

**SISTEM REKOMENDASI PEMILIHAN PRODUK
SKINCARE DENGAN PENDEKATAN *CONTENT-
BASED FILTERING***

(Studi Kasus: *Review* Produk *Skincare* di *Female Daily*)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Program
Studi Statistika



Disusun Oleh:

Farah Ayu Fitriani

18611002

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2022**

**HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR**

Judul : SISTEM REKOMENDASI PEMILIHAN
PRODUK SKINCARE DENGAN
PENDEKATAN *CONTENT-BASED FILTERING*
(Studi Kasus: *Review* Produk *Skincare* di Female
Daily).

Nama Mahasiswa : Farah Ayu Fitriani

NIM : 18611002

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

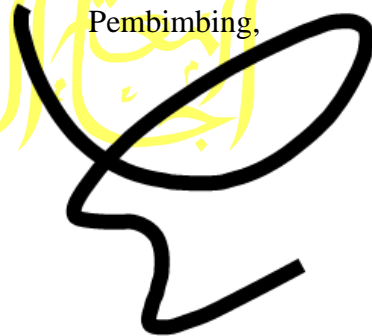
Ketua Prodi Statistika,

Yogyakarta, 13 April 2022

Pembimbing,



(Dr. Edy Widodo, S.Si., M.Si)



(Tuti Purwaningsih, S.Stat., M.Si.)

HALAMAN PENGESAHAN
TUGAS AKHIR

**SISTEM REKOMENDASI PEMILIHAN PRODUK *SKINCARE* DENGAN
PENDEKATAN *CONTENT-BASED FILTERING***
(Studi Kasus: *Review* Produk *Skincare* di *Female Daily*)

Nama Mahasiswa : Farah Ayu Fitriani

NIM : 18611002

**TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL : 26 April 2022**

Nama Penguji


1. Raden Bagus Fajriya Hakim, Dr., S.Si., M.Si.
2. Rahmadi Yotenka, S.Si., M.Sc.
3. Tuti Purwaningsih, S.Stat., M.Si.

Tanda Tangan



Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Prof. Riyanto, S.Pd., M.Si., Ph.D.)

KATA PENGANTAR



Assalamu'alaikum Wr.Wb

Alhamdulillah Robbil 'Alamin, Segala puji dan syukur atas kehadiran Allah SWT karena atas rahmat dan karunia-Nya penulis diberi keimanan, kekuatan, kesabaran, kelancaran serta keselamatan selama menyusun tugas akhir yang berjudul “Sistem Rekomendasi Pemilihan Produk *Skincare* Dengan Pendekatan *Content-Based Filtering*”. Shalawat serta salam tak lupa juga selalu tercurah kepada junjungan kita Nabi Muhammad SAW beserta keluarga, sahabat dan para pengikutnya hingga akhir zaman.

Tugas akhir ini disusun sebagai salah satu syarat untuk mendapatkan gelar sarjana Jurusan Statistika di Universitas Islam Indonesia. Penulis telah banyak mendapat bimbingan dan dukungan dari berbagai pihak dalam penyusunan tugas akhir ini, sehingga pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sedalam-dalamnya kepada:

1. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta jajarannya.
2. Bapak Dr. Edy Widodo, S.Si., M.Si., selaku Ketua Jurusan Statistika beserta seluruh jajarannya.
3. Ibu Tuti Purwaningsih, S.Stat., M.Si., selaku dosen pembimbing yang telah memberi bimbingan dan arahan selama mengerjakan tugas akhir ini.
4. Bapak Achmad Fauzan, S.Pd., M.Si., selaku Dosen Pembimbing Akademik yang telah memberikan arahan dan dukungan selama masa perkuliahan.
5. Seluruh Dosen Program Studi Statistika Universitas Islam Indonesia yang memberikan banyak ilmu serta pengalaman yang berharga selama perkuliahan.
6. Kedua Orang tua tercinta, Papa dan Mama yang telah memberikan segala dukungan, do'a, dan kasih sayang dari awal proses perkuliahan hingga selesai.

7. Para sahabat, khususnya Konni Callista Asyisyifaa dan Siti Salsabila Rifdah Kharisma yang selalu mendukung penulis, mendengarkan keluh kesah dan menemani penulis dari awal perkuliahan sampai dapat menyelesaikan tugas akhir ini.
8. Semua teman – teman seperjuangan Statistika Universitas Islam Indonesia, khususnya Angkatan 2018 yang senantiasa berbagi suka cita, serta menemani selama perkuliahan dan mengerjakan tugas akhir ini.
9. Kesayangan cici; Mbeb, Lady, Katty, Jack, dan John terimakasih sudah selalu ada untuk cici, menemani dan menghibur cici. Terkhusus untuk Jack, semoga bahagia di surga ya sayang.
10. Semua pihak yang telah memberikan dukungan dan bantuan yang tidak dapat penulis sebutkan satu per satu, terima kasih.
11. Yang paling akhir, saya mengucapkan terimakasih kepada diri saya sendiri sudah bertahan sampai saat ini melewati segala cobaan, rintangan, segala permasalahan yang ada dengan kesabaran dan hati yang ikhlas.

Penulis menyadari bahwa masih terdapat kesalahan dan jauh dari kata sempurna, oleh karena itu penulis akan menerima seluruh kritik dan saran yang bersifat membangun. Penulis berharap semoga penelitian ini dapat bermanfaat khususnya untuk penulis dan semua yang membutuhkan. Akhir kata, penulis ingin mengucapkan banyak terima kasih dan semoga Allah SWT selalu memberikan rahmat dan karunia-Nya bagi kita semua, Aamiin.

Wassalamu'alaikum, Wr.Wb.

Yogyakarta, April 2022

Farah Ayu Fitriani

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR.....	ii
HALAMAN PENGESAHAN TUGAS AKHIR	iii
KATA PENGANTAR.....	iv
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN	x
PERNYATAAN	xi
INTISARI.....	xii
ABSTRACT	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	6
BAB III LANDASAN TEORI	10
3.1. <i>E-Commerce</i>	10
3.2. <i>Female Daily</i>	11
3.3. <i>Skincare</i>	12
3.4. <i>Text Mining</i>	13
3.4.1 Vektorisasi Kata	14
3.4.2 <i>Text Pre-processing</i>	14
3.5. TF-IDF	15
3.6. <i>Cosine Similarity</i>	16
3.7. Sistem Rekomendasi	17
3.7.1 <i>Collaborative Filtering</i>	18
3.7.2 <i>Content-Based Filtering</i>	20
BAB IV METODOLOGI PENELITIAN.....	22
4.1. Populasi dan Sampel Data.....	22
4.2. Variabel Penelitian	22
4.3. Metode Pengambilan Data	22
4.4. Metode Analisis Data	23
4.5. Diagram Alir Penelitian	23
BAB V HASIL DAN PEMBAHASAN	25
5.1. Analisis Deskriptif	25
5.2. <i>Pre-processing Data</i>	26
5.2.1 <i>Case Folding</i>	27
5.2.2 <i>Tokenizing</i>	28
5.2.3 <i>Filtering</i>	29
5.2.4 <i>Stemming</i>	30
5.3. Pembobotan <i>Term Frequency-Invers Document Frequency</i> (TF-IDF)	31

5.4. <i>Cosine Similarity</i>	34
5.5. Hasil Rekomendasi.....	35
5.6. Tampilan <i>Website</i>	36
BAB VI PENUTUP.....	38
6.1. Kesimpulan	38
6.2. Saran.....	38
DAFTAR PUSTAKA.....	40
LAMPIRAN	43

