

Penerapan Metode Content Based Filtering pada Sistem Rekomendasi Pemilihan Produk Skincare

Sayid Muhammad Iqbal¹, Irfan Pratama²

Fakultas Teknologi Informasi, Program Studi Sistem Informasi, Universitas Mercu Buana
Yogyakarta

Email: ¹201210084@student.mercubuana-yogya.ac.id, ²irfan@mercubuana-yogya.ac.id

(Naskah masuk: 13 Juli 2024, diterima untuk diterbitkan: 20 Juli 2024)

Abstrak: Kulit memiliki peran penting sebagai organ terluar yang melindungi dan menyelimuti tubuh manusia. Kurangnya memahami faktor penyebab masalah kulit serta tidak mengetahui jenis kulit sehingga menggunakan skincare tanpa pertimbangan matang sering kali terjadi. Penerapan teknologi dalam sistem rekomendasi produk skincare dengan banyaknya pilihan metode yang dapat diaplikasikan semakin berkembang. Sehingga dapat membantu pelanggan untuk mendapatkan rekomendasi sesuai dengan kriteria masalah pada setiap jenis kulit. Content-Based Filtering adalah salah satu metode yang dapat merekomendasikan atas dasar kemiripan atribut dari produk yang telah dinilai oleh penggunaannya. Nilai dari perhitungan yang didapatkan menunjukkan bahwa pasangan dokumen dengan nilai cosine similarity tertinggi adalah antara Q dan D5 dengan nilai 0.411, diikuti oleh Q dan D2 dengan nilai 0.332, Q dan D3 dengan nilai 0.292, Q dan D4 dengan nilai 0.260, dan terakhir Q dan D1 dengan nilai 0.195. Peringkat ini menunjukkan bahwa dokumen D5 paling mirip dengan query Q, sedangkan dokumen D1 memiliki kemiripan paling rendah. Hasil penelitian dapat menunjukkan nilai cosine similarity dari setiap produk yang disarankan.

Kata Kunci – Sistem Rekomendasi; TF-IDF; Cosine Similarity; Skincare;

Application of Content Based Filtering Method on Skincare Product Selection Recommendation System

Abstract: Skin has an important role as the outermost organ that protects and envelops the human body. Lack of understanding the factors that cause skin problems and not knowing the skin type so that using skincare without careful consideration often occurs. The application of technology in skincare product recommendation systems with a large selection of methods that can be applied is growing. So that it can help customers to get recommendations according to the criteria of problems in each skin type. Content-Based Filtering is one method that can recommend on the basis of similarity of attributes of products that have been assessed by its users. The value of the calculation obtained shows that the document pair with the highest cosine similarity value is between Q and D5 with a value of 0.411, followed by Q and D2 with a value of 0.332, Q and D3 with a value of 0.292, Q and D4 with a value of 0.260, and finally Q and D1 with a value of 0.195. This ranking shows that document D5 is most similar to query Q, while document D1 has the lowest similarity. The results of the study can show the cosine similarity value of each suggested product.

Keywords – System Recommendation; TF-IDF; Cosine Similarity; Skincare;

1. PENDAHULUAN

Kemajuan teknologi komputer saat ini sangat membantu manusia, salah satunya dalam dunia Kesehatan kulit [1] Kulit sebagai organ terluar yang melindungi dan menyelimuti tubuh manusia, termasuk wajah, memiliki peran penting dalam menciptakan kesan pertama yang mempengaruhi penampilan, terutama pada remaja Wanita. Merawat kulit wajah sejak usia dini menjadi suatu keharusan karena pada masa remaja, perubahan dan masalah kulit sering kali muncul. Namun, masih banyak remaja Wanita yang kurang memahami faktor penyebab masalah kulit serta tidak mengetahui jenis kulit yang mereka miliki [2] dalam penelitiannya mengatakan keterlibatan masker

gel dapat memberikan efek perlindungan jangka panjang dan efek samping yang mungkin dapat ditimbulkan. Kulit yang mengalami penuaan dini juga menyebabkan nilai kelembaban kulit menjadi berkurang sehingga kulit menjadi kering. Berkurangnya nilai kelembaban kulit dapat diatasi dengan strategi yang efektif adalah dengan melakukan perawatan kulit yang terarah dan konsisten dan dengan memberikan produk-produk skincare yang sesuai dan rutin [3]. Di era digital ini, promosi iklan *skincare* dengan cepat menyebar melalui berbagai platform media sosial. Banyaknya perusahaan *skincare* menggunakan strategi pemasaran yang kuat dengan menggandeng artis atau selebriti terkenal untuk mempromosikan produk mereka. Namun, strategi ini sering kali menimbulkan masalah bagi konsumen seperti berbagai masalah kulit bahkan kerusakan kulit jangka panjang. Hal ini disebabkan oleh tingginya pengaruh artis sehingga banyaknya orang terpengaruh tanpa mempertimbangkan produk lebih lanjut. Salah satu cara efektif untuk mengatasi kesulitan menentukan produk yang tepat yaitu dengan melihat ulasan langsung dari pengguna lain seperti yang tersedia di platform sociolla, karena kolom review merupakan sumber informasi yang sangat bermanfaat bagi pengguna baru.

