PA+++	0.218	1
<i>Precision@3</i>			100%
<i>Precision@5</i>			100%

Berdasarkan tabel 4.14, produk *UV Perfect Instant White SPF50* memiliki nilai *similarity* tertinggi sebesar 0.239, diikuti oleh *Skin Aqua UV Moisture Gel*, *UV Aqua Rich Watery Essence SPF 50+ PA++++*, *Aloe Soothing Sun Cream*, dan *Sun Protect & White Oil Control Serum SPF50 PA+++*. Seluruh produk yang direkomendasikan memiliki nilai relevansi 1, sehingga diperoleh *precision@3* dan *precision@5* sebesar 100%. Hasil ini menunjukkan bahwa

seluruh rekomendasi yang dihasilkan sistem pada skenario ini termasuk ke dalam kategori relevan.

e. Skenario 5

Pada skenario pengujian 5, digunakan kategori *skincare Sun Protection* dengan jenis kulit normal. Sistem secara acak memilih produk acuan pada kategori dan jenis kulit tersebut yaitu *Ultra Light Daily UV Defense SPF 50 PA+++*. Selanjutnya, sistem menghasilkan rekomendasi produk berdasarkan nilai *similarity* tertinggi terhadap produk acuan tersebut.

Tabel 4.15 Hasil Skenario 5

Index	Product	Nilai <i>Similarity</i>	Relevansi
0	Tone Up UV Essence SPF50+ PA++++	0.363	1
1	Sunprise Mild Airy Finish	0.230	1
2	Daily UV Protection Cream Mild	0.215	1
3	Nivea Sun Protect & White Serum	0.196	1
4	Aloe Soothing Sun Cream	0.195	1
<i>Precision@3</i>			100%
<i>Precision@5</i>			100%

Berdasarkan tabel 4.15, produk *Tone Up UV Essence SPF50+ PA++++* memiliki nilai *similarity* tertinggi sebesar 0.363, diikuti oleh *Sunprise Mild Airy Finish*, *Daily UV Protection Cream Mild*, *Nivea Sun Protect & White Serum*, dan *Aloe Soothing Sun Cream*. Seluruh produk yang direkomendasikan memiliki nilai relevansi 1, sehingga diperoleh *precision@3* dan *precision@5* sebesar 100%. Hasil ini menunjukkan bahwa seluruh rekomendasi yang dihasilkan sistem pada skenario ini termasuk ke dalam kategori relevan.

f. Skenario 6

Pada skenario pengujian 6 digunakan kategori *skincare Lotion & Emulsion* dengan jenis kulit *combination*. Sistem secara acak memilih produk acuan pada

kategori dan jenis kulit tersebut yaitu *Dramatically Different Hydrating Jelly*. Selanjutnya, sistem menghasilkan rekomendasi produk berdasarkan nilai *similarity* tertinggi terhadap produk acuan tersebut.

Tabel 4.16 Hasil Skenario 6

Index	Product	Nilai <i>Similarity</i>	Relevansi
0	Orchid Enriched Cream	0.309	1
1	Moisture Bomb Pudding Gel Creme	0.278	1
2	Moist in A Bottle Moisturizer	0.251	1
3	Dramatically Different Moisturizing Gel	0.242	1
4	Gokujoyun Ultimate Moisturizing Milk	0.218	1
Precision@3			100%
Precision@5			100%

Berdasarkan tabel 4.16, produk *Orchid Enriched Cream* memiliki nilai *similarity* tertinggi sebesar 0.309, diikuti oleh *Moisture Bomb Pudding Gel Creme*, *Moist in A Bottle Moisturizer*, *Dramatically Different Moisturizing Gel*, dan *Gokujoyun Ultimate Moisturizing Milk*. Seluruh produk yang direkomendasikan memiliki nilai relevansi 1, sehingga diperoleh *precision@3* dan *precision@5* sebesar 100%. Hasil ini menunjukkan bahwa seluruh rekomendasi yang dihasilkan sistem pada skenario ini termasuk ke dalam kategori relevan.

g. Skenario 7

Pada skenario pengujian 7 digunakan kategori *skincare Face Oil* dengan jenis kulit normal. Sistem secara acak memilih produk acuan pada kategori dan jenis kulit tersebut yaitu *100% Organic Cold-Pressed Rose Hip Seed Oil*. Selanjutnya, sistem menghasilkan rekomendasi produk berdasarkan nilai *similarity* tertinggi terhadap produk acuan tersebut.