DAFTAR TABEL

Tabel 2.1. Tabel Penelitian Sebelumnya	8
Tabel 4.1. Definisi Operasional Variabel.....	22
Tabel 5.1. Contoh dokumen pada tahap <i>case folding</i>	27
Tabel 5.2. Contoh dokumen pada tahap <i>tokenizing</i>	28
Tabel 5.3. Contoh dokumen pada tahap <i>filtering</i>	29
Tabel 5.4. Contoh dokumen pada tahap <i>stemming</i>	31
Tabel 5.5. Contoh hasil perhitungan TF.....	32
Tabel 5.6. Contoh hasil pembobotan TF-IDF	33
Tabel 5.7. 5 Dokumen yang paling mirip dengan dokumen ke-14716.....	35
Tabel 5.8. 5 Rekomendasi produk “Antioxidant Serum”	35
Tabel 5.9. <i>Review</i> produk “Antioxidant Serum” dan “Antioxidant Cleansing Foam”	36

DAFTAR GAMBAR

Gambar 3.1. Logo Female Daily	11
Gambar 3.2. Ilustrasi proses <i>collaborative filtering</i>	19
Gambar 3.3. Ilustrasi Proses <i>Content-Based Filtering</i>	21
Gambar 4.1. <i>Flowchart</i> penelitian.....	23
Gambar 5.1. 10 <i>Brand</i> dengan frekuensi terbanyak.....	25
Gambar 5.2. Rata-rata <i>rating</i> produk <i>skincare</i>	26
Gambar 5.3. Tampilan awal <i>website</i>	37
Gambar 5.4. Tampilan hasil rekomendasi.....	37

DAFTAR LAMPIRAN

Lampiran 1 Data <i>Review</i> Produk <i>skincare</i> di Femaledaily	43
Lampiran 2 <i>Script</i> dan <i>output</i> sistem rekomendasi	47

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 20 April 2022



Farah Ayu Fitriani

INTISARI

SISTEM REKOMENDASI PEMILIHAN PRODUK SKINCARE DENGAN PENDEKATAN CONTENT-BASED FILTERING

(Studi Kasus: *Review* produk *Skincare* di *Female Daily*)

Farah Ayu Fitriani

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Kulit merupakan organ tubuh terbesar yang menutupi seluruh permukaan tubuh manusia. Merawat kesehatan kulit merupakan salah satu bagian dari menjaga kesehatan tubuh yang sangat penting untuk dilakukan. Di Indonesia, tren dalam perawatan kulit atau yang saat ini dikenal dengan sebutan *skincare* terus berkembang hingga saat ini. Kebutuhan informasi bagi masyarakat khususnya dalam memilih produk *skincare* saat ini tentunya sangat dibutuhkan. *Female Daily Network* merupakan sebuah situs media informasi yang menyediakan *review* atau ulasan dari berbagai macam produk *skincare*. *Female Daily* diakui sebagai *platform* kecantikan nomor satu di Indonesia. Permasalahan yang ada saat ini yaitu banyak orang merasa sulit dalam memilih produk *skincare* karena produk yang beredar di pasaran sangat banyak, selain itu juga karena iklan yang hanya memberikan sisi positifnya saja. Berdasarkan permasalahan tersebut, diperlukan solusi berupa model sistem rekomendasi dalam memilih produk *skincare* yang diperoleh dari *review* pengguna. Penelitian ini bertujuan untuk mengetahui rekomendasi suatu produk *skincare* berdasarkan *review* menggunakan *Content-Based Filtering* dan mengetahui hasil rekomendasi produk *skincare* berdasarkan *review* yang diberikan oleh sistem rekomendasi menggunakan *Content-Based Filtering*. Metode *Content-Based Filtering* menggunakan konsep pembobotan TF-IDF dan perhitungan *Cosine Similarity*. Pada proses mendapatkan hasil rekomendasi, digunakan produk “Antioxidant Serum” sebagai contoh pengujian. Didapatkan hasil rekomendasi yang didapatkan yaitu memberikan lima rekomendasi produk terhadap pengguna (*user*) berdasarkan *review* pada produk “Antioxidant Serum” berturut-turut yaitu “Antioxidant Cleansing Foam”, “Antioxidant Toner”, “Antioxidant Soothing Gel”, “Antioxidant Night Cream”, dan “Antioxidant Morning Cream” dengan nilai *cosine similarity* berturut-turut adalah 0.875379, 0.772920, 0.670935, 0.385445, 0.296287.

Kata Kunci : *Skincare*, Sistem Rekomendasi, *Content-Based Filtering*

ABSTRACT

SKINCARE PRODUCT SELECTION RECOMMENDATION SYSTEM USING CONTENT-BASED FILTERING METHOD

(Case Study: Review of Skincare Products at Femaledaily)

Farah Ayu Fitriani

Department of Statistics, Faculty of Mathematics and Natural Sciences
Universitas Islam Indonesia

The skin is the largest organ of the body that covers the entire surface of the human body. Caring for skin health is one part of maintaining a healthy body which is very important to do. In Indonesia, the trend in skin care or what is currently known as skincare continues to grow today. The need for information for the community, especially in choosing skincare products at this time, is certainly very much needed. Female Daily Network is an information media site that provides reviews or reviews of various kinds of skincare products. Female Daily is recognized as the number one beauty platform in Indonesia. The problem that exists today is that many people find it difficult to choose skincare products because there are so many products on the market, but also because advertisements only give the positive side. Based on these problems, a solution is needed in the form of a recommendation system model in choosing skincare products obtained from user reviews. This study aims to determine the recommendation of a skincare product based on a review using Content-Based Filtering and to find out the results of a skincare product recommendation based on a review given by a recommendation system using Content-Based Filtering. The Content-Based Filtering method uses the concept of TF-IDF weighting and the calculation of Cosine Similarity. In the process of getting the recommendation results, the product "Antioxidant Serum" is used as an example of the test. The recommendation results obtained are providing five product recommendations to users based on reviews on the "Antioxidant Serum" product in a row, namely "Antioxidant Cleansing Foam", "Antioxidant Toner", "Antioxidant Soothing Gel", "Antioxidant Night Cream", and "Antioxidant Morning Cream" with cosine similarity values are 0.875379, 0.772920, 0.670935, 0.385445, 0.296287.

Keywords: *Skincare, Recommendation System, Content-Based Filtering*

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Kulit merupakan organ tubuh terbesar yang menutupi seluruh permukaan tubuh manusia (Swari, 2021). Struktur kulit terbagi menjadi tiga lapisan utama yaitu epidermis, dermis, dan hipodermis. Tiap lapisan mempunyai fungsi yang berbeda, diantaranya yaitu membentuk sistem pertahanan kulit, mengatur suhu tubuh, mengendalikan kehilangan cairan, dan sebagainya. Secara keseluruhan, kulit mempunyai peran penting sebagai pelindung tubuh dari cedera dan patogen sehingga merawat kesehatan kulit merupakan salah satu bagian dari menjaga kesehatan tubuh yang sangat penting untuk dilakukan. Menurut *National Institute of Arthritis and Musculoskeletal and Skin Diseases*, menjaga kulit sehat artinya juga menjaga tubuh kita dari beragam bakteri yang bisa membahayakan otot, internal organ, bahkan tulang kita (Amanda, 2021).