Dampak penggunaan teknologi kini dirasakan dalam berbagai hal[4]. Penerapan teknologi dalam sistem rekomendasi produk *skincare* semakin berkembang karena banyaknya pilihan metode yang dapat di aplikasikan ini bisa menciptakan kemudahan untuk pelanggan mendapatkan rekomendasi yang kiranya sesuai dengan kriteria masalah pada kulit mereka. *Content-Based Filtering* sebagai salah satu contoh karena sistem ini rekomendasinya dibangun atas dasar kemiripan salah satu contoh Sistem ini memberikan rekomendasi berdasarkan kesamaan atribut dari produk yang telah dinilai oleh pengguna. Tahapan analisis dengan metode Content-based Filtering dimulai dengan perhitungan vektor dalam teks, pembobotan menggunakan algoritma TF-IDF (term frequency-inverse document frequency), dan cosine similarity[5]

Beberapa penelitian terkait dengan sistem rekomendasi menggunakan *content-based filtering* telah banyak dilakukan. Pada tahun 2024 [6] melakukan penelitian untuk membantu customer dalam memilih minuman yang ingin dipesan. Hasil yang diperoleh dari penelitian ini adalah rating indikasi kecocokan preferensi customer dengan item dalam menu, semakin besar rating maka semakin cocok dengan preferensi customer. Lalu, penelitian dilakukan untuk untuk membangun sistem yang dapat digunakan untuk memberikan rekomendasi kepada mahasiswa dalam mencari informasi magang berbayar . Hasil penelitian ini menampilkan 10 rekomendasi berdasarkan kriteria atau kata kunci yang dicari [7]. Pada tahun 2022 [8] melakukan penelitian untuk membuat sistem rekomendasi hotel di Yogyakarta pada salah satu OTA di Indonesia. Hasil pengujian diperoleh 10 hotel dengan kemiripan atau kesamaan. Penelitian sebelumnya yang dilakukan oleh [9] untuk membuat sistem rekomendasi dengan penerapan metode *Content-Based Filtering* yang dapat mencatat data lahan pertanian yang berpotensi menghasilkan komoditas tanaman pangan. Berdasarkan pengujian terhadap 10 profil pedagang dengan 15 rekomendasi Kelompok tani teratas, didapatkan hasil presisi rata-rata sebesar 78,40%. Penelitian yang dilakukan oleh [10] untuk mengembangkan aplikasi yang dapat merekomendasikan kosmetik emina kepada pelanggan berdasarkan produk yang telah mereka cari sebelumnya. Hasil penelitian ini, produk dengan similaritas tertinggi didapat dengan nilai sebesar 0,7195. Sedangkan penelitian ini dilakukan untuk membuat sistem rekomendasi yang bisa memberikan saran atau rekomendasi laptop berdasarkan minat dan kebutuhan dalam mencari referensi. Hasil penelitian ini menghasilkan Teknik penggabungan 2 metode yang dapat menghasilkan sistem rekomendasi dan untuk penggunaan data, pada hasil eksekusi data metode *content-based filtering* memiliki waktu eksekusi lebih cepat dari metode *collaborative filtering* [11]. Selanjutnya [12] melakukan penelitian untuk membangun sistem rekomendasi produk menggunakan metode *content-based filtering* dan TF-IDF dari model *information Retrieval*. Hasil penelitian ini memberikan Hasil perhitungan TF-IDF dengan bobot nilai 13,845 menampilkan rekomendasi produk terbaik pertama dan memiliki akurasi sebesar 96,5% dalam memberikan rekomendasi pena. Suhendra [13] melakukan penelitian untuk perancangan sistem rekomendasi menggunakan metode *content-based filtering* dengan menggunakan TF-IDF. Hasil penelitian ini telah merancang sistem rekomendasi untuk judul anime yang memungkinkan pengguna mendapatkan judul anime dengan synopsis yang mirip dengan judul anime yang telah

dibaca. Pratiwi et al [14] melakukan penelitian untuk membuat sistem rekomendasi menggunakan metode *content-based filtering* dan algoritma *Random Forest Regression*. Hasil rekomendasi didapat berdasarkan nilai kesamaan tertinggi. Hasil penelitian menggunakan model *Random Forest Regression* memperoleh akurasi 83,750% dengan nilai MAPE 16,249% pada pengujian K=10 dalam *tree random forest*. Ayu et al [15] melakukan penelitian untuk mengembangkan aplikasi yang merekomendasikan produk perawatan kulit kepada konsumen berdasarkan preferensi produk sebelumnya, menggunakan metode *content-based filtering* dan algoritma *apriori*. Hasil penelitian dengan menggunakan 40 data skincare didapatkan hasil rating tertinggi 0,477 dan menggunakan *minimum support* 40% dan *minimum confidence* 40%, aturan posisi yang dihasilkan memiliki nilai *confidence* sebesar 88,89%. Dari temuan tersebut dapat disimpulkan bahwa kombinasi metode *content-based filtering* dan algoritma *apriori* memberikan rekomendasi yang efektif.

