

SISTEM REKOMENDASI PRODUCT EMINA COSMETICS DENGAN MENGUNAKAN METODE CONTENT - BASED FILTERING

By Fatoni Batari Agung Larasati

SISTEM REKOMENDASI PRODUCT EMINA COSMETICS DENGAN MENGGUNAKAN METODE CONTENT - BASED FILTERING

Fatoni Batari Agung Larasati¹, Hery Februariyanti²

¹²Program Studi Teknik Informatika Universitas Stikubank Semarang

Jl. Trilomba Juang No 1 Semarang 50241
Jl. Kendeng V, Bendan Ngisor, Kec. Semarang Sel., Kota Semarang, Jawa Tengah 50241
¹larasatifatoni26@gmail.com, ²herny@edu.unisbank.ac.id

Abstract

Cosmetics is one of the important needs for women that used to care for and beautify themselves. Emina cosmetic is one of PT Paragon Technology and Innovation product wich have a concept cosmetics for teens and young adults. Over time, variants of emina products will certainly increase. Teen and young adults need more selective when choose a product because majority of them has skin problem like dullness, acne, and oily skin. With the increasing number of products customer need the right product for them, that's why they need a recommender system to recommend products according to products that they need. The recommendation process is carried out by comparing the similarities between the products the customer is looking for and descriptions of existing products. Content-based filtering method can be used to recommend products based on the availability of product description. To calculate the similarity between sentences using the cosine similarity algorithm. First, document will be weighted with tf-idf, then similarity value will be calculated with cosine similarity algorithm. Products containing keywords will be calculated and ranking based on the highest to lowest similarity value. In this study, the product with the highest similarity was obtained with a value of 0.7195.

Keywords : Recommender System, Content-Based Filtering, Cosine Similarity Algorithm, tf-idf, Cosmetic

Abstrak

Kosmetik merupakan salah satu kebutuhan penting bagi wanita yang digunakan untuk merawat dan mempercantik diri. Emina cosmetic merupakan produk kosmetik dari PT Paragon Technology and Innovation dengan mengusung konsep kosmetik untuk remaja dan dewasa muda. Seiring berjalannya waktu, produk emina tentunya akan bertambah varian. Pemilihan produk harus lebih selektif karena target market emina cosmetic yang merupakan remaja mayoritas di usianya muncul masalah kulit seperti kusam, jerawat, berminyak dll. Dengan semakin bertambahnya produk dan kebutuhan untuk memilih produk yang tepat bagi customer, maka dibutuhkan adanya sistem yang dapat merekomendasikan produk sesuai dengan produk yang dibutuhkan customer. Proses rekomedasi dilakukan dengan membandingkan kesamaan antara produk yang dicari customer dengan deskripsi produk yang ada. Metode *content-based filtering* adalah metode yang dapat digunakan untuk merekomendasikan produk berdasarkan ketersediaan konten/ deskripsi produk. Untuk menghitung kesamaan antar kalimat menggunakan algoritma *cosine similarity*. Mulanya dokumen akan dilakukan pembobotan dengan *tf idf*, lalu akan dihitung nilai similaritasnya dengan algoritma *cosine similarity*. Produk yang mengandung kata kunci akan dihitung nilai kemiripannya dan dilakukan perangkingan berdasarkan nilai similaritas tertinggi hingga terendah. Dalam penelitian ini, produk dengan similaritas tertinggi didapat dengan nilai sebesar 0,7195.

Kata kunci : Sistem Rekomendasi, Content-based filtering, Algoritma cosine similarity, tf-idf, kosmetik

1. PENDAHULUAN

Berkembangnya dunia kosmetik dewasa ini semakin melesat, terlebih dengan hadirnya *beauty vlogger* dan *beauty influencer* yang juga ikut mempengaruhi gaya hidup masyarakat tentang penggunaan kosmetik. Menurut penggunaannya, kosmetik dibagi menjadi 2, yaitu skincare yaitu jenis kosmetik yang digunakan untuk merawat kebersihan dan kesehatan kulit, contohnya penyegar, pembersih, pelembab,

serum, dll. Sedangkan yang kedua adalah make up, yaitu jenis kosmetik yang digunakan untuk mempercantik dan menutup ketidaksempurnaan pada kulit wajah, sehingga menghasilkan penampilan yang lebih menarik, contoh bedak, pewarna pipi, lipstick, dsb.