Di Indonesia, tren dalam perawatan kulit atau yang saat ini dikenal dengan sebutan *skincare* terus berkembang hingga saat ini. Perkembangan pada berbagai jenis perawatan kulit disebabkan oleh banyak faktor, diantaranya yaitu banyaknya permasalahan kulit yang muncul, perkembangan teknologi, banyaknya penelitian mengenai zat-zat tertentu, dan semakin banyak orang yang menyebarkan berbagai informasi mengenai pentingnya menjaga kesehatan kulit. Meskipun arti dari *skincare* secara umum yaitu perawatan kulit yang mencakup seluruh tubuh, namun kini *skincare* yang banyak digunakan dan dipahami masyarakat lebih banyak mengacu pada perawatan kulit wajah. Terdapat berbagai macam cara dalam melakukan perawatan kulit. Beberapa orang cenderung untuk berkonsultasi kepada dokter kulit atau berkunjung ke klinik kecantikan. Namun, dalam melakukan hal tersebut terdapat beberapa permasalahan yang dialami yaitu biaya yang relatif mahal, tidak adanya klinik kecantikan di daerah tertentu, dan masa pandemi covid-19 yang mengharuskan masyarakat tidak keluar rumah. Permasalahan tersebut juga menjadi penyebab orang – orang mulai beralih untuk melakukan perawatan kulit secara mandiri dengan memilih produk *skincare* yang

dijual di pasaran sebagai solusi untuk menyelesaikan permasalahan pada kulit wajahnya.

Terdapat beberapa kriteria yang perlu diperhatikan dalam memilih produk *skincare*, hal pertama yang harus dilakukan yaitu mengetahui jenis kulit. Terdapat beberapa jenis kulit yaitu normal, berminyak, kering, kombinasi, maupun sensitif. Selain itu, pemilihan *skincare* juga didasarkan oleh permasalahan kulit yang dimiliki sehingga dapat menyesuaikan produk *skincare* yang akan digunakan. Jenis produk *skincare* yang digunakan secara umum yaitu sabun pembersih wajah (*facial wash*), *toner*, pelembab (*moisturizer*), *serum*, dan tabir surya (*sunscreen*). Namun selain penggunaan *skincare* secara umum, tentunya harus didukung dengan penggunaan produk *skincare* pendukung agar dapat memperoleh hasil lebih maksimal pada kulit. Jenis produk *skincare* pendukung diantaranya yaitu *cream*, *lotion*, *oil*, masker wajah, *scrub*, *exfoliator*, *face mist*, dan masih banyak lagi. Banyaknya produk *skincare* yang bermunculan didukung dengan promosi yang menarik dari setiap *brand skincare* membuat banyak orang merasa kesulitan dalam memilih produk *skincare* di pasaran.

Kebutuhan informasi bagi masyarakat khususnya dalam memilih produk *skincare* saat ini tentunya sangat dibutuhkan. Pada masyarakat modern, faktor kecepatan pengaksesan dan ketepatan isi informasi menjadi kunci utama dalam pemenuhan kebutuhan akan informasi (Salsabil, 2019). Berdasarkan riset yang dilakukan oleh Badan Pusat Statistik (BPS) pada kuartal I Tahun 2020 pertumbuhan industri kimia, farmasi, dan obat tradisional termasuk kosmetik tumbuh 5,59% (Rizaty, 2021). Berkat adanya internet, perkembangan teknologi telah memunculkan berbagai peluang baru dalam bisnis, salah satunya yaitu *e-commerce*. *E-Commerce (Electronic Commerce)* adalah segala aktivitas jual beli yang dilakukan melalui media elektronik. Meskipun terdapat berbagai macam media elektronik, kini *e-commerce* lebih sering terjadi melalui internet. Terdapat beberapa jenis *e-commerce* salah satunya yaitu *consumer to consumer (C2C)* yaitu transaksi barang atau jasa yang dilakukan dari konsumen kepada konsumen. C2C terbagi menjadi dua model yakni *marketplace* dan *classified*. Pada model *marketplace*, konsumen sebagai penyedia barang atau jasa membutuhkan sebuah *platform* sebagai wadah transaksi. Contoh *platform* C2C yang terkenal di

Indonesia adalah Tokopedia, Bukalapak, Shopee, Female Daily (*Beauty Studio*), dan sebagainya (Nayoan, 2021).

Female Daily Network merupakan sebuah situs media informasi yang menyediakan *review* atau ulasan dari berbagai macam produk *skincare*. Female Daily diakui sebagai *platform* kecantikan nomor satu di Indonesia. Dilansir dari *website* www.femaledaily.com, ada lebih dari 13.500 produk kecantikan dari 1.000 merek dan lebih dari 100.000 ulasan yang tersedia. Selain menyediakan *review* produk, *platform* Female Daily menyediakan artikel tips & trik, opini ahli, opini konsumen, tutorial kecantikan, diskusi, dan *workshop* kecantikan. Selain itu, Female Daily juga menyediakan *marketplace* penjualan *skincare* yang bernama *Beauty Studio*. *Review* produk oleh pengguna menjadi salah satu faktor yang paling berperan dalam memberikan informasi bagi pengunjung *website* Female Daily. Informasi mengenai pengalaman pengguna yang telah menggunakan suatu produk *skincare* serta *rating* yang diberikan dijadikan referensi oleh pengguna lain untuk memilih produk *skincare* yang akan digunakan.

Masa pandemi Covid-19 menimbulkan perubahan yang sangat signifikan bagi masyarakat salah satunya yaitu perubahan gaya hidup. Masyarakat menjadi lebih banyak melakukan aktivitas di dalam rumah. Selain itu, banyak tempat yang ditutup demi memutus rantai penyebaran virus covid-19. Hal ini membuat masyarakat secara terpaksa harus dapat melakukan hal-hal yang biasanya dilakukan di luar ruangan menjadi di dalam ruangan atau di rumah, salah satunya yaitu melakukan perawatan kulit. Disitulah peran *skincare* sangat dibutuhkan agar orang-orang memiliki kulit yang sehat. Namun, banyak orang merasa sulit dalam memilih produk *skincare* karena produk yang beredar di pasaran sangat banyak, selain itu juga karena iklan yang hanya memberikan sisi positifnya saja. Berdasarkan permasalahan tersebut, diperlukan solusi berupa model sistem rekomendasi dalam memilih produk *skincare* yang diperoleh dari *review* pengguna.

Sistem rekomendasi adalah suatu sistem yang memberikan rekomendasi pada pengguna untuk mendapatkan produk yang diinginkan. Contoh dari sistem rekomendasi yaitu rekomendasi film, musik, dan lain sebagainya. Terdapat berbagai metode dalam membuat sistem rekomendasi yaitu *collaborative-*

filtering, *content-based filtering*, *knowledge-based*, dan *hybrid filtering* (Girsang, 2020). Peneliti menggunakan pendekatan *content-based filtering* karena sistem rekomendasinya dibangun atas dasar kemiripan atribut dari produk yang telah dinilai oleh penggunanya. Langkah dalam analisis menggunakan metode *content-based filtering* dimulai dari perhitungan vektor dalam teks, pembobotan menggunakan algoritma TF-IDF (*term frequency-inverse document frequency*), dan *cosine similarity* (Badriyah, 2018). *Cosine similarity* digunakan untuk mengukur kedekatan antara dua vektor. Pada hasil perhitungan *cosine similarity*, hasil perhitungan tertinggi yang menjadi vektor yang terdekat dan vektor yang ingin dibandingkan. Pada kasus ini, nilai tertinggi adalah rekomendasi produk *skincare* yang akan diberikan kepada pengguna.

Berdasarkan permasalahan tersebut, peneliti akan melakukan penelitian tentang sistem rekomendasi produk *skincare* berdasarkan *review* pengguna *skincare* pada *website* Female Daily dengan pendekatan *content-based filtering*, dengan harapan penelitian ini dapat membantu pengguna *skincare* dalam memilih produk agar sesuai dengan spesifikasi yang dibutuhkan.

1.2. Rumusan Masalah

Berdasarkan penjelasan pada latar belakang, maka rumusan masalah yang dapat dikaji pada penelitian ini adalah sebagai berikut:

1. Bagaimana merekomendasikan suatu produk *skincare* berdasarkan *review* menggunakan *Content-Based Filtering*?
2. Bagaimana hasil rekomendasi produk *skincare* berdasarkan *review* yang diberikan oleh sistem rekomendasi menggunakan *Content-Based Filtering*?