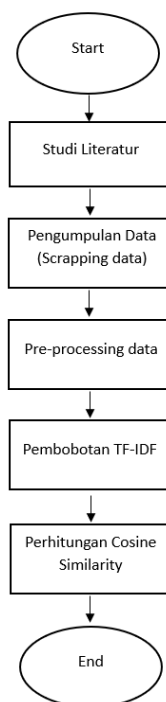
Penelitian ini bertujuan untuk menciptakan sebuah sistem yang bisa dikatakan sesuai dalam perhitungan, perbandingan kecocokan dalam setiap data kemudian bisa menjadi sumber terpercaya sebagai pengambilan Keputusan rekomendasi skincare. Kurangnya edukasi membuat seseorang memakai produk berbahaya tanpa tahu efek samping kedepannya. [16] hal itu menjadi salah satu alasan penelitian ini, agar orang yang belum paham skincare bisa mendapatkan produk tanpa perlu khawatir membuang waktu lama mencari berbagai informasi terkait prduk yang diinginkan.

2. METODE PENELITIAN

2.1. Tahapan Penelitian

Content-Based Filtering adalah sistem rekomendasi yang mempergunakan profil preferensi pengguna dan deskripsi item untuk memberikan rekomendasi berdasarkan karakteristik konten dari item tersebut tepat digunakan dalam situasi di mana terdapat lebih banyak item daripada pengguna dalam sistem. Dalam konteks ini, sistem dapat menentukan ketertarikan dan minat pengguna berdasarkan profil yang dimilikinya. [17]

Dalam penelitian ini, pendekatan *Content-Based Filtering* yang digunakan secara umum memiliki beberapa tahapan agar dapat mencapai tujuan secara sistematis. Dalam implementasi metode ini menggunakan pembobotan *TF-IDF* untuk melakukan pembobotan nilai atribut dan perhitungan *Cosine Similarity*. Alur tahapan penelitian terdapat pada gambar 1.



Gambar 1. Alur tahapan penelitian

Berikut penjelasan dari alur tahapan penelitian diatas:

1. Studi Literatur

Pada tahap ini dilakukan studi literatur dari penelitian terdahulu yang relevan dengan penerapan metode *content-based filtering* dalam implementasi sistem rekomendasi. Serta studi penggunaan *scrapping*, *cosine similarity*, dan pembobotan *TF-IDF* melalui jurnal ilmiah.

2. Pengumpulan Data (*Scrapping* data)

Pengumpulan data yang dilakukan dengan cara memeriksa situs website yang dituju yaitu sociolla dengan menggunakan bantuan software parsehub, yang di ambil seperti data *description* dan *review*

3. *Pre-processing* Data

Dalam tahap ini data yang telah di dapatkan menjalani proses cleaning dengan menghilangkan data-data yang tidak diperlukan seperti tahapan sebagai berikut:

1) *Case folding*

Case folding merupakan proses mengubah semua huruf kapital dalam dokumen menjadi huruf kecil. Karakter tanda baca dan angka akan dihilangkan dan dianggap delimiter.

2) *Tokenizing*

Tokenizing atau parsing adalah proses membagi string input berdasarkan setiap kata yang membentuknya. Pada tahap tokenisasi, kalimat dipisahkan menjadi kata-kata, dan dilakukan penghapusan angka, tanda baca seperti simbol, serta spasi kosong.

3) *Filtering*

Filtering adalah proses untuk mengambil kata-kata kunci dari hasil tokenisasi. Algoritma yang digunakan mencakup penggunaan stoplist (untuk menghapus kata-kata yang tidak penting) atau wordlist (untuk menyimpan kata-kata yang penting)

4) *Steaminging*

Steaminging adalah proses untuk mencari akar kata atau bentuk dasar dari setiap kata setelah proses normalisasi. Pada tahap steaminging, kata-kata dibawa ke bentuk representasi yang sama dan awalan serta akhiran kata dihilangkan.

5) Pembobotan *TF-IDF*

Metode ini melibatkan analisis pola untuk menentukan seberapa sering kata-kata tertentu muncul dalam suatu dokumen. Kata-kata dengan nilai TF (Term Frequency) yang tinggi menunjukkan pentingnya dalam dokumen tersebut, sedangkan DF (Document Frequency) mencerminkan berapa kali kata tersebut muncul di seluruh kumpulan dokumen. Metode ini menghitung frekuensi kemunculan kata dalam berbagai dokumen.[18] Metode ini mengkombinasikan dua konsep untuk menghitung bobot, yaitu frekuensi kemunculan sebuah kata dalam dokumen tertentu dan inversi frekuensi dokumen yang mengandung kata tersebut. Semakin tinggi frekuensi kata dalam dokumen tertentu, semakin penting kata tersebut dalam dokumen tersebut. Perhitungan bobot setiap *term* di dokumen menggunakan rumus [19]