Semakin banyak orang yang menggunakan kosmetik membuat para perusahaan di bidang kecantikan semakin lomba-lomba dalam menghadirkan inovasi. Emina Cosmetics merupakan salah satu brand

21
kosmetik yang diproduksi oleh PT Paragon Technology and Innovation yang merupakan salah satu perusahaan manufaktur kosmetik terbesar di Indonesia. PT Paragon Technology and Innovation memproduksi 3 brand kosmetik, yaitu Wardah Cosmetic, Make Over Cosmetic, dan Emina Cosmetic. Dalam penelitian ini, penulis hanya merancang sitem rekomendasi untuk Brand Emina Cosmetic. Dengan mengusung konsep kosmetik untuk remaja, emina menghadirkan berbagai jenis produk dengan *packaging* yang unik dan harga yang terjangkau sesuai dengan target marketnya. Kulit remaja yang cenderung masih alami dan rentan terhadap masalah kulit seperti kusam, kering, berjerawat, ataupun berminyak. Dengan beragam jenis *make up* dan *skincare* dari brand emina cosmetics, terkadang membuat para customer emina yang kebanyakan adalah remaja dan dewasa muda bingung dalam memilih jenis kosmetik yang diperlukan oleh kulit mereka, maka diperlukan adanya sebuah system rekomendasi yang dapat memberikan saran ataupun rekomendasi kepada para customer untuk memilih product Emina Cosmetics berdasarkan ketertarikan dan kebutuhan dalam pencarian referensi

Dalam membangun system rekomendasi untuk menentukan produk yang sesuai kebutuhan customer, terdapat 8 beberapa metode yang sering digunakan yaitu *Collaborative Filtering*, *Content-Based Filtering*, dan *Hybrid*. Pada *collaborative filtering*, menggunakan riwayat product yang disukai atau riwayat penilaian sebagai dasar untuk menentukan rekomendasi. Pada *content-based* menggunakan kesamaan produk untuk ditawarkan kepada pembeli seperti judul atau deskripsi. Sedangkan metode *Hybrid* menggabungkan metode dua atau lebih metode untuk menghasilkan rekomendasi yang lebih baik. Pada penelitian ini, penulis menggunakan metode *content-based filtering* yang menggunakan ketersediaan konten sebuah item sebagai basis dalam pemberian rekomendasi (Ricci, 2011). Berdasarkan item yang pernah dipilih oleh customer, system dapat merekomendasikan item baru yang memiliki kemiripan deskripsi konten dengan item lama. Jenis barang dibagi berdasarkan vektor komponen pembentuknya, misalnya kategori face make up, kategori lip make up, kategori skincare dll. Dengan adanya sitem rekomendasi, perusahaan berupaya untuk memberikan informasi produk lebih mudah kepada pelanggan untuk memilih produk yang sedang dicari maupun yang direkomendasikan oleh sistem.

2. Tinjauan Pustaka

2
Pengujian akurasi sistem rekomendasi berbasis *content-based filtering* pada dosen pembimbing

dalam Sintesis memanfaatkan dokumen UPP dan dokumen publikasi atau penelitian calon dosen pembimbing sebagai dasar penentuan 2 rekomendasi. Penelitian ini meninjau pengaruh *stopword* dalam tingkat akurasi penentuan rekomendasi dokumen dengan klasifikasi dokumen yang digunakan adalah UPP dan dokumen publikasi atau penelitian calon dosen pembimbing. Penelitian ini menyimpulkan bahwa system rekomendasi dengan proses *stopword removal* masih lebih unggul dibandingkan sistem rekomendasi tanpa proses *stopword removal*. [1]

Penelitian yang merancang sebuah sistem rekomendasi untuk alternative produk kepada customer berdasarkan karakteristik dan history transaksi menggunakan algoritma *cosine similarity* diterapkan untuk menghitung nilai kesamaan suatu produk. Metode *content-based filtering* digunakan untuk menyediakan rekomendasi customer yang sesuai dengan karakteristik perusahaan dan transaksi pembelian yang terjadi. [23]

Penelitian yang menggunakan metode *content-based filtering* lainnya dibuat dengan algoritma apriori yang dapat menganalisis pola-pola kombinasi item. Sistem dapat memberikan rekomendasi produk berdasarkan transaksi belanja yang pernah dilakukan oleh pembeli di e-commerce. Sistem dapat menampilkan produk rekomendasi untuk setiap user member yang aktif. Dan pada setiap detail produk yang ditampilkan, dapat direkomendasikan produk lain berdasarkan prosentasi kemunculan dari produk lain dari user yang sedang menampilkan detail produk pada saat itu. Semakin besar nilai support dan confidence yang diberikan maka akan semakin sedikit rekomendasi yang ditampilkan dan semakin kecil nilai support dan confidence yang diberikan maka akan semakin banyak rekomendasi yang ditampilkan. [3]