1.3. Batasan Masalah

Agar pembahasan penulisan tepat mencapai sasaran, maka penulis membatasi masalah yang akan dibahas yaitu sebagai berikut:

1. Data yang digunakan dalam penelitian adalah data *review* produk *skincare* di *Femaledaily* yang diperoleh pada tanggal 6 Desember 2021 sampai dengan tanggal 30 Desember 2021.

2. Data *review* produk *skincare* mencakup 25 kategori dengan jumlah data sebanyak 20915.
3. Data diolah dengan menggunakan Bahasa Pemrograman *Python* dengan metode analisis yang digunakan yaitu sistem rekomendasi dengan menggunakan pendekatan *content-based filtering*. *Microsoft Excel* digunakan untuk menyimpan hasil *scraping* data yang dilakukan menggunakan *software ParseHub*.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ada, maka penelitian ini bertujuan sebagai berikut:

1. Mengetahui rekomendasi suatu produk *skincare* berdasarkan *review* menggunakan *Content-Based Filtering*
2. Mengetahui hasil rekomendasi produk *skincare* berdasarkan *review* yang diberikan oleh sistem rekomendasi menggunakan *Content-Based Filtering*

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Dapat membantu para pengguna dalam menentukan produk *skincare* yang akan digunakan sesuai dengan spesifikasi yang diinginkan.
2. Dapat membantu para pengguna dan perusahaan *skincare* dalam mengetahui beberapa produk yang mempunyai kemiripan terhadap suatu produk.
3. Memberikan pengetahuan bagaimana cara kerja sistem rekomendasi dengan metode *content-based filtering*.

BAB II

TINJAUAN PUSTAKA

Dalam penulisan penelitian ini tentunya membutuhkan acuan berupa penelitian terdahulu yang berguna untuk menambah pengetahuan dan wawasan bagi peneliti sehingga dapat menyempurnakan penelitian sebelumnya. Selain itu, penelitian terdahulu juga berguna bagi peneliti dalam mengkaji keterkaitan antara penelitian terdahulu dengan penelitian yang akan dilakukan dalam menghindari tindakan duplikasi. Berikut merupakan penelitian-penelitian terdahulu yang dijadikan sebagai acuan dalam penelitian:

Penelitian pertama yang dilakukan oleh (Putri, Muchayan, & Kamisutara, 2020) membahas mengenai sistem rekomendasi pemilihan pena yang dapat dijadikan alat bantu pemilihan pena terhadap calon pelanggan serta dapat meminimalisir terjadinya *miss stock*. Data yang dipergunakan terdiri dari 258 kode produk yang masing-masing mempunyai delapan kategori serta 33 kata kunci pembentuk sesuai dengan *product knowledge* perusahaan. Penelitian ini menggunakan metode *Content-Based Filtering* dan *Term Frequency Inverse Document Frequency* (TF-IDF) dari model *Information Retrieval* (IR). Hasil dari penelitian ini yaitu nilai bobot di perhitungan TF-IDF sebesar 13,854 ketika menampilkan rekomendasi produk terbaik pertama dan mempunyai keakuratan sebesar 96,5% dalam memberikan rekomendasi pena.

Penelitian kedua dari (Larasati & Henry, 2021) membahas mengenai sistem rekomendasi produk Emina *Cosmetics* sesuai dengan ketertarikan dan kebutuhan *customer*. Penelitian ini bertujuan untuk merancang sebuah aplikasi yang dapat memberikan rekomendasi kosmetik emina kepada customer berdasarkan produk yang dicari sebelumnya. Proses rekomendasi dilakukan dengan membandingkan kesamaan antara produk yang dicari *customer* dengan deskripsi produk. Penelitian ini menggunakan metode *content-based filtering* dimana metode tersebut digunakan untuk merekomendasikan produk berdasarkan ketersediaan konten atau deskripsi produk. Hasil dari penelitian ini yaitu produk dengan nilai similaritas tertinggi sebesar 0,7195.

Penelitian ketiga yang dilakukan oleh (Alkaff, Khatimi, & Eriady, 2020) membahas mengenai penerapan sistem rekomendasi menggunakan metode *Content Based Filtering* dalam memberikan rekomendasi buku berdasarkan kemiripan *item* yang dianalisis dari fitur yang terdapat pada *item* tersebut dengan *Weighted Tree Similarity*. Penelitian ini bertujuan untuk mempermudah pengunjung Perpustakaan Daerah Kalimantan Selatan untuk mendapatkan buku yang sesuai dengan keinginannya. Hasil dari penelitian tersebut yaitu nilai *precision* dari lima skenario pengujian sebesar 88%.

Penelitian keempat dari (Fiarni, Maharani, & Calista, 2019) mengenai sistem rekomendasi pada perusahaan pewarna tekstil PT. X untuk menyarankan produk baru atau alternatif kepada calon pelanggan berdasarkan karakteristik dan riwayat transaksi dari setiap pelanggan. Metode yang digunakan pada penelitian ini yaitu *cosine similarity* untuk menghitung skor kesamaan produk dan *content-based filtering* untuk menghitung skor model rekomendasi pelanggan. Hasil dari penelitian tersebut yaitu nilai *similarity* sebesar 93,47%.

Penelitian kelima yang dilakukan oleh (Putra & Santika, 2020) yang bertujuan untuk membuat sistem yang dapat menyajikan rekomendasi musik sesuai dengan preferensi *user* sehingga tingkat kenyamanan *user* akan meningkat. Penelitian ini menggunakan metode *content-based filtering* dengan menghasilkan tingkat kemiripan rata-rata hingga 0,6684 dan hasil *performance testing* dan *system testing* yang diperoleh menyatakan bahwa sistem rekomendasi dapat berjalan baik dengan rata-rata *response time* mencapai 3,5 detik.

Penelitian keenam dari (Safitri, Helilintar, & Wahyuniar, 2021) yang bertujuan untuk merancang sebuah aplikasi yang dapat memberikan rekomendasi produk *skincare* kepada konsumen berdasarkan produk yang disukai sebelumnya menggunakan metode *Content-Based Filtering* dan Algoritma Apriori dengan cara menghitung nilai kemiripan konten suatu *item* yang menghasilkan *rating* produk dan juga menghitung nilai minimum *support* serta nilai *minimum confidence* untuk menentukan aturan asosiasi suatu kombinasi *item*. Hasil dari penelitian ini yaitu nilai *rating* tertinggi sebesar 0,447 dan *confidence* sebesar 88,89%.