1

$$tfidf_{t,d} = tf_{t,d} \times idf_t \quad (1)$$

Dengan idf, diperoleh dari

$$idf_t = \log\left(\frac{N}{df_t}\right) \quad (2)$$

4. Perhitungan cosine similarity

Cosine similarity adalah algoritma yang digunakan dalam text mining untuk mengelompokkan dokumen atau teks berdasarkan cara membandingkan kedekatan antara dokumen A dan B. Skor similarity diperoleh dengan mengukur skor similarity antara dua vektor, yaitu vektor kueri dan vektor dokumen. Semakin besar nilai relevansi maka semakin mirip kueri dan dokumen. Perhitungan ini digunakan untuk mencari nilai kemiripan antar review prosuk satu dan yang lainnya untuk menghasilkan rekomendasi yang sesuai. Perhitungan *cosine similarity* menggunakan rumus [20]

$$\begin{aligned} \text{Cosine similarity} &= S_c(A, B) := \text{Cos}(\theta) \\ &= \frac{A \cdot B}{\|A\| \|B\|} \\ &= \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \end{aligned} \quad (3)$$

Keterangan:

A = Vektor A yang akan di bandingkan kemiripannya

B = Vektor B yang akan di bandingkan kemiripan

3. HASIL DAN PEMBAHASAN

3.1. Grouping dan penggabungan review

Proses dimulai dengan pengelompokan ulasan berdasarkan nama produk (`product_name`) dalam DataFrame `df`, di mana setiap kelompok ulasan digabungkan menjadi satu teks menggunakan fungsi `groupby` dan `apply`. Hasilnya disimpan dalam `combined_df`, yang kemudian digabungkan kembali dengan DataFrame asli (`df`) berdasarkan `product_name` untuk mempertahankan informasi tambahan seperti merek produk, kategori, subkategori, deskripsi, rating bintang, dan rekomendasi. Setelah itu, baris duplikat berdasarkan `product_name` dihapus untuk menghasilkan DataFrame akhir `df`, yang merupakan representasi terbaru dan terstruktur dari data produk dengan ulasan yang terintegrasi, siap untuk analisis lebih lanjut. Contoh gambar sebelum dan sesudah hasil review di gabungkan menjadi satu berdasarkan 1 produk yang sama dapat dilihat pada Gambar 1 dan Gambar 2.

	A	B	C	D	E	F
1	category	subcategory	product_name	price	description	user_name
2	Treatment	Serum & Essence	All In One Miracle Serum	Rp. 89.900	All in one mircale serum ini mengandung niacinamide, centella asiatica, glutathione, micro gol wulanraahma	
3	Treatment	Serum & Essence	All In One Miracle Serum	Rp. 89.900	All in one mircale serum ini mengandung niacinamide, centella asiatica, glutathione, micro gol sofiahdynt	
4	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g wulanraahma	
5	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g sofiahdynt	
6	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g popy_mercylin	
7	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g putrimayask	
8	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g Intanndya	
9	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g reginasandra	
10	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g arshya	
11	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g ayufatimah27	
12	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g liagracia	
13	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g lrmachan	
14	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g ayunurintanofficial	
15	Moisturizer	Cream	Glowing Night Cream	Rp. 89.900	Night cream ini mengandung cucumber extract, niacinamide, alpha arbutin, shea butter, dan g Intanwati_	

Gambar 2. Sebelum grouping dan penggabungan review

Gambar 3. Sesudah grouping dan penggabungan review

3.2. Data preprocessing

Dalam data processing, penulis menambahkan langkah-langkah berikut untuk fungsi preprocess_text: pertama, menghapus emoji dengan remove_emoji; kedua, menghilangkan angka menggunakan re.sub(r'\d+', "", text); ketiga, merubah teks menjadi huruf kecil menggunakan text.lower(); keempat, memecah teks menjadi token dengan nltk.word_tokenize(text); kelima, melakukan filtering kata-kata dengan hanya mempertahankan kata-kata yang terdiri dari alfabet dan tidak termasuk dalam daftar stop_words; keenam, melakukan stemming pada kata-kata menggunakan stemmer.stem(word) untuk mengembalikan kata-kata ke bentuk dasarnya; dan terakhir, menggabungkan kata-kata yang telah di-stem menjadi satu string dengan " ".join(stemmed_words). Langkah-langkah ini membentuk proses yang komprehensif untuk membersihkan teks dari elemen yang tidak relevan dan mempersiapkannya untuk analisis teks lebih lanjut.