Tabel 4.17 Hasil Skenario 7

Index	Product	Nilai <i>Similarity</i>	Relevansi
0	Luna Sleeping Night Oil	0.288	1
1	Avocado Serum Oil	0.271	1
2	Oils of Life Intensely Revitalizing Facial Oil	0.268	1
3	Rosehip Oil	0.240	1
4	Lifting Face Treatment Oil	0.228	1
Precision@3			100%
Precision@5			100%

Berdasarkan Tabel 4.17, produk *Luna Sleeping Night Oil* memiliki nilai *similarity* tertinggi sebesar 0.288, diikuti oleh *Avocado Serum Oil*, *Oils of Life Intensely Revitalizing Facial Oil*, *Rosehip Oil*, dan *Lifting Face Treatment Oil*. Seluruh produk yang direkomendasikan memiliki nilai relevansi 1, sehingga diperoleh *precision@3* dan *precision@5* sebesar 100%. Hasil ini menunjukkan bahwa seluruh rekomendasi yang dihasilkan sistem pada skenario ini termasuk ke dalam kategori relevan.

h. Skenario 8

Pada skenario pengujian 8 digunakan kategori *skincare Face Oil* dengan jenis kulit *oily*. Sistem secara acak memilih produk acuan pada kategori dan jenis kulit tersebut yaitu *Grapeseed Serum Oil*. Selanjutnya, sistem menghasilkan rekomendasi produk berdasarkan nilai *similarity* tertinggi terhadap produk acuan tersebut.

Tabel 4.18 Hasil Skenario 8

Index	Product	Nilai <i>Similarity</i>	Relevansi
0	Tamanu Oil	0.295	1
1	Pure Tamanu Oil	0.239	1
2	Jjoba Oil	0.215	1
3	Cleopatra's Rose Facial Hydrate	0.184	1
4	Grapeseed Oil	0.171	1
Precision@3			100%
Precision@5			100%

Berdasarkan tabel 4.18, produk *Tamanu Oil* memiliki nilai *similarity* tertinggi sebesar 0.295, diikuti oleh *Pure Tamanu Oil*, *Jojoba Oil*, *Cleopatra's Rose Facial Hydrate*, dan *Grapeseed Oil*. Seluruh produk yang direkomendasikan memiliki nilai relevansi 1, sehingga diperoleh *precision@3* dan *precision@5* sebesar 100%. Hasil ini menunjukkan bahwa seluruh rekomendasi yang dihasilkan sistem pada skenario ini termasuk ke dalam kategori relevan.

i. Skenario 9

Pada skenario pengujian 9 digunakan kategori *skincare Skin Soothing Treatment* dengan jenis kulit *dry*. Sistem secara acak memilih produk acuan pada kategori dan jenis kulit tersebut yaitu *Jeju Aloe Ice Refreshing Soothing Gel*. Selanjutnya, sistem menghasilkan rekomendasi produk berdasarkan nilai *similarity* tertinggi terhadap produk acuan tersebut.

Tabel 4.19 Hasil Skenario 9

Index	Product	Nilai <i>Similarity</i>	Relevansi
0	Jeju Aloe 99% Fresh Soothing Gel	0.343	1
1	Aloe Revital Soothing Gel	0.261	1
2	Missha Premium Cica Aloe Soothing Gel	0.213	1
3	Aloe Vera Gel	0.213	1
4	Aloe-HA Aloe Vera Gel	0.207	1
<i>Precision@3</i>			100%
<i>Precision@5</i>			100%

Berdasarkan tabel 4.19, produk *Jeju Aloe 99% Fresh Soothing Gel* memiliki nilai *similarity* tertinggi sebesar 0.343, diikuti oleh *Aloe Revital Soothing Gel*, *Missha Premium Cica Aloe Soothing Gel*, *Aloe Vera Gel*, dan *Aloe-HA Aloe Vera Gel*. Seluruh produk yang direkomendasikan memiliki nilai relevansi 1, sehingga diperoleh *precision@3* dan *precision@5* sebesar 100%. Hasil ini

menunjukkan bahwa seluruh rekomendasi yang dihasilkan sistem pada skenario ini termasuk ke dalam kategori relevan.

j. Skenario 10

Pada skenario pengujian 10 digunakan kategori *skincare Acne Treatment* dengan jenis kulit normal. Sistem secara acak memilih produk acuan pada kategori dan jenis kulit tersebut yaitu 2.5 (*Benzolac*). Selanjutnya, sistem menghasilkan rekomendasi produk berdasarkan nilai *similarity* tertinggi terhadap produk acuan tersebut.

