Website Sistem Rekomendasi dengan Content Based Filtering pada Produk Perawatan Kulit

Stefanie Angelina Gunarto^{a1}, Eric Sugiharto Honggara^{a2}, Devi Dwi Purwanto^{a3}

*Departemen Sistem Informasi, Institut Sains dan Teknologi Terpadu Surabaya
Jl. Ngagel Jaya Tengah 73-77, Surabaya, Jawa Timur*

¹stefanie1@mhs.stts.edu

²ericsugiharto@stts.edu

³devi@stts.edu

Abstrak

Pada saat ini produksi dari produk perawatan kulit sangat banyak dan memiliki bahan yang bervariasi. Karena banyaknya produk tersebut maka muncullah kebutuhan dari penjual produk kecantikan untuk memperoleh sebuah rekomendasi dengan menggunakan rekomendasi content based filtering agar customer tetap melakukan pembelian dalam e-commerce yang digunakan. Rekomendasi content based filtering tersebut akan digunakan untuk membantu mencari produk kecantikan dengan bahan baku yang mirip untuk menggantikan sebuah produk yang diinginkan pelanggan namun tidak memiliki stok. Pertama produk akan dibandingkan dengan produk lain berdasarkan komposisi mereka. Kedua, akan dilakukan pemisahan terminology yang terdapat pada konten yang ada dalam produk. Ketiga, data akan diproses dengan membuat index dan melakukan stop word removal. Keempat, algoritma Term Frequency-Invers Document Frequency(TF-IDF) digunakan untuk membuat profil pengguna. Kelima akan dilakukan normalisasi dengan metode cornell smart system. Keenam, melakukan pembobotan berdasarkan profil pengguna. Ketujuh, menghitung cosine similarity dan menentukan rank score dari seluruh produk. Terakhir, menampilkan produk dengan nilai cosine similarity lebih besar dari 10%. Rekomendasi ini akan digunakan dalam sebuah website, diuji dengan blackbox testing oleh masyarakat. Hasilnya adalah sebuah website yang mampu memberikan rekomendasi produk perawatan kulit pada pembeli dengan bahan baku yang serupa dengan yang diinginkan oleh pembeli. Sebagian besar dari masyarakat yang mencoba pun setuju bahwa sistem rekomendasi ini membantu mereka dalam membeli produk perawatan kulit.

Kata kunci: Content Based Filtering, Sistem Rekomendasi, Produk perawatan kulit

Recomennder System Website using Content Based Filtering for Skincare Products

Abstract

These days, there are plethora of skincare with various ingredients. Because there are so many products in the market, there is a need for a recommender system to be used in order to make user become a loyal user to the used e-commerce app. A recommender system using content based filtering is chosen because the usefulness to find another skincare product which has almost identical ingredients. This recommender system will be useful when customer want to buy a product, but the product stock is empty. First, the product will be compared with every product based on their compositions. Secondly, each word terminology in the content of the ingredients will be separated. Third, data will be processed by utilizing index and doing stop word removal. Fourth, Term Frequency-Invers Document Frequency(TF-IDF) algorithm is used to create user persona. Fifth, data will be normalized using cornell smart systems. Sixth, rank all item based on user persona. Lastly, give user the recommended items with cosine similarity higher than 10%. This recommender system will be used on a website and tested using blackbox testing. The result is as follows, users agree that the website able to give recommendation of alternative skincare products which has almost identical ingredients. Thus, the website with recommender system helps them to buy skincare products.

Keywords: Content Based Filtering, Recommender System, Skincare product

I. PENDAHULUAN

Skincare adalah istilah yang sudah familiar pada generasi milenial terutama kaum perempuan. Dimana pada

masa sekarang masyarakat sangat mementingkan penampilan dengan menggunakan skincare untuk memastikan penampilan terjaga setiap harinya. Secara umum, skincare adalah produk yang cara kerja dan bahan-

bahan aktif di dalamnya difungsikan untuk merawat kulit. Bahan-bahan aktif yang dimaksud dalam produk skincare adalah bahan-bahan yang bekerja langsung untuk mengatasi kulit peggunganya.