Table 1. data preprocessing

No	Nama produk	sebelum	sesudah
D1	01 Radiance Skin Exfoliate Serum/Peeling Serum	Bangga sama produk buatan Sidoarjo satu ini. Untuk harganya yg di bawah 30.000, ini oke banget. Kandungannya ada AHA 3%, BHA 1%, dan PHA 2%, plus Witch Hazel. Tiap abis pake ini kalo malem, paginya auto mulus, dan glowingnya cukup nggak lebay. Ngga ada fragrancenya jd bakal kerasa bau chemical tp ngga ganggu kok. Nyoba yg ini krn udah cocok banget sama serum pertama lysca yg whitening. Ini ngebantu ngempesin jerawat yg meradang, tp kalo klaim exfo'nya aku belum ngerasain krn baru nyoba belum nyampe sebulan sih. So far cocok n pengen coba jg varian yg lain :) Akhirnya menemukan exfoliating serum dengan kandungan yang ngga main-main yaitu AHA 3%, BHA 1%, PHA 2% dan juga Witch Hazel Extract. Meskipun dengan 3 kandungan exfoliasi dari AHA, BHA dan PHA, ada juga kandungan witch hazel extract yang bisa sangat bagus untuk soothing agent dan anti inflammation. Dengan kandungan yang hebat seperti ini, harganya pun sangat ekonomis. Untuk exfo serum seperti ini, menurutku paling pas dipakainya pas malam hari supaya	bangga produk buat sidoarjo harga yg oke banget kandung aha bha pha plus witch hazel abis pake kalo malem pagi auto mulus glowingnya nggak lebay ngga fragrancenya jd rasa bau chemical tp ngga ganggu nyoba yg krn udah cocok banget serum lysca yg whitening ngebantu ngempesin jerawat yg radang tp kalo klaim exfo'nya ngerasain krn nyoba nyampe bulan sih so far cocok n ken coba jg varian yg temu exfoliating serum kandung ngga main aha bha pha witch hazel extract kandung exfoliasi aha bha pha kandung witch hazel extract bagus soothing agent anti inflammation kandung hebat harga ekonomis exfo serum turut pas pakai pas malam ngga iritasi pas pakai

	<p>ngga iritasi. Pas aku pakai ini, di awal ada sedikit banget tingling sensation yang aku rasakan. Tapi itu sangat wajar. Besoknya aku notice kalo wajah jadi lebih bouncy dan lebih fresh. Cocok banget buat para teman-teman newbie yang mau nyobain exfoliating serum, you guys should try this! Exfoliating serum yang menurutku sangat gentle dan mild, tekstur-nya kaya air alias watery. Serum ini bagus banget sih buat ratakan tekstur kulit, bye bruntus dan hopefully bisa mudarin bekas jerawat aku juga. 😊 Serum ini leave on ya alias ga perlu dibilas. Ga ada kerasa cekat-cekit atau tingling walaupun dia mengandung AHA BHA & PHA, cocok bgt buat pemula yang mau coba exfo serum. 👍 ini aku beli karena iseng gais, belinya sama serum yang bounce skin ituu. bagus sih, bagus banget bahkan. kayaknya kalo buat pemula jangan dulu deh. soalnya aku pake ada rasa tingling nya. tapi kalo dah terbiasa enakeun. kulit jadi lembut banget setelah exfo. semoga kedepannya makin diperbanyak varian serumnya dan skincarenya 😊😊😊</p>	<p>banget tingling sensation rasa wajar besok notice kalo wajah bouncy fresh cocok banget teman newbie nyobain exfoliating serum you guys should try this exfoliating serum turut gentle mild tekstur kaya air alias watery serum bagus banget sih ratakan tekstur kulit bye bruntus hopefully mudarin bekas jerawat serum leave on ya alias ga bilas ga rasa cekat-cekit tingling kandung aha bha pha cocok bgt mula coba exfo serum beli iseng gais beli serum bounce skin ituu bagus sih bagus banget kayak kalo mula deh pake tingling nya kalo dah biasa enakeun kulit lembut banget exfo moga depan banyak varian serum skincarenya</p>
--	---	---

Teks yang sudah diproses menghilangkan kata-kata yang tidak relevan seperti "dan", "adalah", dan "sesudah", serta menggabungkan kalimat-kalimat yang memiliki arti yang sama atau serupa. Hasilnya menjadi lebih padat dan lebih fokus pada inti informasi. Misalnya, deskripsi tentang manfaat produk dan kandungan aktifnya telah disederhanakan untuk menyampaikan informasi secara lebih langsung tanpa kehilangan makna pentingnya.

3.3. TF-IDF

Pada data “Deskripsi produk dan review produk” yang telah di gabungkan, untuk menemukan hasil dari kata kunci, di lakukan pembobotan yang kemudia pada setiap dokumen mengandung nilai term dari jumlah seringnya kata muncul namun sementara diberi nilai 1.