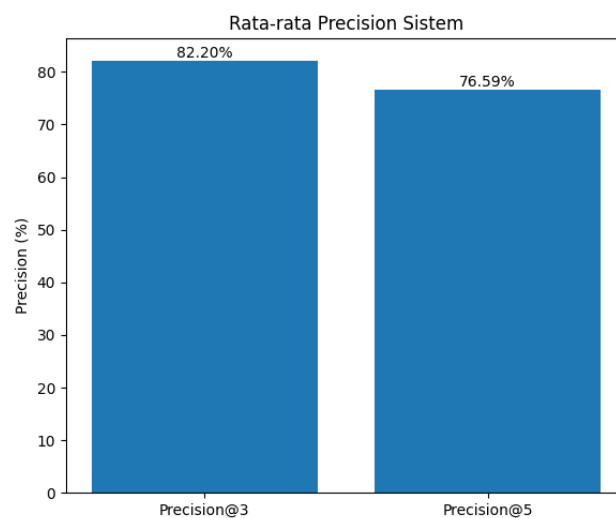
Tabel 4.20 Hasil Skenario 10

Index	Product	Nilai <i>Similarity</i>	Relevansi
0	Tea Tree Oil	0.460	1
1	Acne Pimple Medication	0.287	1
2	Lotion for Acne	0.273	1
3	AHA BHA PHA 30 Days Miracle Starter Kit	0.243	1
4	Acne Solution	0.225	1
<i>Precision@3</i>			100%
<i>Precision@5</i>			100%

Berdasarkan tabel 4.20, produk *Tea Tree Oil* memiliki nilai *similarity* tertinggi sebesar 0.460, diikuti oleh *Acne Pimple Medication*, *Lotion for Acne*, *AHA BHA PHA 30 Days Miracle Starter Kit*, dan *Acne Solution*. Seluruh produk yang direkomendasikan memiliki nilai relevansi 1, sehingga diperoleh *precision@3* dan *precision@5* sebesar 100%. Hasil ini menunjukkan bahwa seluruh rekomendasi yang dihasilkan sistem pada skenario ini termasuk ke dalam kategori relevan.

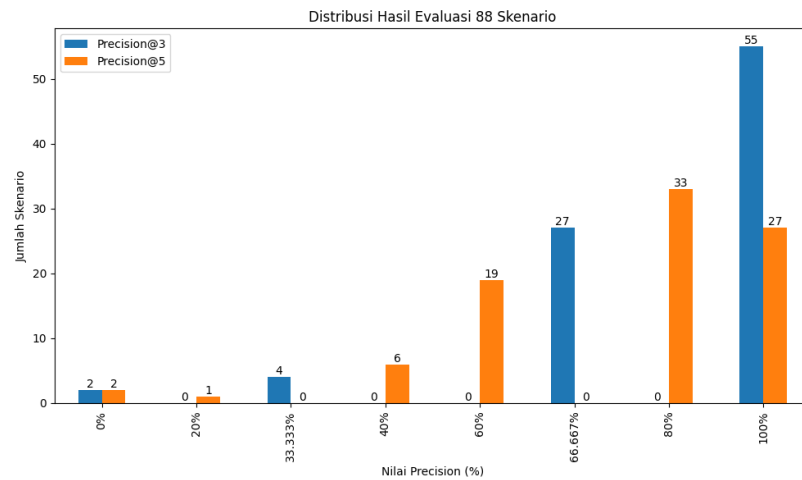
Berdasarkan 10 skenario terbaik yang telah ditampilkan, seluruh produk yang direkomendasikan memiliki nilai relevansi 1 sehingga menghasilkan nilai *precision@3* dan *precision@5* sebesar 100%. Adapun untuk hasil evaluasi terhadap

seluruh 88 skenario menunjukkan bahwa sistem memperoleh nilai rata-rata *precision@3* sebesar 82,197% dan *precision@5* sebesar 76,591%. Nilai tersebut menunjukkan bahwa sebagian besar produk yang direkomendasikan oleh sistem termasuk ke dalam kategori relevan. Visualisasi rata-rata *precision* sistem dapat dilihat pada gambar 4.15.