Berdasarkan survei yang dilaksanakan pada bulan April 2019 oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), diketahui bahwa jumlah pengguna internet di Indonesia sebesar 171,7 juta jiwa (64,8%) dari total populasi penduduk Indonesia yang berjumlah 264 juta jiwa. Pengguna internet tersebut meningkat sebesar 27,9 juta jiwa (10,2%) dibandingkan pengguna internet pada tahun sebelumnya. Kontribusi terbesar atas peningkatan jumlah pengguna internet adalah kaum muda dengan rentang usia 15-35 tahun. Dimana kaum muda Indonesia menggunakan internet untuk mencari informasi / barang-barang yang mereka butuhkan tanpa harus pergi ke tokonya yang dikenal dengan *e-commerce*. Namun produk perawatan kulit sangat bergantung pada jenis kulit yang dimiliki oleh masing-masing orang, sehingga ketika produk telah habis, membutuhkan usaha yang cukup besar untuk mencari produk lain yang bisa cocok dengan kulit pengguna. Hal inilah yang menyebabkan banyak orang menuju toko fisik untuk mendapatkan rekomendasi produk pengganti dengan bahan baku yang mendekati dari produk yang dikehendakinya.

Berdasarkan masalah yang ada, maka timbul ide untuk membuat sebuah website penjualan produk perawatan kulit dilengkapi sistem rekomendasi dengan content based filtering. Rekomendasi dengan content based filtering akan memberikan rekomendasi berdasarkan tingkat kemiripan antara produk pembandingan dan produk-produk yang dibandingkan untuk membantu customer melihat produk yang serupa sehingga mendapat rekomendasi produk yang sesuai dengan preferensi bahan yang diinginkan [1] [2].

II. METODOLOGI

Metode yang digunakan adalah *content-based filtering* yang digunakan untuk memberikan hasil rekomedasi bedasarkan perbandingan konten yang dimiliki oleh produk [3] [4]. Hasil dari metode ini adalah rekomendasi produk yang memiliki tingkat kemiripan paling tinggi ke yang memiliki tingkat kemiripan terendah dengan produk yang dijadikan sebagai pembandingan sehingga customer tertarik membeli produk lain pada website tersebut tanpa mencari dari toko lain [5]. Produk yang direkomendasikan oleh sistem adalah produk yang mirip dengan produk yang diinginkan oleh user di masa lampau [6]. Pada pembangunan website ini, metode content based filtering diterapkan dengan menggunakan algoritma TF-IDF (*Term Frequency – Inverse Document Frequency*) adalah salah satu algoritma yang dapat digunakan untuk menganalisa hubungan antara sebuah frase/kalimat dengan sekumpulan dokumen [7]. Berikut ini adalah langkah-langkah penerapan content-based filtering [8].

- 1) Produk yang tengah dilihat pengguna adalah input yang akan digunakan sebagai produk pembandingan.

- 2) Memisahkan *terminology* yang terdapat pada konten seluruh produk, termasuk produk yang digunakan sebagai pembandingan.
- 3) Menghitung TF, TF adalah jumlah *terminology* muncul pada suatu produk. Berikut ini adalah formula yang dapat digunakan untuk menghitung TF.
 - TF biner (binary TF), hanya memperhatikan apakah suatu kata atau term ada atau tidak dalam dokumen, jika ada diberi nilai satu (1), jika tidak diberi nilai nol (0).
 - TF normalisasi, menggunakan perbandingan antara frekuensi sebuah term dengan nilai maksimum dari keseluruhan atau kumpulan frekuensi term yang ada pada suatu dokumen. Normalisasi data adalah proses membuat variabel memiliki rentang nilai yang sama sehingga data dapat diolah dengan lebih mudah.

$$TF = \{1 + \log (f_{t,d}), f_{t,d} > 0, f_{t,d} = 0 \quad (1)$$

Dimana nilai $f_{t,d}$ adalah frekuensi term (t) pada document (d). Jadi jika suatu kata atau term terdapat dalam suatu dokumen sebanyak 5 kali maka diperoleh bobot = $1 + \log (5) = 1.699$. Tetapi jika term tidak terdapat dalam dokumen tersebut, bobotnya adalah nol (0).