Penelitian yang menggunakan dua metode yaitu *content-based filtering* dan *collaborative filtering* bertujuan untuk memberikan saran ataupun rekomendasi laptop kepada customer berdasarkan ketertarikan dan kebutuhan dalam pencarian referensi. Algoritma yang digunakan untuk metode *collaborative filtering* yaitu Adjusted-cosine similarity untuk menghitung kemiripan antar customer, dan algoritma weighted sum untuk perhitungan prediksinya, sedangkan untuk metode *content based filtering* algoritma yang digunakan adalah tf-idf untuk pencarian ketersediaan konten yang 9 da. [4]

Penelitian mengenai system temu kembali informasi pencarian buku perpustakaan dengan metode *content-based filtering*, dimana pengunjung perpustakaan nantinya akan menginput kata kunci buku yang akan dicari, selanjutnya dengan metode *cosine similarity*,

system akan menampilkan buku apa saja yang direkomendasikan berdasarkan kata kunci yang diinput pengunjung.[5]

2.1 Data Mining

Data mining merupakan suatu proses pada database yang besar menggunakan kecerdasan buatan, teknik matematika, statistik, dan machine learning untuk meringkas dan menentukan informasi yang berguna dan pengetahuan yang terkait. Data mining merupakan suatu proses menggali data dari pengetahuan yang berbeda dan menghasilkan sebuah informasi-informasi penting yang dapat dipakai dalam meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Secara teknis, data mining dapat disebut juga sebagai suatu proses untuk mencari korelasi atau pola dari beratus atau beribu field dari sebuah database terkait yang sangat besar” [6]

2.2 Sistem Rekomendasi

Sistem rekomendasi adalah suatu alat atau teknik perangkat lunak yang digunakan untuk melakukan prediksi terhadap sesuatu objek. Sistem rekomendasi dapat memberikan saran-saran item yang bermanfaat bagi pengguna (Ricci, et al., 2011). Saran tersebut dapat digunakan untuk menentukan pengambilan keputusan oleh user, seperti menentukan item yang akan dibeli, menentukan musik yang akan didengarkan, film apa yang akan ditonton, berita mana yang akan dibaca, dsb.

Sistem rekomendasi memiliki 2 inputan yang berbeda, yaitu implicit input dan explicit input (Hu, et al., 2008). Pada implicit input, didapat dengan cara mengamati kebiasaan pengguna, seperti riwayat pemesanan, riwayat penelusuran, pola pencarian, dll. Sedangkan pada explicit input didapat dengan hasil penilaian yang diberikan oleh user, seperti pemberian rating, pemberian tanda favorite item, ataupun thumbs-up, down pada item tertentu

Tipe sistem rekomendasi dibagi dalam beberapa klasifikasi yaitu: content-based recommendation, collaborative-based recommendation, hybrid-based recommendation dan beberapa peneliti menambahkan metode knowledge based recommendation, seperti pada gambar 2.1 yang menunjukkan tabel klasifikasi sistem rekomendasi.

2.3 Content Based Filtering

Berbeda dengan Collaborative Filtering, Content-Based Filtering tidak melibatkan pengguna lain dalam menentukan rekomendasi, namun hanya pengguna itu sendiri. Berdasarkan apa yang

dicari user, algoritma ini hanya akan memilih item dengan konten yang mirip untuk direkomendasikan. Hal ini akan membuat keragaman rekomendasi lebih sedikit, tetapi akan berfungsi baik walaupun pengguna tidak memberikan penilaian (rating) Sebagai contoh, katakanlah pengguna A sangat menyukai film-film dystopian dan komedi gelap. Pengguna B juga menikmati film-film dystopian tetapi tidak pernah menonton komedi gelap. content-based filtering hanya akan terus merekomendasikan film dystopian atau sejenisnya. Tentu saja ada banyak kategori yang dihitung kesamaannya, contoh dalam film dapat dibangun sistem rekomendasi berdasarkan genre saja, atau mungkin sutradara, aktor utama, dsb.

2.4 TF-IDF

Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada information retrieval. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat (Robertson, 2006:45). Metode ini akan menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut.

Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen, dan frekuensi keseluruhan dokumen yang mengandung kata tersebut ya rendah pada kumpulan dokumen. Metode ini akan menghitung bobot setiap token t di dokumen d dengan rumus berikut :

$$IDF = (D/DF) \\ W = TF * (IDF + 1)$$

Keterangan :

25 : jumlah kemunculan kata atau term dalam dokumen

IDF = inverse document frequency

D = jumlah semua dokumen

DF = jumlah dokumen yang mengandung kata (term)

W : bobot setiap dokumen

10

Setelah bobot (W) masing-masing dokumen diketahui, maka dilakukan proses pengurutan dimana semakin besar nilai W, semakin besar tingkat similaritas dokumen tersebut terhadap kata kunci, demikian sebaliknya.

2.5 Cosine Similarity

17 Metode *cosine similarity* merupakan metode untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah vector dengan menggunakan keywords (kata kunci) 18 dari sebuah dokumen sebagai ukuran. *Cosine Similarity* dapat diimplementasikan untuk menghitung nilai kemiripan antar kalimat dan menjadi salah satu teknik untuk mengukur kemiripan teks yang populer. Pada saat metode cosine dibanding 19 dengan algoritma lain yang serupa seperti jaccard, dan k-nearest neighbor (K-NN) yang digunakan pada proses klasifikasi dokumen teks dengan hasil akhir dari percobaan 33 kali dengan key yang berbeda dan total 6326 dokumen didapat metode cosine yang nilai kemiripannya tertinggi yaitu 41% dari metode jaccard 19% dan k-nearest neighbor (K-NN) 40%, karena metode *cosine similarity* mempunyai konsep normalisasi panjang vektor data dengan membandingkan Ngram yang sejajar 26 sama lain dari 2 pembanding Cosine Similarity digunakan untuk melakukan perhitungan kesamaan dokumen. Rumus *cosine similarity* adalah sebagai berikut

$$\cos \alpha = \frac{A \cdot B}{|A| \vee B \vee} = \frac{\sum_{i=1}^n A_i X B_i}{\sqrt{\sum_{i=1}^n}}$$

6

A = Vektor A, yang akan dibandingkan kemiripannya

B = Vektor B, yang akan dibandingkan kemiripannya

A • B = dot product antara vektor A dan vektor B

|A| = panjang vektor A |B| = panjang vektor B

|A||B| = cross product antara |A| dan |B|

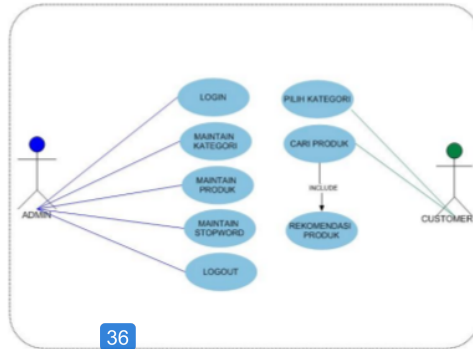
30

3. METODOLOGI PENELITIAN

3.1. Use Case Diagram

Use case diagram sistem rekomendasi Emina Cosmetics dengan metode *content-based filtering* diperlihatkan pada gambar 3.1. Sistem rekomendasi emina cosmetics memiliki 2 aktor, yaitu admin dan customer. Pada aktor admin, memiliki 5 hak akses, yaitu melakukan login untuk masuk ke sistem, maintain kategori, maintain kosmetik, maintain *stopword*, dan

logout. Admin dapat menghapus, mengedit, maupun menambah item yang diinginkan dari menu kategori, kosmetik, dan *stopword*. Pada aktor customer, user dapat mencari produk yang diinginkan dengan memilih kategori dan memasukan keyword produk, selanjutnya sistem akan menampilkan produk yang direkomendasikan.



36

Gambar 3.1 Use Case Diagram

3.2. Pengumpulan Data

Objek dalam penelitian ini adalah Brand Emina Cosmetics yang saat skripsi ini dibuat memiliki 87 varian produk yang berbeda.

Penulis melakukan wawancara langsung dengan assistant trainer di PT Paragon Technology and Innovation, dimana assistant trainer yang memiliki data product knowledge brand kosmetik emina yang merupakan data yang diolah dalam system rekomendasi.