Tabel 2.1. Tabel Penelitian Sebelumnya

Tahun	Nama	Judul	Hasil Penelitian
2020	Mariani Widia Putri, Achmad Muchayan, Made Kamisutara	Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode <i>Content-Based Filtering</i> dan TF-IDF	Menghasilkan sistem rekomendasi pemilihan pena yang dapat dijadikan alat bantu dalam pemilihan pena oleh calon pelanggan serta dapat meminimalisir terjadinya <i>miss stock</i> . Sistem rekomendasi yang dibuat menghasilkan akurasi sebesar 96,5% sehingga dapat dikatakan sangat akurat dan sesuai dengan kebutuhan pelanggan.
2021	Fatoni Batari Agung Larasati, Herny Februariyanti	Sistem Rekomendasi <i>Product Emina Cosmetics</i> Dengan Menggunakan Metode <i>Content-Based Filtering</i>	Menghasilkan sistem rekomendasi produk <i>Emina Cosmetics</i> berdasarkan deskripsi produk dengan nilai similaritas tertinggi senilai 0,7195.
2020	Muhammad Alkaff, Husnul Khatimi, Andi Eriady	Sistem Rekomendasi Buku Menggunakan <i>Weighted Tree Similarity</i> dan <i>Content Based Filtering</i>	Menghasilkan sistem rekomendasi buku di Perpustakaan Daerah Provinsi Kalimantan Selatan dalam menemukan alternatif buku apabila buku yang dicari tidak ada. Berdasarkan hasil pengujian yang dilakukan pada 5 skenario pengujian dihasilkan nilai <i>precision</i> sebesar 88%.
2019	Cut Fiarni, Herastia Maharani, Nathania Calista	<i>Product Recommendation System Design Using Cosine Similarity and Content-Based Filtering Methods</i>	Menghasilkan sistem rekomendasi produk tekstil di PT. X dalam memilih produk yang sesuai dengan karakteristik calon pelanggan dan memberikan saran produk apabila produk yang diinginkan tidak tersedia. Sistem rekomendasi tersebut menghasilkan akurasi sebesar 93,47% sehingga dapat dikatakan bahwa algoritma sudah baik digunakan.

Tahun	Nama	Judul	Hasil Penelitian
2020	Aldy Istna Putra, Reva Ragam Santika	Implementasi <i>Machine Learning</i> dalam Penentuan Rekomendasi Musik dengan Metode <i>Content-Based Filtering</i>	Menghasilkan sistem rekomendasi musik yang sesuai dengan preferensi <i>user</i> dengan nilai tingkat kemiripan musik rata-rata hingga 0,6684 sehingga mampu menghasilkan rekomendasi yang sesuai dengan preferensi <i>user</i> , tetapi dengan data yang cukup banyak diperlukan algoritma yang lebih efektif.
2021	Dwi Ayu Nur Safitri, Risa Helilintar, Lilia Sinta Wahyuniar	Sistem Rekomendasi <i>Skincare</i> Menggunakan Metode <i>Content-Based Filtering</i> dan Algoritma Apriori	Menghasilkan sistem rekomendasi produk <i>skincare</i> sesuai dengan kebutuhan konsumen berdasarkan produk yang disukai sebelumnya. Sistem rekomendasi ini menghasilkan nilai <i>confidence</i> sebesar 88,89% sehingga dapat dikatakan bahwa metode tersebut mampu memberikan hasil rekomendasi yang cukup baik.

Perbedaan penelitian sebelumnya dengan penelitian ini adalah dalam hal objek yang digunakan kebanyakan berasal dari deskripsi produk, namun pada penelitian ini digunakan *review* produk yang diberikan oleh *user*. Selain itu, terdapat penelitian terdahulu yang menggunakan *weighted tree similarity* dalam pengukuran kemiripan teks, sedangkan pada penelitian ini menggunakan *cosine similarity*. Pada penelitian ini juga menggunakan tahap *pre-processing* yang lebih kompleks.

BAB III

LANDASAN TEORI

3.1. *E-Commerce*

E-Commerce berasal dari bahasa Inggris yang merupakan penggabungan dua buah kata, yaitu kata *E* yang merupakan singkatan dari *Electronic* dan kata *Commerce*. Menurut bahasa, *Electronic* adalah ilmu elektronik (muatan listrik), alat-alat elektronik, atau semua hal yang berhubungan dengan dunia elektronika dan teknologi. Sedangkan *Commerce* adalah perdagangan dan pergiagaan (Echols & Shadily, 2003). *E-Commerce* adalah proses pembelian, penjualan, atau pertukaran barang, jasa, dan informasi melalui jaringan komputer termasuk internet. Menurut (Kalakota & Whinston, 1997) istilah *e-commerce* dapat dilihat dari empat perspektif yang berbeda yaitu (Turban, King, Lee, Liang, & Turban, 2002):

1. Komunikasi

E-Commerce dapat berarti penyediaan barang, jasa, informasi, atau pembayaran melalui jaringan komputer atau alat elektronik lainnya.

2. Proses Bisnis

E-Commerce adalah aplikasi dari teknologi dengan tujuan mengotomatisasi transaksi bisnis dan langkah-langkah dalam melaksanakan pekerjaan (*workflow*).

3. Pelayanan

E-Commerce merupakan sebuah alat yang dapat memenuhi kebutuhan perusahaan, konsumen, dan manajemen agar meminimalisir biaya pelayanan, meningkatkan kualitas pelayanan kepada konsumen, dan meningkatkan kecepatan pelayanan konsumen.

4. *Online*

E-Commerce merupakan tempat dilaksanakannya proses jual beli produk atau informasi melalui internet dan layanan *online* lainnya.

3.2. Female Daily

Female Daily didirikan pada tahun 2005 oleh Hanifa Ambadar yang awal mulanya berasal dari *blog* pribadi dengan konten *fashion* dan kecantikan. Pada tahun 2006, Hanifa bertemu dengan Affi Assegaf yang mempunyai pekerjaan sebagai *blogger* dan karyawan di perusahaan kecantikan MAC *Cosmetics*. Setahun kemudian, Hanifa bekerja sama dengan Affi untuk sepakat dalam mengembangkan *blog* Female Daily. Pada saat itu, mereka mengamati potensi *blog* pribadi yang bisa menjadi bisnis atau media yang sangat berpengaruh seperti yang sudah terjadi di luar negeri seperti *blog Stylebubble* dan *Bryanboy*. Pada 2009, Female Daily mulai membuka kantor, membuat PT, dan merekrut karyawan menggunakan dana hasil iklan. *Blog* pertama yang diluncurkan yaitu *Mommies Daily* yang anggotanya masih menggunakan fasilitas forum *FashioneseDaily* untuk berdiskusi seputar kehamilan, dan lain-lain. Karena melihat banyak topik di forum yang di luar topik *fashion* dan *beauty*, maka Hanifa dan Affi memutuskan untuk memindahkan forum tersebut ke *domain* Female Daily pada 2010.



Gambar 3.1. Logo Female Daily

Female Daily kini telah menjadi jaringan media *digital* dan komunitas wanita terbesar di Indonesia yang berada di bawah badan hukum PT. *Daily Dinamika Kreasi*. Female Daily mempunyai visi: “*To be the company that best understand women, empower women, and connect women to each other*”. Dilansir dari *website Femaledaily.com*, Female Daily memiliki 2 juta *unique user* per bulan, lebih dari 355 ribu pengikut di *social media*, 7.5 juta *pageviews* per bulan, 18 ribu *forum threads*, 4.5 juta *forum threads*, dan 450 anggota baru per hari. Pada *website* Female Daily memuat berbagai pilihan yang berhubungan dengan kecantikan yaitu *review* produk, tips & trik, opini ahli & konsumen, tutorial kecantikan, diskusi, pelatihan kecantikan, dan lain-lain. Selain itu, Female Daily juga menyediakan *mall online* bernama *Beauty Studio* yang menyediakan

berbagai macam produk *skincare* dari berbagai merek dalam negeri maupun luar negeri. Dengan adanya Female Daily, orang dapat berbagi pengalaman tentang produk *skincare* bahkan dapat menjadi solusi dalam menentukan produk yang akan dibeli sesuai dengan keadaan kulit calon pelanggan. Per tahun 2022, terdapat lebih dari 13.500 produk kecantikan dari 1.000 merek dan lebih dari 100.000 ulasan. Female Daily pun hadir dalam bentuk aplikasi *mobile* untuk memudahkan pengguna ketika akan melakukan pembelian *online* atau memberikan *review* produk tanpa harus membuka *website*.