Tabel 2. Pembobotan term 5 dokumen teratas

Term	TF					
	Bersih	Cocok	Facial	jerawat	kering	kulit
Q	1	1	1	1	1	1
D1	0	4	0	2	0	7
D2	0	3	0	0	3	20
D3	0	1	0	5	1	8
D4	2	2	0	16	0	5
D5	1	6	0	4	1	12

DF (Document Frequency) adalah jumlah dokumen yang mengandung kata kunci. Dalam menghitung nilai IDF menggunakan rumus $IDF = \log(n/DF)$ dimana n adalah jumlah total dokumen dan DF adalah jumlah dokumen yang mengandung kata tersebut. IDF memberikan bobot lebih tinggi pada kata-kata yang jarang muncul, sehingga kata-kata unik mendapatkan nilai IDF yang lebih tinggi.

Tabel 3. Perhitungan IDF

Term	DF	D/DF	IDF
Bersih	3	2	0.301
Cocok	6	1	0.000
Facial	1	6	0.778
Jerawat	4	1,5	0.176
Kering	3	2	0.301
Kulit	6	1	0.000

Untuk mendapatkan hasil perhitungan wdt, yaitu menerapkan rumus wdt adalah hasil dari perkalian antara nilai tf dan idf kemudian dapat menghitung bobot wdt yang mempertimbangkan kedua faktor untuk setiap kata dalam dokumen.

Tabel 4. Perhitungan Wdt

Term	Wdt=TF.IDF					
	Bersih	Cocok	Facial	Jerawat	Kering	Kulit
Q	0.301	0.000	0.778	0.176	0.301	0.000
D1	0.000	0.000	0.000	0.352	0.000	0.000
D2	0.000	0.000	0.000	0.000	0.903	0.000
D3	0.000	0.000	0.000	0.880	0.301	0.000
D4	0.602	0.000	0.000	2.816	0.000	0.000
D5	0.301	0.000	0.000	0.704	0.301	0.000

Saklar (Switching Score) mengukur seberapa banyak sebuah term muncul dalam satu dokumen dibandingkan dengan dokumen lainnya. Langkah 1: Hitung Total Frequency (TF) untuk setiap term, Langkah 2: Hitung Average Frequency (AF) untuk setiap term: Langkah 3: Hitung Switching Score (SS).

Tabel 5. Perhitungan saklar

	Bersih	Cocok	Facial	Jerawat	Kering	Kulit
Q	0.5	0.647	5	0.786	0	0.887
D1	1	0.412	1	0.571	1	0.207
D2	1	0.059	1	1	2	1.265
D3	1	0.647	1	0.071	0	0.094
D4	2	0.294	1	2.429	1	0.443
D5	0.5	1.118	1	0.143	0	0.359

Untuk menghitung panjang vektor (vector length) dari setiap dokumen, kita menggunakan rumus Euclidean norm, adalah akar kuadrat dari jumlah kuadrat nilai $WdtW_{dt}$ setiap term dalam dokumen tersebut.

Dokumen Q

$$\sqrt{(0.301^2) + (0.000^2) + (0.778^2) + (0.176^2) + (0.301^2) + (0.000^2)}$$

$$= \sqrt{0.090601 + 0 + 0.605284 + 0.030976 + 0.090601 + 0}$$

$$= \sqrt{0.817462}$$

$$= \sqrt{0.904}$$

Dokumen D1

$$\sqrt{(0.000^2) + (0.000^2) + (0.000^2) + (0.352^2) + (0.000^2) + (0.000^2)}$$

$$= \sqrt{0 + 0 + 0 + 0.123904 + 0 + 0}$$

$$= \sqrt{0.123904}$$

$$= \sqrt{0.352}$$

Tabel 6. Perhitungan Vektor

Term	Panjang Vektor
Q	0.904
D1	0.352
D2	0.903
D3	0.930
D4	2.879
D5	0.822

3.4. Cosine similarity

Cosine Similarity digunakan setelah penghitungan TF-IDF dalam analisis teks untuk mengevaluasi seberapa mirip dua dokumen berdasarkan representasi vektor mereka dalam ruang n-dimensi. TF-IDF digunakan untuk menormalisasi pentingnya kata-kata dalam dokumen relatif terhadap seluruh korpus, yang menghasilkan vektor yang merepresentasikan bobot kata-kata. Cosine Similarity kemudian mengukur sudut kosinus antara vektor-vektor ini, mengabaikan panjang vektor dan fokus pada arah dan hubungan relatif. Hal ini memungkinkan untuk membandingkan dokumen berdasarkan penggunaan kata-kata kunci yang relevan, memfasilitasi clustering dokumen, pencarian dokumen serupa, serta memberikan rekomendasi berbasis konten dengan efisien dalam pengolahan data teks besar.