Gambar 4.15 Grafik Rata-Rata *Precision* Sistem

Berdasarkan gambar 4.15, nilai *precision@3* lebih tinggi dibandingkan *precision@5*. Hal ini menunjukkan bahwa produk-produk yang berada pada tiga rekomendasi teratas memiliki tingkat relevansi yang lebih tinggi dibandingkan ketika jumlah rekomendasi diperluas menjadi lima produk. Selanjutnya, distribusi hasil evaluasi pada seluruh 88 skenario ditampilkan pada gambar 4.16. Visualisasi ini menunjukkan sebaran nilai *precision@3* dan *precision@5* yang diperoleh sistem pada setiap skenario pengujian.



Gambar 4.16 Grafik Distribusi Hasil Evaluasi Sistem

Berdasarkan Gambar 4.16, sebanyak 55 skenario memperoleh nilai *precision@3* sebesar 100%, sedangkan 27 skenario memperoleh nilai *precision@3* sebesar 66,667%. Pada *precision@5*, terdapat 27 skenario yang memperoleh nilai 100%, 33 skenario memperoleh nilai 80%, dan 19 skenario memperoleh nilai 60%. Hasil tersebut menunjukkan bahwa sebagian besar skenario pengujian memiliki tingkat relevansi yang tinggi, di mana rekomendasi yang dihasilkan sistem didominasi oleh produk-produk yang termasuk kategori relevan.

4.6 Pembahasan

Berdasarkan hasil pengujian dan evaluasi yang telah dilakukan pada 88 skenario yang merupakan kombinasi kategori *skincare* dan jenis kulit pengguna, sistem rekomendasi berbasis *Content-Based Filtering* memperoleh nilai rata-rata *precision@3* sebesar 82,197% dan *precision@5* sebesar 76,591%. Nilai tersebut menunjukkan bahwa sebagian besar produk yang direkomendasikan oleh sistem termasuk dalam kategori relevan berdasarkan *ground truth* yang dibangun dari

atribut Recommend pada data ulasan pengguna. Hasil ini menunjukkan bahwa metode *Content-Based Filtering* mampu memanfaatkan informasi yang terkandung dalam teks ulasan untuk menemukan produk-produk yang memiliki karakteristik serupa.

Berdasarkan distribusi hasil evaluasi, sebagian besar skenario memperoleh nilai *precision* yang tinggi. Sebanyak 55 skenario memperoleh *precision@3* sebesar 100%, sedangkan pada *precision@5* terdapat 27 skenario yang memperoleh nilai 100% dan 33 skenario memperoleh nilai 80%. Hasil tersebut menunjukkan bahwa sistem mampu menghasilkan rekomendasi yang relevan pada berbagai kategori *skincare* dan jenis kulit yang berbeda. Meskipun demikian, masih terdapat beberapa skenario dengan nilai *precision* yang lebih rendah. Hal ini menunjukkan bahwa tingkat kemiripan berdasarkan ulasan pengguna belum selalu mampu merepresentasikan seluruh karakteristik produk secara sempurna.

Hasil pengujian juga memperlihatkan bahwa nilai *precision@3* cenderung lebih tinggi dibandingkan *precision@5*. Kondisi ini menunjukkan bahwa produk-produk yang berada pada urutan teratas hasil rekomendasi memiliki tingkat relevansi yang lebih tinggi dibandingkan produk pada urutan berikutnya. Dengan kata lain, sistem lebih efektif dalam merekomendasikan produk yang paling relevan pada posisi teratas, sedangkan ketika jumlah rekomendasi diperluas hingga lima produk, kemungkinan munculnya produk yang kurang relevan menjadi lebih besar.