3.3. Skincare

Kata *skincare* berasal dari bahasa Inggris yang memiliki arti perawatan kulit. Secara luas, *skincare* dapat diartikan sebagai serangkaian perawatan kulit dari berbagai produk yang digunakan untuk menjaga kesehatan dan kecantikan kulit. Untuk memiliki kulit yang sehat diperlukan nutrisi dari dalam dan luar tubuh. Nutrisi dari dalam yaitu dari makanan dan minuman yang dikonsumsi sedangkan nutrisi dari luar yaitu berasal dari *skincare* (Na'imah, 2021). Saat ini, penggunaan *skincare* merupakan suatu kebiasaan yang tidak pernah terlewatkan bagi setiap orang baik laki – laki maupun perempuan. Hal ini dikarenakan *skincare* membawa banyak manfaat bagi kulit, diantaranya yaitu memberikan nutrisi untuk kulit, memperlambat penuaan, serta memperbaiki masalah kulit yang dialami seperti jerawat, bekas jerawat, bintik hitam, keriput, dan lainnya.

Dalam memilih *skincare* tentunya disesuaikan dengan jenis kulit dan kondisi atau masalah kulit yang dialami. Umumnya terdapat lima produk *skincare* dasar yang sering digunakan yaitu *facial wash*, toner, serum, pelembab, dan *sunscreen* (Ikhsania, 2020). *Facial wash* atau sabun pembersih wajah merupakan langkah pertama yang dilakukan sebelum menggunakan produk lain. Membersihkan wajah dapat membuka pori – pori, melembutkan kulit, dan membuat kulit bersih dari kotoran. Produk toner merupakan cairan yang mengandung bahan aktif untuk mengatasi masalah kulit tertentu. Sebagian besar toner membantu melembabkan kulit sehingga kulit dapat menyerap bahan aktif dalam produk lain dengan lebih baik. Namun, ada pula toner yang mampu menyeimbangkan pH kulit, mengencangkan pori-pori, dan mengatasi jerawat.

Serum merupakan *gel* bertekstur ringan yang mengandung sejumlah bahan aktif yang bertujuan untuk mengatasi masalah kulit yang lebih spesifik seperti jerawat, flek hitam, atau kulit kusam. Untuk menjaga sel – sel kulit agar terhindar dari kerusakan dan iritasi, diperlukan pelembab karena kulit hanya bisa menjalankan fungsinya dengan baik ketika kondisinya lembab. Produk yang paling penting digunakan pada pagi hari yaitu *sunscreen* atau tabir surya. Paparan sinar matahari dalam jangka panjang dapat menyebabkan kerusakan pada kulit mulai dari penuaan dini hingga kanker. Maka dari itu, diperlukan *sunscreen* agar dapat melindungi kulit dari penuaan akibat paparan sinar matahari sehingga memperlambat tanda – tanda penuaan (Perwitasari, 2021). Perlu diketahui bahwa tidak semua orang memiliki produk *skincare* yang sama, penggunaan *skincare* harus disesuaikan dengan jenis dan kebutuhan kulit. Dengan demikian, manfaat dari *skincare* yang digunakan bisa didapatkan secara optimal.

3.4. Text Mining

Text mining juga dikenal sebagai Teks Data Mining (TDM) atau *Knowledge Discovery in Text* (KDT) dapat didefinisikan sebagai penemuan informasi baru dan tidak diketahui sebelumnya oleh komputer, yang secara otomatis mengekstrak informasi dari sumber-sumber teks tak terstruktur yang berbeda. Hal terpenting dari proses ini adalah menggabungkan informasi yang berhasil diekstraksi dari berbagai sumber (Prasetyo, 2012). Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen sehingga sumber data yang digunakan dalam *text mining* berupa sekumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. *Text mining* mendukung proses *knowledge discovery* pada kumpulan dokumen yang berukuran besar.

Pada prinsipnya, *text mining* adalah bidang ilmu multidisipliner yang melibatkan *information retrieval* (IR), *text analysis*, *information extraction* (IE), *clustering*, *categorization*, *visualization*, *database technology*, *natural language processing* (NLP), *machine learning*, dan *data mining*. *Text mining* juga dapat disebut salah satu bentuk aplikasi kecerdasan buatan (*artificial intelligence*). *Text mining* mempunyai tujuan yang sama dengan *data mining* yaitu untuk

memperoleh informasi dan pengetahuan dari sekumpulan data yang besar yang berbentuk *database* (Setiaji, 2014). Perbedaan dari *text mining* dan *data mining* terletak pada sumber data yang digunakan. Pada *data mining*, pola-pola diekstrak dari basis data yang terstruktur, sedangkan pada *text mining*, pola-pola diekstrak dari data tekstual (*natural language*). Terdapat empat tahapan dalam *text mining*, yaitu pemrosesan awal terhadap teks (*text preprocessing*), transformasi teks (*text transformation*), pemilihan fitur (*feature selection*), dan penemuan pola (*pattern discovery*).

3.4.1 Vektorisasi Kata

Vektorisasi merupakan proses membuat vektor dengan nilai berupa angka-angka kemunculan *term* (kata unik) dalam dokumen. Sebelum melakukan vektorisasi harus dilakukan tokenisasi yaitu pemisahan teks menjadi kata-kata. Menurut (Aulia & Patriya, 2019), tokenisasi merupakan proses pemecahan dokumen menjadi komponen-komponen individual.

3.4.2 Text Pre-processing

Text preprocessing merupakan langkah yang digunakan untuk mempelajari suatu teks data dengan menemukan fitur-fitur yang mewakili setiap kata untuk fitur yang ada pada dokumen. Sebelum menentukan fitur-fitur yang mewakili, diperlukan tahap *pre-processing* yang dilakukan secara umum dalam *text mining* pada dokumen, yaitu *case folding*, *tokenizing*, *filtering*, *stemming*, *tagging*, dan *analyzing*.

a. Case Folding

Case folding merupakan proses mengubah semua huruf kapital dalam dokumen menjadi huruf kecil. Huruf yang diterima yaitu huruf “a” sampai dengan “z”. Karakter selain huruf (tanda baca dan angka) akan dihilangkan dan dianggap delimiter. Delimiter adalah urutan satu atau lebih karakter yang digunakan untuk menentukan batas pemisah (Raghavan & Schutze, 2009).

b. Tokenizing

Tokenizing atau yang bisa disebut juga dengan *parsing* adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Pada tahap ini, terjadi pemotongan kalimat menjadi sebuah kata dengan melakukan analisa terhadap sekumpulan data dengan memisahkan kata tersebut dan

menentukan struktur dari setiap kata. Pada tahap ini juga dilakukan proses *removing number*, *removing punctuation* seperti simbol dan tanda baca, serta *removing whitespace*.

c. *Filtering*

Filtering adalah tahap mengambil kata-kata penting dari hasil *tokenizing*. Algoritma yang digunakan yaitu *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting) yang merupakan kata-kata yang tidak deskriptif. Contoh *stopwords* adalah “dan”, “yang”, “di”, dan seterusnya.

d. *Stemming*

Stemming merupakan proses pencarian kata *root* atau kata dasar dari setiap kata hasil dari proses normalisasi. Pada tahap ini dilakukan proses pengambilan berbagai bentuk kata kedalam suatu representasi yang sama serta menghilangkan imbuhan (awalan dan akhiran). Tahap ini sering kali dipakai untuk teks bahasa Inggris dan sulit diterapkan pada teks bahasa Indonesia. Hal ini dikarenakan bahasa Indonesia tidak memiliki rumus bentuk baku yang permanen. Pada teks bahasa Inggris, proses yang diperlukan hanya proses menghilangkan sufiks. Namun pada teks bahasa Indonesia harus menghilangkan sufiks, prefiks dan konflik.