Data produk selanjutnya di klasifikasikan sesuai dengan jenis produk diantaranya adalah :

1. Face Make Up
Berisi produk make up wajah
2. Lip Make Up
Berisi produk make up bibir
3. Eye Make Up
Berisi produk make up mata
4. Cheek Make Up
Berisi produk make up pipi
5. Body Care
Berisi produk perawatan tubuh
6. Lip Care
Berisi produk perawatan bibir
7. Skin care
Berisi produk perawatan wajah
8. Nail
Berisi produk untuk kuku
9. Other

Berisi produk / tools kecantikan

3.3 Activity Diagram Rekomendasi

Customer akan mendapatkan rekomendasi produk pada menu rekomendasi. Pada awalnya, customer akan memilih kategori terlebih dahulu. Kategori diinput oleh admin dengan klasifikasi data yang dibahas pada poin ke 3.2 Pengumpulan data. Selanjutnya customer menginput kata kunci yang diinginkan. Apabila data valid maka sistem akan menyimpan data kata kunci dalam tabel term, kunci, temp, term, dan data rekomendasi produk akan muncul di interface rekomendasi. Data valid disini adalah apabila produk yang kata kunci nya yang diinput oleh customer tersedia dalam system. Sebagai contoh, customer menginputkan kata kunci "krim pemutih wajah" dalam halaman rekomendasi. Kata kunci tersebut tidak akan menghasilkan rekomendasi karena tidak tersedia produk dengan deskripsi "krim pemutih wajah" dalam system rekomendasi emina cosmetics ini. ²⁹ Activity diagram rekomendasi ditunjukkan dalam gambar 3.2 dibawah ini.

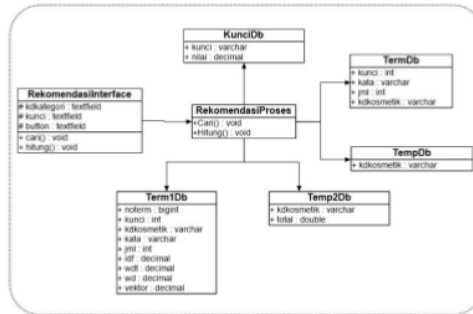
REKOMENDASI



Gambar 3.2 Activity Diagram Rekomendasi

3.4 Class Diagram Rekomendasi

Class diagram rekomendasi menjelaskan alur dalam proses rekomendasi produk. Pada class diagram rekomendasi, customer akan memilih kategori produk dan menginput kata kunci yang diinginkan. Kata kunci yang diinput customer akan tersimpan pada tabel kunci, term, temp, term1, dan temp2.



3.3 Class Diagram Rekomendasi

4. HASIL DAN PEMBAHASAN

Proses rekomendasi emina cosmetics dengan kata kunci "loose powder untuk kulit berminyak" dengan metode *content based filtering* menggunakan algoritma *cosine similarity* dilakukan dengan tahapan sebagai berikut :

1. Text Preprocessing

Lakukan text preprocessing pada deksripsi produk dengan tujuan agar data yang dipakai dapat diproses menjadi angka dengan TF-IDF dan *cosine similarity*. Tahapan dalam proses ini adalah tokenisasi dan *stopword* removal. Preprocessing dilakukan pada item yang mengandung kata kunci saja sehingga proses pencarian tidak memakan waktu yang lama, dimana deskripsi produk yang mengandung kata kunci pada peneliti ²⁸ ini ada 10 dokumen . Hasil dari preprocessing dapat dilihat dalam tabel 4.1

Tabel 4.1 Hasil Preprocessing

Doc	Kode	Nama Produk	Setelah Preprocessing
Q	-	Loose powder untuk kulit ²⁰ erawat	loose powder kulit jerawat
D1	S0001	Bare with me mineral compact powder 01 fair, 14 gr	complexion powder wajah halus bebas kilap seharian warna kulit merata hasil ringan bebas kilap memiliki varian warna fair light beige amber ebony cocok tampil natural memiliki kandungan soft focus agent kesan halus membantu menyerap minyak praktis dibawa bentuk compact
D2	S0002	Emina	complexion powder

		bare with me mineral loose powder 01 fair 8 g	wajah halus bebas kilap seharian warna kulit merata memiliki tekstur ringan halus tahan memiliki 4 pilihan warna fair light beige amber ebony cocok tampil natural memiliki kandungan soft focus agent kesan halus membantu menyerap minyak
D3	S0003	Emina sebum fighter loose powder 8 g	complexion powder mengandung partikel sebum absorbent menyerap minyak berlebih cocok kulit berjerawat rentan berjerawat bedak translucent berwarna less pigmen cocok warna kulit setting powder ringan kulit bebas kilap shine free
D4	S0004	Emina bright stuff loose powder 55 g	complexion powder mencerahkan tampilan wajah kulit wajah halus micro smooth particle dilengkapi glitter tampilan wajah glowing natural matte finish tekstur ringan nyaman mudah dibaurkan efek mencerahkan natural kulit wajah tingkat coverage diatur sheer to full tahan
D5	S0005	Beauty bliss bb cream light, 20 ml	primer makeup wajah dewy halus kelembaban menutupi kemerahan kulit wajah pelembab foundation ringan dipakai sehari-hari diperkaya vitamin c vitamin e menjadikan kulit halus bercahaya mengandung spf 32 melindungi kulit wajah sinar