Menghitung cosine similarity antara dokumen Q dengan dokumen lainnya:

Cosine Similarity antara Q dan D1:

$$\frac{(0.301 \times 0) + (0 \times 0) + (0.778 \times 0) + (0.176 \times 0.352) + (0.301 \times 0) + (0 \times 0)}{0.904 \times 0.352}$$

$$= \frac{0+0+0+0.061952+0+0}{0.318208}$$

$$= \frac{0.061952}{0.318208}$$

$$\approx 0.195$$

Cosine Similarity antara Q dan D2:

$$\frac{(0.301 \times 0) + (0 \times 0) + (0.778 \times 0) + (0.176 \times 0) + (0.301 \times 0.903) + (0 \times 0)}{0.904 \times 0.903}$$

$$= \frac{0+0+0+0+0.271203+0}{0.816912}$$

$$= \frac{0.271203}{0.816912}$$

$$\approx 0.332$$

Tabel 7. Hasil Cosine similarity

Pasangan Dokumen	Cosine Similarity
Q dan D1	0.195
Q dan D2	0.332
Q dan D3	0.292
Q dan D4	0.260
Q dan D5	0.411

Berdasarkan perhitungan cosine similarity sebelumnya, kita dapat merangking dokumen dengan kesamaan dokumen query (Q). Dokumen dengan nilai cosine similarity tertinggi akan ditempatkan di peringkat pertama karena memiliki kesamaan terbesar dengan dokumen query.

Tabel 8. Hasil peringkat

Pasangan Dokumen	Cosine Similarity	Peringkat
Q dan D5	0.411	1
Q dan D2	0.332	2
Q dan D3	0.292	3
Q dan D4	0.260	4
Q dan D1	0.195	5

3.5. Prototype

1. Tampilan kriteria produk

Pada tahapan awal user diminta untuk menginputkan data kriteria produk yang ingin di cari, dan memasukan jumlah data rekomendasi yang ingin di tampilkan



Gambar 4. Tampilan kriteria produk

2. Tampilan hasil kriteria produk

Pada tampilan selanjutnya akan mengeluarkan sebuah tabel yang menampilkan hasil urutan terbaik dari nama produk, merk produk, kategori produk dan serta sub kategori produk yang di dapatkan dari proses pencocokan dengan menggunakan pendekatan content-based filtering dalam konteks setiap komen review serta deskripsi produk.

Cat Rekomendasi			
Nama Produk	Merek Produk	Kategori	Sub-Kategori
sunshield Acne gel	Nourish Beauty Care	Treatment	Acne Treatment
Hydrocolloid Acne Patch	Tecno Beauty	Treatment	Acne Treatment
Clear Tinted Care Plus Spot Patch	OSD	Treatment	Acne Treatment
Acne Patch	Akasmol	Treatment	Acne Treatment
Acne Clear Patch	Kiss Beauty	Treatment	Acne Treatment
Peptide Anti-Aging Serum	E-velite	Treatment	Serum & Essence
Bioluminesc Toner	Body Patch Skincare	Cleanser	Toner
Acne Lotion	Akademi Aestheti	Treatment	Acne Treatment
Blue Hydrocolloid Pimple Patch	Mel For Makeup	Treatment	Acne Treatment
AcneWala Microsanda Technology Pimple Patch	REDA	Treatment	Acne Treatment

Gambar 5. Tampilan hasil kriteria produk

3. Tampilan hasil rekomendasi produk

Pada bagian selanjutnya akan menampilkan diagram batang yang mempresentasikan hasil seberapa besar perbedaan antara nilai cosine yang di dapat kan dari setiap produk atas kecocokan yang di inputkan, semakin tinggi nilai cosine semakin tinggi rangking untuk rekomendasi produk itu



Gambar 6. Tampilan hasil rekomendasi produk

4. KESIMPULAN

Berdasarkan analisis dari lima sampel data pengujian, penelitian ini menunjukkan bahwa pembobotan TF-IDF dan algoritma cosine similarity berhasil mendeteksi kemiripan antar dokumen, dengan tahapan data preprocessing yang sangat berpengaruh terhadap nilai kemiripan hasil. Nilai dari perhitungan yang didapatkan menunjukkan bahwa pasangan dokumen dengan nilai cosine similarity tertinggi adalah antara Q dan D5 dengan nilai 0.411, diikuti oleh Q dan D2 dengan nilai 0.332, Q dan D3 dengan nilai 0.292, Q dan D4 dengan nilai 0.260, dan terakhir Q dan D1 dengan nilai 0.195. Peringkat ini menunjukkan bahwa dokumen D5 paling mirip dengan query Q, sedangkan dokumen D1 memiliki kemiripan paling rendah. Penerapan metode content-based filtering dalam implementasi sistem rekomendasi pemilihan produk skincare terbukti dapat membantu pengguna memudahkan serta mengurangi risiko kesalahan dalam memilih produk yang sesuai dengan kebutuhan kulit mereka.

UCAPAN TERIMA KASIH

Terima kasih banyak atas bimbingan dan dorongan yang Bapak Irfan Pratama berikan selama proses penelitian ini. Setiap arahan dan nasihat dari Bapak telah menjadi pilar yang kokoh dalam memandu langkah-langkah saya. Dengan kesabaran dan pengetahuan yang mendalam, Bapak telah membantu saya mengatasi setiap tantangan. Semangat Bapak dalam mengajarkan nilai-nilai keilmuan dan keteladanan telah menginspirasi saya untuk terus berusaha menjadi lebih baik. Terima kasih atas dedikasi Bapak dalam membimbing saya menuju kelancaran dan kesuksesan penelitian ini.