3.5. TF-IDF

TF-IDF (*Term Frequency Inverse Document Frequency*) merupakan proses pembobotan data berupa kata menjadi numerik. Metode TF-IDF digunakan untuk menentukan seberapa jauh keterhubungan kata (*term*) terhadap dokumen dengan memberikan bobot setiap kata. Dengan menggunakan frekuensi kemunculan kata (TF) merupakan petunjuk sejauh mana kata tersebut mewakili isi sebuah dokumen. Jadi, semakin besar kemunculan kata dalam sebuah dokumen maka akan memberikan nilai kemiripan yang semakin besar.

Dalam perhitungan bobot menggunakan TF-IDF terdapat beberapa tahapan yaitu sebagai berikut (Luqyana, Cholissudin, & Perdana, 2018):

- a. *Term Frequency* (TF) merupakan perhitungan awal dengan menghitung pembobotan setiap kata dengan melihat jumlah kemunculan *term* atau kata

pada sebuah dokumen dengan masing – masing bobot setiap kata berbeda bernilai 1 dengan rumus sebagai berikut:

$$TF_{i,j} = \frac{n_{i,j}}{\sum n_{i,j}} \quad (3.1)$$

Keterangan :

$tf_{i,j}$ = *term frequency* (kata ke-*i* dalam dokumen ke-*j*)

$n_{i,j}$ = banyaknya kata ke-*i* dalam dokumen ke-*j*

- b. *Inverse Document Frequency* (IDF) yaitu menghitung *term* pada setiap dokumen. Sebelum mencari nilai IDF maka terlebih dahulu mencari nilai DF yang merupakan perhitungan jumlah frekuensi dokumen yang mengandung kata yang sama. IDF digunakan untuk mengurangi bobot suatu *term* jika banyak tersebar atau lebih banyak muncul di seluruh dokumen sehingga dianggap sebagai *term* umum yang dinilai tidak penting.

$$IDF_j = \ln \frac{D}{df_j} \quad (3.2)$$

Keterangan :

IDF_j = *Inverse Document Frequency* (kata ke-*j*)

D = Jumlah dokumen

df_j = Jumlah frekuensi dokumen yang mengandung *term* (kata ke- *j*)

- c. Perhitungan bobot TF-IDF

$$W_{ij} = TF_{ij} \times IDF_j \quad (3.3)$$

$$W_{ij} = TF_{ij} \times \left(\ln \left(\frac{D}{df_j} \right) + 1 \right)$$

Keterangan :

$W_{i,j}$ = Bobot TF-IDF (kata ke-*i* dalam dokumen ke-*j*)

$TF_{i,j}$ = *Term frequency* (kata ke-*i* dalam dokumen ke-*i*)

D = Jumlah dokumen

df_j = Jumlah frekuensi dokumen yang mengandung *term* (kata ke- *j*)

3.6. *Cosine Similarity*

Cosine Similarity adalah ukuran kesamaan yang lebih umum digunakan dalam *information retrieval* dan merupakan ukuran sudut antara vektor dokumen D_a (titik (ax,bx)) dan D_b (titik (ay,by)). Tiap vektor tersebut mempresentasikan

setiap kata dalam setiap dokumen (teks) yang dibandingkan dan membentuk sebuah segitiga, sehingga dapat diterapkan hukum kosinus (Imbar, Adelia, Ayub, & Rehatta, 2014). *Cosine similarity* juga merupakan algoritma di dalam *text mining* yang berfungsi untuk mengklasifikasikan suatu dokumen atau teks dengan konsep normalisasi panjang vektor dengan membandingkan antara dokumen A dan B. Rumus yang digunakan untuk menghitung *cosine similarity* yaitu sebagai berikut (Yusuf & Cherid, 2020):

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

Keterangan :

A = Vektor A, yang akan dibandingkan kemiripannya

B = Vektor B, yang akan dibandingkan kemiripannya

$A \cdot B$ = *Dot product* antara vektor A dan vektor B

$|A|$ = Panjang vektor A

$|B|$ = Panjang vektor B

$|A||B|$ = *Cross product* antara $|A|$ dan $|B|$

Nilai *cosine similarity* berada pada rentang 0-1. Jika nilai *cosine similarity* semakin mendekati 0, maka semakin tidak mirip antara teks A dan B. Sebaliknya, apabila nilai *cosine similarity* semakin mendekati 1, maka semakin mirip antara teks A dan B (Samuel, Natan, Fitria, & Syafiqoh, 2018)

3.7. Sistem Rekomendasi

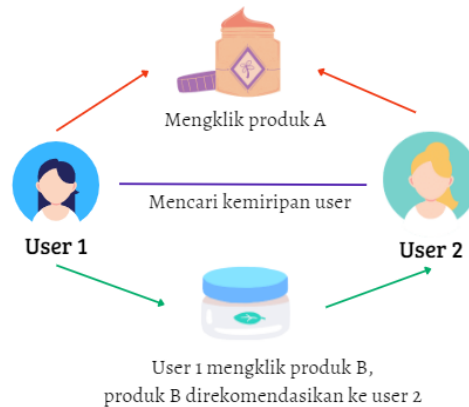
Sistem rekomendasi merupakan suatu sistem yang dapat memberikan informasi dan rekomendasi yang membantu *user* (pengguna) dalam membuat keputusan berdasarkan data yang telah ada sebelumnya (Ferio, Intan, & Rostianingsih, 2018). Menurut (Sebastia, Garcia, Onaindia, & Guzman, 2009), sistem rekomendasi merupakan sebuah alat personalisasi yang menyediakan sebuah informasi berupa daftar *item* yang sesuai dengan keinginan masing-masing pengguna. Sistem rekomendasi menyimpulkan preferensi pengguna dengan menganalisa ketersediaan data pengguna, informasi tentang pengguna dan lingkungannya. Oleh karena itu, sistem rekomendasi akan menawarkan kemungkinan dari penyaringan informasi personal sehingga hanya informasi yang

sesuai dengan kebutuhan dan preferensi pengguna yang akan ditampilkan di sistem dengan menggunakan sebuah teknik atau model rekomendasi.

Pada sistem rekomendasi, terdapat beberapa *item* yang akan disaring untuk direkomendasikan kepada pengguna berdasarkan profil pengguna, skala *rating*, *item* yang sudah dibeli, dan lain-lain sehingga menghasilkan beberapa *item* yang direkomendasikan kepada pengguna. Saat ini banyak media sosial yang sudah menerapkan sistem rekomendasi, seperti *twitter*, *facebook*, *youtube*, *netflix*, dan lain-lain. Terdapat beragam metode yang digunakan untuk membuat sistem rekomendasi, diantaranya yaitu *collaborative filtering* (rekomendasi berdasarkan ulasan dari pengguna lain atau diri sendiri), *content-based filtering* (rekomendasi berdasarkan kemiripan atribut dari *item* yang disukai), *knowledge-based* (rekomendasi berdasarkan kondisi nilai atribut yang telah ditentukan oleh pengguna), dan *hybrid filtering* (kombinasi dari metode rekomendasi yang lain untuk menghasilkan rekomendasi yang lebih akurat).

3.7.1 Collaborative Filtering

Collaborative filtering merupakan proses penyaringan atau pengevaluasian *item* menggunakan opini orang lain (Schafer, 2007). *Collaborative filtering* melakukan penyaringan data berdasarkan kemiripan karakteristik konsumen sehingga mampu memberikan informasi yang baru kepada konsumen karena sistem memberikan informasi berdasarkan pola satu kelompok konsumen yang hampir sama. Kelebihan dari metode *collaborative filtering* yaitu sistem rekomendasi tidak akan berhenti bekerja walaupun konten sulit dianalisa. Kekurangan dari metode ini yaitu wajib memiliki parameter *rating* sehingga jika ada *item* baru, sistem tidak akan merekomendasikan *item* tersebut (Wijaya & Alfian, 2018). Adapun ilustrasi proses dari metode *collaborative filtering* unruk merekomendasikan suatu *item* adalah sebagai berikut:



Gambar 3.2. Ilustrasi proses *collaborative filtering*

Berdasarkan **Gambar 3.2.** jika *user 1* dan *user 2* mengklik produk A, maka sistem akan mencari kemiripan antar *user*, didapatkan bahwa *user 2* mirip dengan *user 1*, maka disaat *user 1* mengklik produk B, produk B akan direkomendasikan ke *user 2*.