Pada penelitian sebelumnya menggunakan pendekatan serupa, diperoleh nilai *cosine similarity* sebesar 0.87 pada produk Antioxidant Serum (Fitriani, 2022). Nilai tersebut merepresentasikan tingkat kemiripan antar produk, bukan tingkat

relevansi hasil rekomendasi. Penelitian tersebut belum menggunakan metrik evaluasi yang melibatkan *ground truth* untuk mengukur relevansi hasil rekomendasi. Oleh karena itu, kinerja sistem pada penelitian tersebut belum dapat dibandingkan secara langsung dengan penelitian ini yang telah menggunakan metrik evaluasi terukur berupa *precision@K*.

Sistem yang dibangun memiliki beberapa keunggulan, di antaranya mampu memberikan rekomendasi berdasarkan pengalaman pengguna lain yang tertuang dalam teks ulasan. Selain itu, penggunaan jenis kulit sebagai filter kontekstual membantu sistem menghasilkan rekomendasi yang lebih sesuai dengan kebutuhan pengguna. Namun demikian, sistem ini juga memiliki keterbatasan, yaitu bergantung pada kualitas teks ulasan yang tersedia. Jika ulasan yang diberikan tidak informatif, maka hasil rekomendasi menjadi kurang optimal. Selain itu, sistem belum mampu memahami konteks makna secara mendalam karena hanya mengandalkan kemiripan kata, sehingga pada beberapa kasus masih menghasilkan rekomendasi yang kurang sesuai.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa metode *Content-Based Filtering* mampu menghasilkan rekomendasi produk *skincare* yang relevan berdasarkan kategori produk dan jenis kulit pengguna. Nilai rata-rata *precision@3* sebesar 82,197% dan *precision@5* sebesar 76,591% menunjukkan bahwa sebagian besar rekomendasi yang dihasilkan termasuk dalam kategori relevan berdasarkan data ulasan pengguna.

4.7 Integrasi Islam

Dalam kaitannya dengan Islam, Al-Qur'an tidak secara langsung membahas mengenai sistem rekomendasi. Namun, konsep dalam sistem rekomendasi memiliki keterkaitan dengan nilai-nilai Islam, khususnya dalam memberikan petunjuk, kemudahan, serta membantu sesama dalam kebaikan. Sistem rekomendasi *skincare* yang dibangun dalam penelitian ini bertujuan untuk membantu pengguna dalam memilih produk yang sesuai dengan jenis kulitnya, sehingga memberikan manfaat serta menghindari dampak negatif dari penggunaan produk yang tidak tepat. Hal ini sejalan dengan ajaran Islam yang mendorong manusia untuk memberikan manfaat kepada sesama.

Integrasi nilai Islam dalam penelitian ini dapat dijelaskan melalui tiga aspek utama, yaitu hubungan manusia dengan Allah (*Muamalah Ma'a Allah*), hubungan manusia dengan sesama (*Muamalah Ma'a An-Nas*), dan hubungan manusia dengan lingkungan (*Muamalah Ma'al Bi'ah*). Dalam aspek hubungan manusia dengan Allah (*Muamalah Ma'a Allah*), penelitian ini berkaitan dengan upaya manusia dalam memberikan manfaat kepada orang lain melalui pemanfaatan ilmu pengetahuan dan teknologi. Hal ini sesuai dengan firman Allah SWT dalam QS. Ali Imran ayat 110:

كُنْتُمْ خَيْرَ أُمَّةٍ أُخْرِجَتْ لِلنَّاسِ تَأْمُرُونَ بِالْمَعْرُوفِ وَتَنْهَوْنَ عَنِ الْمُنْكَرِ وَتُؤْمِنُونَ بِاللَّهِ وَلَوْ ءَاَمَنَ أَهْلُ الْكِتَابِ لَكَانَ

خَيْرًا لَهُمْ مِّنْهُمْ الْمُؤْمِنُونَ وَأَكْثَرُهُمُ الْفَاسِقُونَ ﴿١١٠﴾

“Kamu (umat Islam) adalah umat terbaik yang dilahirkan untuk manusia (selama) kamu menyuruh (berbuat) yang makruf, mencegah dari yang mungkar, dan beriman kepada Allah. Seandainya Ahlulkitab beriman, tentulah itu lebih baik bagi

mereka. Di antara mereka ada yang beriman dan kebanyakan mereka adalah orang-orang fasik.” (QS. Ali Imran ayat 110)