- 4) Menghitung IDF, IDF adalah sebuah perhitungan dari bagaimana term didistribusikan secara luas pada koleksi dokumen yang bersangkutan.

$$idf_i = \log \left(\frac{n}{df_i} \right) \quad (2)$$

Dimana df adalah jumlah produk yang mengandung *terminology* i dan n adalah jumlah seluruh produk.

- 5) Bobot kemungkinan minat pengguna terhadap setiap produk berdasarkan profil pengguna.

$$w_{i,d} = tf_{i,d} * \log \left(\frac{n}{df_i} \right) \quad (3)$$

Dimana tf adalah nilai *terminology frekuensi* yang menunjukkan apakah suatu *terminology* muncul atau tidak, dan n adalah jumlah produk selain produk yang digunakan sebagai pembandingan.

- 6) Menghitung *cosine similarity* setiap produk untuk mengetahui tingkat kemiripan antara produk pembandingan dan produk yang dibandingkan.

$$CosSim(d_i, q) = \frac{d_i \rightarrow q}{|d_i \rightarrow q|} = \frac{\sum_{i=1}^n \frac{1}{i=1} (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^n \frac{1}{i=1} w_{ij}^2 \cdot \sum_{i=1}^n \frac{1}{i=1} w_{iq}^2}} \quad (4)$$

- 7) Menentukan *rank score* seluruh produk dengan mengalikan *cosine similarity* produk pembandingan dengan *cosine similarity* setiap produk.
- 8) Menampilkan produk yang memiliki nilai *Cosine Similarity* $\geq 10\%$ sebagai produk yang direkomendasikan. *Cosine Similarity* $\geq 10\%$ menunjukkan bahwa produk yang dibandingkan memiliki tingkat kesamaan tertentu dengan produk yang digunakan sebagai pembandingan.

Sistem akan diuji coba dengan menggunakan metode blackbox testing [9] dan kuisioner yang diberikan untuk melihat seberapa baik sistem rekomendasi diberikan pada pengguna.

III. HASIL DAN PEMBAHASAN

Pada bagian ini, akan dijelaskan mengenai hasil uji coba yang telah dilakukan pada website. Uji coba yang dilakukan adalah uji coba perangkat lunak yang dilakukan setelah website ini di-hosting. Uji coba perangkat lunak adalah proses yang dirancang untuk memastikan program komputer menjalankan apa yang seharusnya dilakukan. Berikut ini adalah perhitungan dan hasil uji coba pada website penjualan produk perawatan kulit dilengkapi sistem rekomendasi dengan content based filtering untuk fitur rekomendasi produk.

- a. Produk yang tengah dilihat pengguna adalah input yang akan digunakan sebagai produk pembanding. Berikut ini adalah tabel yang berisikan contoh data produk yang dijadikan produk pembanding.

TABEL I
CONTOH PRODUK PEMBANDING

id	nama_produk	Komposisi_produk	tipe_kulit	kategori_produk
17	Garnier Light Complete Super UV Spot Proof Sunscreen Matte Finish 30ml	Niacinamide; Disodium EDTA; Dipropylene Glycol; Maltodextrin; Menthoxypropan ediol	dry skin	Facial Sunscreen

- b. Memisahkan terminology yang terdapat pada konten seluruh produk, termasuk produk yang digunakan sebagai pembanding. Berikut ini adalah tabel contoh produk yang akan dibandingkan.