			matahari warna light
D6	S0006	City chic cc cake peach, 12 gr	complexion powder wajah halus bebas kilap seharian warna kulit merata membantu memperbaiki tekstur kulit bertahap kombinasi unik cc cream powder hasil halus diperkaya apricot antioksidan membantu rehidrasi kulit mudah diaplikasikan longlasting ringan mudah dibaur tahan bebas kilap
D7	S0007	Emina city chic cc cream natural 20 ml	hadir color changing technology mengubah warna krim putih warna sesuai warna kulit shade pilihan light natural menyamarkan pori pori garis halus kerutan wajah mengandung vitamin c vitamin e berperan antioksidan menjaga sel bahaya radikal bebas menjadikan wajah halus mengandung ekstrak aloe vera allantoin agen pelembab menghidrasi kulit
D8	S0009	Emina bare with me mineral cushion 01 light 15 g	dewy matte finish tampilan matte lembab efek kulit sehat segar kandungan oil absorber kulit berminyak efek 1 tingkat cerah perlindungan sinar uva uvb tekstur ringan tahan buildable coverage
D9	S0010	Emina refill bare with me mineral cushion 01 light 15 g	refill cushion dewy matte finish tampilan matte lembab efek kulit sehat segar kandungan oil absorber kulit berminyak efek 1

			tingkat cerah perlindungan sinar uva uvb tekstur ringan tahan buildable coverage
D10	S0095	Emina daily matte loose powder 01 light beige 20 g	complexion powder daily matte loose powder ringan bare with me cocok sehari memiliki coverage rendah bare with me selling pointnya tampilan matte bebas kilap ringan sasaran daily matte customer muda

2. Pembobotan TF-IDF

Pembobotan dilakukan pada deskripsi produk yang mengandung kata kunci, setiap dokumen yang mengandung 16 term diberi nilai 1. Pembobotan setiap term dapat dilihat pada tabel 4.2.

Tabel 4.2 Bobot Term

term	TF			
	Loose	Powder	Kulit	Jerawat
Q	1	1	1	1
D1		1	1	
D2		1	1	
D3		1	1	1
D4		1	1	
D5			1	
D6		1	1	
D7			1	
D8			1	
D9			1	
D10	1	1		

Nilai DF merupakan jumlah dari term yang ditemukan pada setiap dokumen. Hitung nilai IDF dengan rumus $IDF = \log(n/DF)$. Hasil perhitungan idf seperti yang ditampilkan pada tabel 4.3.

Tabel 4.3 Hasil Perhitungan idf

term	DF	D/DF	IDF
Loose	2	5,5	0,7404
Powder	7	1,5714	0,1963
Kulit	10	1,1	0,0414
Jerawat	2	5,5	0,7404

Lakukan perhitungan Wdt dengan rumus $Wdt = tf \cdot idf$, dan didapatkan hasil seperti pada tabel 4.4.

Tabel 4.4 Perhitungan Wdt

term	Wdt=TF.IDF			
	Loose	Powder	Kulit	Jerawat
Q	0,7404	0,1963	0,0414	0,7404
D1	0,0000	0,1963	0,0414	0,0000
D2	0,0000	0,1963	0,0414	0,0000
D3	0,0000	0,1963	0,0414	0,7404
D4	0,0000	0,1963	0,0414	0,0000
D5	0,0000	0,0000	0,0414	0,0000
D6	0,0000	0,1963	0,0414	0,0000
D7	0,0000	0,0000	0,0414	0,0000
D8	0,0000	0,0000	0,0414	0,0000
D9	0,0000	0,0000	0,0414	0,0000
D10	0,7404	0,1963	0,0000	0,0000

3. Perhitungan cosine similarity

Hitung kemiripan vector [kosmetik] query q dengan setiap dokumen yang ada menggunakan rumus cosine similarity. Hitung hasil perkalian skalar antara Q dan 10 dokumen lain. Hasilnya perkalian dari setiap dokumen dengan Q dijumlahkan dan hasilnya diperlihatkan seperti pada tabel 4.5