Menurut Tafsir Al-Muyassar, umat Islam disebut sebagai umat terbaik karena menjadi orang-orang yang paling bermanfaat bagi manusia, dengan mengajak kepada kebaikan dan mencegah keburukan, serta beriman kepada Allah dengan keimanan yang diwujudkan dalam amal nyata. Sementara itu, dalam Tafsir Jalalain dijelaskan bahwa umat terbaik adalah umat yang memberikan petunjuk kepada manusia menuju jalan yang benar serta membawa kemaslahatan dalam kehidupan. Dalam konteks penelitian ini, sistem rekomendasi berperan sebagai sarana untuk memberikan petunjuk kepada pengguna dalam memilih produk *skincare* yang sesuai, sehingga termasuk dalam bentuk upaya memberikan manfaat kepada sesama manusia.

Dalam aspek hubungan manusia dengan sesama (*Muamalah Ma'a An-Nas*), sistem rekomendasi berfungsi sebagai sarana untuk membantu pengguna dalam mengambil keputusan yang lebih baik berdasarkan pengalaman pengguna lain. Hal ini sejalan dengan firman Allah SWT dalam QS. Al-Maidah ayat 2:

وَلَا تَعَاوَنُوا عَلَى الْإِثْمِ وَالْعُدْوَانِ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ شَدِيدُ الْعِقَابِ ﴿٢﴾

“Tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong-menolong dalam berbuat dosa dan permusuhan. Bertakwalah kepada Allah, sesungguhnya Allah sangat berat siksaan-Nya.” (QS. Al-Maidah ayat 2)

Menurut Tafsir Al-Muyassar, ayat ini menekankan bahwa umat Islam diperintahkan untuk saling membantu dalam kebaikan dan ketakwaan, serta dilarang bekerja sama dalam perbuatan dosa dan pelanggaran. Dalam penelitian ini,

sistem rekomendasi merupakan bentuk implementasi dari tolong-menolong dalam kebaikan, karena membantu pengguna menemukan produk yang sesuai dengan kebutuhan kulit mereka, sehingga memberikan manfaat dan menghindari dampak negatif.

Dalam aspek hubungan manusia dengan lingkungan (*Muamalah Ma'al Bi'ah*), sistem rekomendasi membantu pengguna untuk lebih bijak dalam memilih produk sehingga dapat mengurangi penggunaan produk yang tidak diperlukan atau tidak cocok. Hal ini sejalan dengan firman Allah SWT dalam QS. Al-A'raf ayat 31:

يٰۤاَيُّهَا اٰدَمُ خُذْوَ زِيْنَتَكَمْ عِنْدَ كُلِّ مَسْجِدٍ وَكُلُوْا وَاشْرَبُوْا وَلَا تُسْرِفُوْا اِنَّهٗ لَا يُحِبُّ الْمُسْرِفِيْنَ ﴿٣١﴾

“Wahai anak cucu Adam, pakailah pakaianmu yang indah pada setiap (memasuki) masjid dan makan serta minumlah, tetapi janganlah berlebihan. Sesungguhnya Dia tidak menyukai orang-orang yang berlebihan.” (QS. Al-A'raf ayat 31)

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa sistem rekomendasi produk *skincare* menggunakan metode *Content-Based Filtering* berhasil dibangun dengan memanfaatkan teks ulasan pengguna sebagai representasi konten. Sistem menghasilkan rekomendasi berdasarkan kategori *skincare* dan jenis kulit pengguna menggunakan TF-IDF dan *cosine similarity*. Berdasarkan evaluasi menggunakan *precision@3* dan *precision@5* pada uji coba seluruh kategori *skincare* dan jenis kulit pada dataset, diperoleh nilai rata-rata *precision@3* sebesar 82,197% dan *precision@5* sebesar 76,591%. Hasil tersebut menunjukkan bahwa sebagian besar produk yang direkomendasikan oleh sistem termasuk dalam kategori relevan berdasarkan data ulasan pengguna. Selain itu, hasil pengujian memperlihatkan bahwa nilai *precision@3* cenderung lebih tinggi dibandingkan *precision@5*. Hal ini mengindikasikan bahwa produk pada rekomendasi teratas memiliki tingkat relevansi yang lebih tinggi dibandingkan produk pada urutan berikutnya. Secara keseluruhan, sistem yang dibangun dapat membantu pengguna dalam memilih produk *skincare* yang sesuai dengan kategori produk dan jenis kulitnya sehingga dapat membantu pengguna dalam menentukan pilihan produk *skincare* yang lebih relevan.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya.