TABEL II
CONTOH PRODUK DIBANDINGKAN

id	nama_produk	Komposisi_produk	tipe_kulit	kategori_produk
6	Senka Perfect Whip 120gr	Gluceryl Stearate SE; Disodium EDTA; Dipropylene Glycol; Lauric Acid; Myristic Acid	dry skin	Face wash
2	Aloe BHA Skin Toner 200ml	1,2-Hexanediol; Psidium Guajava Fruit Extract; Snail Secretion Filtrate; Aspalathus Linearis Extract; Protulaca Oleracea Extract	dry skin, oily skin, acne skin	Toner
12	PURITO Centella Unscented Serum 60ml	1,2-Hexanediol; Glycereth-26; Niacinamide; Disodium EDTA; Dipotassium Glycyrrhizate	dry skin, oily skin	Serum
	COSRX Low pH Good Morning	1,2-Hexanediol; Pueraria Labata Root Extract; Disodium EDTA;	dry skin, oily skin, acne skin	Face wash

	Cleanser 50ml	Betaine Salicylate; Sodium Lauroyl Methyl Isethionate		
21	SECRET KEY Starting Treatment Eye Cream 30gr	1,2-Hexanediol; Glyceryl Stearate; Niacinamide; Disodium EDTA; Portulaca Oleracea Extract	dry skin	Eye Care

- c. Berikutnya melakukan langkah pemrosesan data yang terdiri dari tokenisasi yaitu proses pembuatan index dan stop word removal yaitu penghapusan kata yang muncul lebih dari sekali dari bahan yang ada. Stop word removal ini berguna untuk menghemat tempat penyimpanan dan waktu yang dibutuhkan untuk melakukan komputasi [10]. Berikut ini adalah tabel hasil pemrosesan pemisahan terminology.

TABEL III
TERMINOLOGY

Terminology
Acne Skin
Dry Skin
Oily Skin
Niacinamide
Disodium EDTA
Dipropylene Glycol
Maltodextrin
Menthoxypropanediol
Gluceryl Stearate SE
Lauric Acid
Myristic Acid
1,2-Hexanediol
Glycereth-26
Dipotassium Glycyrrhizate
Glyceryl Stearate
Portulaca Oleracea Extract
Psidium Guajava Fruit Extract
Snail Secretion Filtrate
Aspalathus Linearis Extract
Pueraria Labata Root Extract
Betaine Salicylate
Sodium Lauroyl Methyl Isethionate

Menggunakan algoritma Term Frequency-Invers Document Frequency (TF-IDF) untuk membuat profil pengguna. TF adalah jumlah terminology pada suatu produk, sementara IDF dapat dihitung menggunakan persamaan berikut ini.

$$idf_i = \log \left(\frac{n}{df_i} \right)$$

Dimana df adalah jumlah produk yang mengandung terminology i dan n adalah jumlah seluruh produk. Berikut ini adalah tabel proses perhitungan nilai tf dan df.

TABEL IV
PERHITUNGAN NILAI tf DAN df

Terminology	Q	tf					df	idf = log(n/df)
		D1	D2	D3	D4	D5		
Acne Skin	0	0	1	0	1	0	2	0,39794
Dry Skin	1	1	1	1	1	1	5	0
Oily Skin	0	0	1	1	1	0	3	0,221849
Niacinamide	1	0	0	1	0	1	2	0,39794

Disodium EDTA	1	1	0	1	1	1	4	0,09691
Dipropylene Glycol	1	1	0	0	0	0	1	0,69897
Maltodextrin	1	0	0	0	0	0	0	0
Menthoxypropandiol	1	0	0	0	0	0	0	0
Gluceryl Strearate SE	0	1	0	0	0	0	1	0,69897
Lauric Acid	0	1	0	0	0	0	1	0,69897
Myristic Acid	0	1	0	0	0	0	1	0,69897
1,2-Hexanediol	0	0	1	1	1	1	4	0,09691
Glycereth-26	0	0	0	1	0	0	1	0,69897
Dipotassium Glycyrrhizate	0	0	0	1	0	0	1	0,69897
Glyceryl Strearate	0	0	0	0	0	1	1	0,69897
Portulaca Oleracea	0	0	1	0	0	1	2	0,39794
Extract Betaine Salicylate	0	0	0	0	1	0	1	0,69897
Psidium Guajava Fruit	0	0	1	0	0	0	1	0,69897
Extract Snail Secretion	0	0	1	0	0	0	1	0,69897
Filtrate Aspalathus Linearis	0	0	1	0	0	0	1	0,69897
Extract Pueraria Labata Root	0	0	0	0	1	0	1	0,69897
Extract Sodium Lauroyl Methyl Isethionate	0	0	0	0	1	0	1	0,69897