Tabel 4.5 Perkalian Skalar

term	WD*Wdi				Total
D1	0,0000	0,0385	0,0017	0,0000	0,0402

D2	0,0000	0,0385	0,0017	0,0000	0,0402
D3	0,0000	0,0385	0,0017	0,5481	0,5883
D4	0,0000	0,0385	0,0017	0,0000	0,0402
D5	0,0000	0,0000	0,0017	0,0000	0,0017
D6	0,0000	0,0385	0,0017	0,0000	0,0402
D7	0,0000	0,0000	0,0017	0,0000	0,0017
D8	0,0000	0,0000	0,0017	0,0000	0,0017
D9	0,0000	0,0000	0,0017	0,0000	0,0017
D10	0,5481	0,0385	0,0000	0,0000	0,5866

$$= 0,0402 / 0,2137$$

$$= 0,1881$$

$$\text{Similarity (Q,D3)} = 0,5883 / (1,0660 * 0,7670)$$

$$= 0,5883 / 0,8176$$

$$= 0,7195$$

$$\text{Similarity (Q,D5)} = 0,0017 / (1,0660 * 0,0412)$$

$$= 0,0017 / 0,0439$$

$$= 0,0387$$

Urutan produk yang menjadi rekomendasi dari pencarian dengan kata kunci "loose powder untuk kulit berminyak" dapat dilihat pada tabel 4.7. Produk yang memiliki nilai similaritas yang sama diurutkan berdasarkan abjad. Implementasi pada sistem menghasilkan interface seperti pada gambar 4.1.

Hitung panjang vector setiap dokumen, termasuk Q dengan mengakarkan penjumlahan wdt yang berada pada kolom total. Hasil perhitungan panjang vector dapat dilihat dalam tabel 4.6.

Tabel 4.6 Perhitungan Panjang Vektor

term	Panjang Vektor				Total	Akar
Q	0,5481	0,5481	0,5481	0,5481	1,1364	1,0660
D1	0,0000	0,0385	0,0017	0,0000	0,0402	1,0660
D2	0,0000	0,0385	0,0017	0,0000	0,0402	1,0660
D3	0,0000	0,0385	0,0017	0,5481	0,5883	1,0660
D4	0,0000	0,0385	0,0017	0,0000	0,0402	1,0660
D5	0,0000	0,0000	0,0017	0,0000	0,0017	1,0660
D6	0,0000	0,0385	0,0017	0,0000	0,0402	1,0660
D7	0,0000	0,0000	0,0017	0,0000	0,0017	1,0660
D8	0,0000	0,0000	0,0017	0,0000	0,0017	1,0660
D9	0,0000	0,0000	0,0017	0,0000	0,0017	1,0660
D10	0,5481	0,0385	0,0000	0,0000	0,5866	1,0660

Tabel 4.7 Hasil Rekomendasi

Doc	Nama Produk	Nilai Cosine	Ranking
D3	Emina sebum fighter loose powder 8 g	0,7195	1
D10	Emina daily matte loose powder 01 light beige 20 g	0,7195	2
D6	Bare with me mineral compact powder 01 fair, 14 gr	0,1881	3
D2	City chic cc cake peach, 12 gr	0,1881	4
D4	Emina bare with me mineral loose powder 01 fair 8 g	0,1881	5
D5	Emina bright stuff loose powder 55 g	0,1881	6
D8	Beauty bliss bb cream light, 20 ml	0,0387	7
D7	Emina bare with me mineral cushion 01 light 15 g	0,0387	8
D9	Emina city chic cc cream natural 20 ml	0,0387	9
D1	Emina refill bare with me mineral cushion 01 light 15 g	0,0387	10

Terapkan rumus *cosine similarity* dengan menghitung kemiripan antar dokumen D1 sampai D10. Contoh perhitungan hanya dituliskan dalam D1, D3, dan D5 karena pada dokumen yang lain yaitu D2, D4, D6, D7, D8, D9, dan D10 memiliki nilai perhitungannya yang sama dengan dokumen yang dicontohkan dibawah ini.

$$\text{Similarity (Q,D1)} = 0,0402 / (1,0660 * 0,2005)$$



Gambar 4.1 Hasil Rekomendasi

5. Kesimpulan dan Saran

Metode *content based filtering* menggunakan kesamaan produk untuk ditawarkan kepada pembeli seperti judul atau deskripsi produk, metode ini tidak memerlukan parameter semacam rating untuk menghasilkan suatu rekomendasi. Algoritma *cosine similarity* cocok dipergunakan pada data yang tidak terdapat nilai subjektif seperti similaritas antara teks berdasarkan kemiripan kata dalam teks. Dalam penelitian sistem rekomendasi emina cosmetics dengan metode *content-based filtering* menghasilkan 10 produk rekomendasi dengan hasil akhir perhitungan cosine tertinggi dengan nilai similaritas 0,7195.