1. Mengembangkan sistem dengan menggabungkan metode lain seperti *Collaborative Filtering* atau pendekatan *hybrid* untuk meningkatkan kualitas rekomendasi dan mengatasi keterbatasan metode *Content-Based Filtering*.
2. Mempertimbangkan faktor tambahan seperti harga, kandungan bahan aktif, serta tingkat rating produk agar rekomendasi yang dihasilkan menjadi lebih personal dan informatif.
3. Pengolahan teks dapat ditingkatkan dengan menggunakan teknik yang lebih canggih seperti *word embedding* atau *deep learning* untuk memahami makna konteks secara lebih mendalam.
4. Sistem dapat dikembangkan dalam bentuk aplikasi yang lebih interaktif dengan antarmuka (*user interface*) pengguna yang lebih baik agar dapat digunakan secara luas oleh masyarakat.

DAFTAR PUSTAKA

- Aggarwal, C. C. (2016). *Recommender Systems*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-29659-3>
- Ardiana, V. P., Atina, V., & Nurmalitasari, N. (2024). Sistem Rekomendasi Pemilihan Basic Skincare Menggunakan Metode Content Based Filtering. *G-Tech: Jurnal Teknologi Terapan*, 8(4), 2154–2167. <https://doi.org/10.70609/gtech.v8i4.4966>
- Aulia, N. P., Sari, S. S., & Saragih, D. K. (2025). Telaah Q.S. Al-Baqarah Ayat 222 Tentang Kesehatan Dan Kebersihan. *Young Journal of Social Sciences and Humanities*, 1(2).
- Ayuningrum, E., Azhar, Y., & Marthasari, G. I. (2024). Sistem Rekomendasi Produk Skincare Korea Berbasis Web Menggunakan Metode Collaborative Filtering. *Jurnal Repositor*, 4(4). <https://doi.org/10.22219/repositor.v4i4.32284>
- Azmi, A. H., Mahardika, N. I., Prasetya, R. S. A., & Sari, A. P. (2024). *Sistem Rekomendasi Film Berbasis Konten Menggunakan Teknik Cosine Similarity dan TF-IDF*. 1(4).
- Bałchanowski, M., & Boryczka, U. (2023). A Comparative Study of Rank Aggregation Methods in Recommendation Systems. *Entropy*, 25(1), 132. <https://doi.org/10.3390/e25010132>
- B.Thorat, P., M. Goudar, R., & Barve, S. (2015). Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System. *International Journal of Computer Applications*, 110(4), 31–36. <https://doi.org/10.5120/19308-0760>
- Cheryl Ayu Melyani. (2022). Hotel Recommendation System with Content-Based Filtering Approach (Case Study: Hotel in Yogyakarta on Nusatrip Website). *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 15(1). <https://doi.org/10.36456/jstat.vol15.no1.a5375>
- Fajriansyah, M., Adikara, P. P., & Widodo, A. W. (2021). Sistem Rekomendasi Film Menggunakan Content Based Filtering. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(6).
- Fitria, A. (2024). *Sistem Rekomendasi Pekerjaan Menggunakan Pendekatan Content-Based Filtering*. Universitas Islam Negeri Maulana Malik Ibrahim Malang.