e. Melakukan normalisasi terhadap Q dan tf dengan metode cornell smart system. Normalisasi data adalah proses membuat variabel memiliki rentang nilai yang sama sehingga data dapat diolah dengan lebih mudah.

TABEL V NORMALISASI

Q	tf				
	D1	D2	D3	D4	D5
0	0	1	0	1	0
1	1	1	1	1	1
0	0	1	1	1	0
1	0	0	1	0	1
1	1	0	1	1	1
1	1	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0
0	1	0	0	0	0
0	1	0	0	0	0
0	1	0	0	0	0
0	0	1	1	1	1
0	0	0	1	0	0
0	0	0	1	0	0
0	0	0	0	0	1
0	0	1	0	0	1
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0

0	0	0	0	1	0
0	0	0	0	1	0
0	0	0	0	1	0

f. Bobot kemungkinan minat pengguna terhadap setiap produk berdasarkan profil pengguna. Berikut ini adalah persamaan yang digunakan.

$$w_{i,d} = tf_{i,d} * \log\left(\frac{n}{df_i}\right)$$

Berikut ini adalah tabel proses perhitungan bobot.

TABEL VI PERHITUNGAN NILAI W

Qid	W				
	D1	D2	D3	D4	D5
0	0	0,39794	0	0,39794	0
0	0	0	0	0	0
0	0	0,221849	9	0,221849	0
0,39794	0	0	0,39794	0	0,39794
0,09691	0,09691	0	0,09691	0,09691	0,09691
0,69897	0,69897	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0,69897	0	0	0	0
0	0	0,09691	0,09691	0,09691	0,09691
0	0	0	0,69897	0	0
0	0	0	0,69897	0	0
0	0	0	0	0	0,69897
0	0	0,39794	0	0	0,39794
0	0	0,69897	0	0	0
0	0	0,69897	0	0	0
0	0	0,69897	0	0	0
0	0	0	0	0,69897	0
0	0	0	0	0,69897	0
0	0	0	0	0,69897	0

Dimana *tf* adalah nilai *terminology frekuensi* yang menunjukkan apakah suatu *terminology* muncul atau tidak, dan *n* adalah jumlah produk selain produk yang digunakan sebagai pembanding.

g. Menghitung *cosine similarity* setiap produk untuk mengetahui tingkat kemiripan antara produk pembanding dan produk yang dibandingkan [11]. Berikut ini adalah persamaan yang dapat digunakan untuk perhitungan *cosine similarity*.

$$CosSim(d_i, q) = \frac{d_i \rightarrow q}{|d_i \rightarrow| \cdot |q \rightarrow|} = \frac{\sum_{i=1}^1 \frac{1}{i-1} (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^1 \frac{1}{i-1} w_{ij}^2 \cdot \sum_{i=1}^1 \frac{1}{i-1} w_{iq}^2}}$$

TABEL VII PERHITUNGAN COSINE SIMILARITY

Perhitungan Cosine Similarity					
Q ²	D1 ²	D2 ²	D3 ²	D4 ²	D5 ²
0	0	0,158356	0	0,158356	0
0	0	0	0	0	0
0	0	0,049217	0,049217	0,049217	0
0,158356	0	0	0,158356	0	0,158356
0,009392	0,009392	0	0,009392	0,009392	0,009392
0,488559	0,488559	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0,488559	0	0	0	0
0	0	0,009392	0,009392	0,009392	0,009392
0	0	0	0,488559	0	0
0	0	0	0,488559	0	0
0	0	0	0	0	0,488559
0	0	0,158356	0	0	0,158356
0	0	0,488559	0	0	0
0	0	0,488559	0	0	0

kulit diperlukan seorang pakar. Hal ini selaras dengan kesimpulan dari penelitian yang juga pada topik sejenis [15].