DAFTAR PUSTAKA

- [1] F. R. Dewi, N. L. Azizah, and H. Hindarto, "Implementasi Fuzzy Tsukamoto Dan Algoritma Genetika Pada Pemilihan Skincare," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 5, no. 2, pp. 95–102, Apr. 2023, doi: 10.47233/jteksis.v5i2.785.
- [2] Y. G. Benedicta, A. Pranayama, R. P. Sutanto, and D. Komunikasi, "PERANCANGAN MEDIA EDUKASI UNTUK MEMBANTU REMAJA WANITA DALAM MENGENAL JENIS KULIT SEBELUM MENGGUNAKAN SKINCARE."
- [3] K. Widyani Astuti, N. P. A D Wijayanti, A. A. D Lestari, I. G. A P Y Artha, I. A. G Pradnyani, and dan I. G A D Ratnayanti, "UJI PENDAHULUAN NILAI KELEMBABAN KULIT MANUSIA PADA PEMAKAIAN SEDIAAN MASKER GEL PEEL OFF KULIT BUAH MANGGIS".
- [4] K. H. C. Utama, A. P. Wardhanie, and R. Santoso, "Aplikasi Pemesanan Menu Makanan Menggunakan Metode FCFS Pada Grande Garden Cafe," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 13, no. 1, Jan. 2024, doi: 10.30591/smartcomp.v13i1.5296.
- [5] "SISTEM REKOMENDASI PEMILIHAN PRODUK SKINCARE DENGAN PENDEKATAN CONTENT-BASED FILTERING."
- [6] R. Prihandi, "SISTEM REKOMENDASI MENU MINUMAN DENGAN METODE CONTENT-BASED FILTERING BERBASIS ANDROID PADA MUBTADA KOPI," 2024.
- [7] R. U. Ichlas Amal Sukma Putra and J. Sasongko Wibowo, "Sistem Rekomendasi Magang Berbayar Menggunakan Metode Content-based Filtering".
- [8] "SISTEM REKOMENDASI HOTEL DENGAN PENDEKATAN CONTENT-BASED FILTERING."
- [9] P. Nastiti, "Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan," *Teknika*, vol. 8, no. 1, pp. 1–10, Jun. 2019, doi: 10.34148/teknika.v8i1.139.
- [10] "5. SISTEM REKOMENDASI PRODUCT EMINA COSMETICS DENGAN MENGGUNAKAN METODE CONTENT - BASED FILT".
- [11] A. E. Wijaya and D. Alfian, "SISTEM REKOMENDASI LAPTOP MENGGUNAKAN COLLABORATIVE FILTERING DAN CONTENT-BASED FILTERING," *Jurnal Computech & Bisnis*, vol. 12, no. 1, pp. 11–27, 2018.
- [12] S. K. Dirjen *et al.*, "Terakreditasi SINTA Peringkat 4 Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF," 2018.
- [13] A. R. Suhendra, G. N. Anom, C. Putra, and M. Cs, "Website Rekomendasi Anime Dengan Menggunakan Pendekatan Content-Based Filtering Berdasarkan Sinopsis," 2022.
- [14] D. A. Pratiwi and A. Qoiriah, "Sistem Rekomendasi Wedding Organizer Menggunakan Metode Content-Based Filtering Dengan Algoritma Random Forest Regression," *Journal of Informatics and Computer Science*, vol. 03, 2022.
- [15] D. Ayu, N. Safitri, R. Helilintar, and L. S. Wahyuniar, "Sistem Rekomendasi Skincare Menggunakan Metode Content-Based Filtering dan Algoritma Apriori."
- [16] A. Ramadhani, N. Br Sinulingga, H. Thamrin, A. Zulfikar Siregar, U. Islam Negeri Sumatera Utara, and K. Kunci, "Dalihan Na Tolu Jurnal Hukum, Politik dan Komunikasi Indonesia Meninjau Usaha Skincare Abal-Abal Perspektif Hukum Islam," 2022, [Online]. Available: <https://jurnal.seaninstitute.or.id/index.php/Jhui>
- [17] S. K. Dirjen *et al.*, "Terakreditasi SINTA Peringkat 2 Implementasi Metode Content Based Filtering Pada Aplikasi Pencarian Taman Penitipan Anak," *masa berlaku mulai*, vol. 1, no. 3, pp. 163–169, 2017.
- [18] S. K. Dirjen *et al.*, "Terakreditasi SINTA Peringkat 4 Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF," 2018.

- [19] A. Riyani, M. Zidny Naf'an #2, and A. Burhanuddin, "Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen," 2019.
- [20] R. Al Rasyid, D. Handayani, and U. Ningsih, "Penerapan Algoritma TF-IDF dan Cosine Similarity untuk Query Pencarian Pada Dataset Destinasi Wisata," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, p. 2024, 2024, doi: 10.35870/jti.