- Fitriani, F. A. (2022). *Sistem Rekomendasi Pemilihan Produk Skincare Dengan Pendekatan Content Based Filtering*.
- Geni, F. P. L. (2025). *Potensi Pasar Industri Kecantikan di Indonesia*.
- Hariatiningsih, L. R. (2020). *Penggunaan Skincare Dan Penerapan konsep Beauty 4.0 Pada Media Sosial (Studi Netnografi Wanita Pengguna Instagram)*. 11(2).
- Jannatin, A. A., & Ratnasari, C. I. (2025). Personalized Skincare Recommendation System Based on Ontology and User Preferences. *Bit-Tech*, 8(1), 1109–1119. <https://doi.org/10.32877/bt.v8i1.2857>
- Khaerunisa, R., & Husain, F. (2025). Experience Dan Persepsi Perempuan Terhadap Dampak Kesehatan Kulit Dalam Keputusan Penggunaan Produk Skincare. *Sosioglobal : Jurnal Pemikiran Dan Penelitian Sosiologi*, 9(2), 113–132. <https://doi.org/10.24198/jsg.v9i2.61544>
- Kusumaningrum, S. D. (2021). Kajian Pustaka dalam Penentuan Tipe dan Permasalahan Kulit Wajah. *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, 1(1). <https://doi.org/10.20885/snati.v1i1.3>
- Lubis, Y. I., Napitupulu, D. J., & Dharma, A. S. (2020). *Implementasi Metode Hybrid Filtering (Collaborative dan Content-based) untuk Sistem Rekomendasi Pariwisata*.
- Magfiroh, D. N. (2024). *Sistem Rekomendasi Pemilihan Skincare Personal Berdasarkan Analisis Kulit Wajah Menggunakan Metode Item-Based Collaborative Filtering*. Universitas Islam Negeri Maulana Malik Ibrahim Malang.
- Mustha'anah, H. (2020). *Skincare Review* [Dataset]. Kaggle. <https://www.kaggle.com/datasets/hafidahmusthaanah/skincare-Review> (Diakses 11 Agustus 2025)
- Nayak, M., Ligade, V. S., & Prabhu, S. S. (2023). Awareness level regarding adverse reactions caused by cosmetic products among female patients: A cross-sectional study. *Journal of Cosmetic Dermatology*, 22(9), 2512–2519. <https://doi.org/10.1111/jocd.15734>
- Noor, N. M., Muhamad, N. J., Sahabudin, N. A., & Mustafa, Z. (2018). Development of Skin Care Routine Support System. *Advanced Science Letters*, 24(10), 7830–7833. <https://doi.org/10.1166/asl.2018.13026>
- Oliveira, R., Ferreira, J., Azevedo, L. F., & Almeida, I. F. (2023). An Overview of Methods to Characterize Skin Type: Focus on Visual Rating Scales and Self-Report Instruments. *Cosmetics*, 10(1), 14. <https://doi.org/10.3390/cosmetics10010014>

- Putri, M. W., Muchayan, A., & Kamisutara, M. (2020). Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF. *JOINTECS (Journal of Information Technology and Computer Science)*, 5(3), 229. <https://doi.org/10.31328/jointecs.v5i3.1563>
- Roziqiin, N. (2024). *Sistem Rekomendasi Pemilihan Film Animasi Menggunakan Metode Content-Based Filtering*. Universitas Islam Negeri Maulana Malik Ibrahim Malang.
- Septiani, D., & Isabela, I. (2022). *Analisis Term Frequency-Inverse Document Frequency (TF-IDF) dalam Temu Kembali Informasi pada Dokumen Teks*. 01(2).
- Suhailah, E., & Hartatik, H. (2023). Pembuatan Sistem Rekomendasi Pariwisata Yogyakarta Menggunakan Triangle Multiplaying Jaccard. *Journal Automation Computer Information System*, 3(2), 115–126. <https://doi.org/10.47134/jacis.v3i2.62>
- Triansyah Putra, M. A. & Ronggo Alit. (2025). Sistem Rekomendasi Destinasi Wisata di Kabupaten Mojokerto dengan Metode Hybrid Filtering Berbasis Website. *Journal of Informatics and Computer Science (JINACS)*, 7(01), 57–64. <https://doi.org/10.26740/jinacs.v7n01.p57-64>
- Wulandari, S. A. (2019). Pengelompokan Jenis Kulit Normal, Berminyak dan Kering Menggunakan 4-Connectivity dan 8-Connectivity Region Properties Berdasarkan Ciri Rerata Bound. *Jurnal Transformatika*, 17(1), 78–87. <https://doi.org/10.26623/transformatika.v17i1.1341>
- Yanisa Putri, K. S., I Made Agus Dwi Suarjaya, & Wayan Oger Vihikan. (2024). Sistem Rekomendasi Skincare Menggunakan Metode Content Based Filtering dan Collaborative Filtering. *Decode: Jurnal Pendidikan Teknologi Informasi*, 4(3), 764–774. <https://doi.org/10.51454/decode.v4i3.601>