Daftar Pustaka

- [1] R. P. Koesmyta, "Pembuatan E-Commerce Berbasis Web Dengan Sistem Rekomendasi Content-Based Filtering Menggunakan Algoritma Apriori," *Diss. Universitas 17 Agustus 1945 Surabaya*, 2020.
- [2] S. A. W. Y. a. I. S. T. Badriyah, "'Recommendation system for property search using content based filtering method,'" in *2018 International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta, Indonesia, 2018.
- [3] R. v. Meteren and v. M. Someren, "Using Content-Based Filtering for Recommendation," *Proceedings of the machine learning in the new information age : Proceedings of the machine learning in the new information age*, vol. 30, pp. 47-56, 2000.
- [4] R. A. Nugroho, A. M. Polina and Y. D. Mahendra, "Tourism Site Recommender System Using Item-Based Collaborative Filtering Approach," *international Journal of Applied Science and Smart Technologies*, vol. 2, no. 2, pp. 209-216, 2020.
- [5] M. Muhammad and S. Sugiyanto, "Item Based Collaborative Filtering Based on Highest Item Similarity," *International Journal of Artificial Intelligence Research*, vol. 6, no. 1, 2022.
- [6] P. Lops, M. De Gemmis and G. Semeraro, Content-based recommender systems : State of the art and trends. In *Recommender Systems Handbook*, Springer, 2011, pp. 73-105.
- [7] M. N. Saadah, R. W. Atmagi, D. S. Rahayu and A. Z. Arifin, "Information Retrieval of Text Document With Weighting TF-IDF and LCS," *Jurnal Ilmu Komputer dan Informasi*, vol. 6, no. 1, pp. 34-37, 2013.
- [8] H. Christian, M. P. Agus and D. Suhartono, "Single Document Automatic Text Summarization using Term Frequency-Inverse Document Frequency(TF-IDF)," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 7, no. 4, pp. 285-294, 2016.
- [9] S. Nidhra and J. Dondeti, "BLACK BOX AND WHITE BOX TESTING TECHNIQUES –A LITERATURE REVIEW," *International Journal of Embedded Systems and Applications (IJESA)*, vol. 2, no. 2, pp. 29-50, 2012.
- [10] C. Silva and B. Ribeiro, "The Importance of stop word removal on recall values in text categorization," in *Proceedings of the International Joint Conference on Neural Networks*, Portland, 2003.
- [11] B. Li and L. Han, "Distance Weighted Cosine Similarity Measure for Text Classification," in *Intelligent Data Engineering and Automated Learning - IDEAL 2013*, Hefei, 2013.
- [12] J. C. Patty, E. T. Kirana and M. S. D. K. Giri, "Recommendations System fo Purchase of Cosmetic Using Content-Based Filtering," *International Journal of Computer Engineering and Information Technology*, vol. 10, no. 1, pp. 1-5, 2018.
- [13] G. Chandrasekaran and D. J. Hemanth, "An Intelligent Frmaework for Online Product Recommendation Using Collaborative Filtering," in *Proceedings of International Conference on Sustainable Expert Systems*, Nepal, 2021.
- [14] F. B. A. Larasati and H. Februyanti, "SISTEM REKOMENDASI PRODUCT EMINA COSMETICS DENGAN MENGGUNAKAN METODE CONTENT - BASED FILTERING," *Jurnal Manajemen Informatika & Sistem Informasi*, vol. 4, no. 1, pp. 45-54, 2021.
- [15] F. L. L. V. FC and L. , "Web-Based Expert System Application in Determining Facial Care," in *International Conference on Environment and Technology*, 2020.