Untuk penelitian selanjutnya, produk yang disediakan dalam sistem rekomendasi dapat ditambah menjadi lebih banyak sehingga mendukung hasil rekomendasi produk yang akurat dan semakin beragam serta dapat dikembangkan menjadi platform mobile android.

Daftar Pustaka:

- [1] Parwita, W. G. S. 2019. Pengujian Akurasi Sistem Rekomendasi Berbasis Content-Based Filtering. *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, 14(1), 224-232.
- [2] Fiarni, C., & Maharani, H. 2019. Product Recommendation System Design Using Cosine Similarity and Content-based Filtering Methods. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, 3(2), 42-48.
- [3] Badriyah, T., Fernando, R., & Syarif, I. 2018. Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori. *Konferensi Nasional Sistem Informasi (KNSI) 2018*.
- [4] Wijaya, A., & Alfian, D. 2018. Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering Dan Content-Based Filtering. *Jurnal Computech & Bisnis*, 12(1), 11-27.
- [5] Kurniawati, S. B. 2018. Sistem Temu Kembali Informasi Pencarian Buku Perpustakaan dengan Metode Cosine Similarity (Studi Kasus di Perpustakaan Otoritas Jasa Keuangan Kantor Regional 3 Jawa Tengah dan DIY). *Skripsi, Program Studi Teknik Informatika FTI Unisbank, Semarang*.
- [6] Sun, R., Pelangi, K. C., & Lasena, Y. (2020). PENERAPAN DATA MINING UNTUK MENENTUKAN POTENSI HUJAN HARIAN DENGAN MENGGUNAKAN ALGORITMA K NEAREST NEIGHBOR (KNN). *Jurnal Manajemen Informatika dan Sistem Informasi*, 3(1), 8-15.

SISTEM REKOMENDASI PRODUCT EMINA COSMETICS DENGAN MENGGUNAKAN METODE CONTENT - BASED FILTERING

ORIGINALITY REPORT

23%

SIMILARITY INDEX

PRIMARY SOURCES

1	eprints.upnjatim.ac.id Internet	85 words — 2%
2	e-journals.unmul.ac.id Internet	79 words — 2%
3	digilib.uinsgd.ac.id Internet	70 words — 2%
4	Ying Zhao, Mohammad Noori, Wael A Altabey, Zhishen Wu. "Fatigue damage identification for composite pipeline systems using electrical capacitance sensors", <i>Smart Materials and Structures</i> , 2018 Crossref	65 words — 2%
5	arfianhidayat.com Internet	52 words — 1%
6	docplayer.info Internet	50 words — 1%
7	remeberprefectday.wordpress.com Internet	39 words — 1%
8	repository.its.ac.id Internet	36 words — 1%
9	eprints.unisbank.ac.id Internet	34 words — 1%

10	eprints.itn.ac.id Internet	30 words — 1%
11	www.unisbank.ac.id Internet	27 words — 1%
12	repository.usu.ac.id Internet	27 words — 1%
13	media.neliti.com Internet	26 words — 1%
14	fti.uajy.ac.id Internet	26 words — 1%
15	sinta3.ristekdikti.go.id Internet	25 words — 1%
16	id.123dok.com Internet	24 words — 1%
17	ojs.unikom.ac.id Internet	23 words — 1%
18	www.neliti.com Internet	22 words — 1%
19	biznethotspot.com Internet	20 words — < 1%
20	sparobayafashion.blogspot.com Internet	18 words — < 1%
21	blog.w3loker.com Internet	18 words — < 1%
22	sintadev.ristekdikti.go.id Internet	17 words — < 1%
23	www.coursehero.com Internet	17 words — < 1%

24	jurnal.ugm.ac.id Internet	13 words — < 1%
25	skripsi.narotama.ac.id Internet	12 words — < 1%
26	inacl.id Internet	11 words — < 1%
27	fr.scribd.com Internet	11 words — < 1%
28	rafizanisafahmi.blogspot.com Internet	10 words — < 1%
29	widuri.raharjo.info Internet	10 words — < 1%
30	dokumen.tips Internet	10 words — < 1%
31	www.crossref.org Internet	9 words — < 1%
32	id.scribd.com Internet	9 words — < 1%
33	www.viadewisposts.com Internet	9 words — < 1%
34	cintaihidup.com Internet	9 words — < 1%
35	e-journal.uajy.ac.id Internet	8 words — < 1%
36	lppm-stmikhandayani.ac.id Internet	8 words — < 1%

EXCLUDE QUOTES OFF

EXCLUDE MATCHES OFF

EXCLUDE
BIBLIOGRAPHY OFF