Menurut (Konguasa, 2018), metode *collaborative filtering* dibagi menjadi dua, yaitu *user-based collaborative filtering* dan *item-based collaborative filtering*. Metode *user-based collaborative filtering* atau bisa disebut dengan *nearest-neighbour* atau *memory-based filtering* memberikan rekomendasi kepada *user item-item* yang disukai atau dirating oleh *user* lain yang memiliki banyak kemiripan dengannya. Kelebihan dari pendekatan *user-based collaborative filtering* adalah dapat menghasilkan rekomendasi yang berkualitas baik. Sedangkan kelemahannya yaitu kompleksitas perhitungan akan semakin bertambah seiring dengan bertambahnya pengguna sistem, semakin banyak pengguna yang menggunakan sistem maka proses perekomendasi akan semakin lama. Metode *item-based collaborative filtering* memberikan rekomendasi berdasarkan kemiripan antar *item*. Metode ini didasari atas adanya kesamaan antara pemberian *rating* terhadap suatu *item* dengan *item* yang pernah dirating pengguna lain. *Item* yang telah di *rating* oleh *user* akan menjadi acuan untuk mencari sejumlah *item* lainnya yang berkorelasi dengan *item* yang telah di *rating user*. Metode ini sering disebut *model-based*. Metode ini dapat dijadikan solusi untuk beberapa permasalahan pada *user-based* yaitu masalah keterbatasan atau *sparsity* dan skalabilitas serta masalah waktu dan memori.

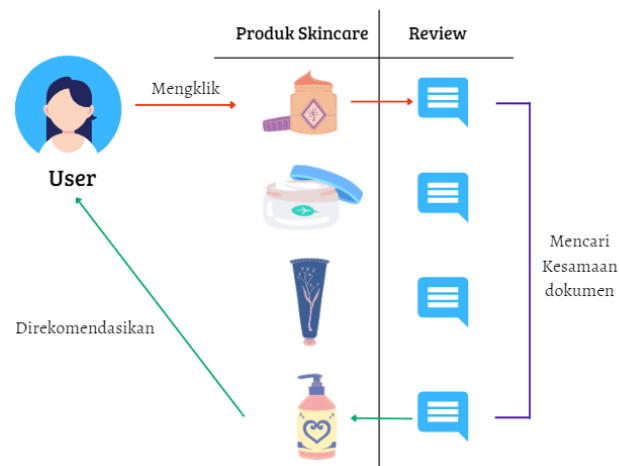
Terdapat dua bentuk *output* yang dihasilkan dari proses *collaborative filtering* yaitu prediksi dan rekomendasi (Masruri & Mahmudy, 2007). Prediksi merupakan nilai numerik dimana $P_{a,j}$ adalah nilai prediksi *rating item* j yang mungkin disukai oleh pengguna aktif (U_a). Nilai prediksi ini digunakan dengan skala yang sama dengan nilai yang disediakan. Rekomendasi adalah daftar N *item* yang mungkin akan disukai oleh pengguna U_a . Daftar *item* yang direkomendasikan biasanya terdiri dari *item* yang belum pernah dibeli atau di *rating* oleh pengguna aktif. *Output* dari algoritma ini juga dikenal sebagai *Top-N Recommendation*.

3.7.2 Content-Based Filtering

Metode *Content-based filtering* menghasilkan suatu rekomendasi berdasarkan hasil analisa kemiripan *item* yang telah dinilai oleh pengguna. *Content-based filtering* membentuk profil penggunanya berdasarkan atribut pembentuk suatu *item*. Pada suatu dokumen, atribut pembentuknya adalah kata-kata yang terdapat pada dokumen tersebut. Metode ini bersifat *user independence* yaitu tidak bergantung pada situasi apakah *item* tersebut merupakan *item* baru (belum pernah dipilih oleh pengguna manapun) maupun bukan *item* baru. Kelemahan dari metode *content-based filtering* adalah terbatasnya rekomendasi hanya pada *item-item* yang mirip sehingga kecil kesempatan untuk mendapatkan *item* yang tidak terduga (Mondi, Wijayanto, & Winarno, 2019).

Metode *content-based filtering* memungkinkan pengguna untuk membuat prediksi berdasarkan analisis *item* atau metadata yang saling terkait. Teknik ini memungkinkan kita untuk mengetahui semua *item* yang digunakan oleh pengguna di masa lampau dan dapat dijadikan sebagai landasan untuk menentukan seberapa mirip preferensi mereka dengan *item* yang ada. Berdasarkan penjelasan diatas, dapat disimpulkan bahwa metode ini merupakan salah satu teknik yang digunakan untuk menganalisis *item* serta melakukan ekstraksi fitur yang dapat menggambarannya. Setelah fitur diekstraksi, secara otomatis sistem akan melakukan pengecekan *item* yang mirip dengan preferensi pengguna. Kualitas data yang tersedia merupakan faktor penentu dalam kualitas hasil yang dihasilkan oleh sistem, karena teknik ini bergantung pada analisis konten (Girsang, 2020).

Adapun ilustrasi proses dari metode *content-based filtering* unruk merekomendasikan suatu *item* adalah sebagai berikut:



Gambar 3.3. Ilustrasi Proses *Content-Based Filtering*

Berdasarkan **Gambar 3.3.** jika *user* meng-klik produk A, maka sistem akan langsung mengecek dokumen A, dalam hal ini yaitu *review* produk A. selanjutnya sistem mencari dokumen yang paling mirip dengan dokumen A, kemudian didapatkan produk D yang paling mirip dengan dokumen A. Maka sistem akan merekomendasikan produk D kepada *user*.

Metode rekomendasi berbasis konten biasanya didasarkan pada gagasan bahwa sistem harus merekomendasikan *item* baru tergantung pada kesamaan atau ketidaksamaan mereka dengan preferensi pribadi pengguna. Sebagai contoh, konsep kesamaan adalah salah satu bagian mendasar dari algoritma yang diperlukan untuk mengatur dan merekomendasikan film. Untuk contoh kasus dalam pemilihan produk *skincare*, kesamaan dalam konten *review* produk dapat menjadi pendekatan yang layak untuk menyediakan rekomendasi produk kepada pengguna. Contoh konten dalam produk *skincare* misalnya tekstur, *brand*, kegunaan, efek, dan lain sebagainya.

Konsep *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) digunakan dalam pencarian informasi sistem dan juga mekanisme *content-based filtering* untuk menentukan kepentingan relatif dari sebuah dokumen, artikel, *item* berita, dan lain-lain. Setelah menghitung skor TF-IDF, peneliti menentukan *item* apa saja yang lebih dekat dengan *item* yang dipilih dengan menggunakan metode *cosine similarity*.

BAB IV

METODOLOGI PENELITIAN

4.1. Populasi dan Sampel Data

Populasi dalam penelitian ini yaitu seluruh pengguna produk *skincare* dan ulasan para pengguna produk di *website* <https://femaledaily.com/>. Sampel yang digunakan adalah data empat kategori *skincare* yaitu *moisturizer*, *treatment*, *mask*, dan *cleanser* dengan total sub-kategori (*face wash*, *oil*, *toner*, dll) sebanyak 22. Data diambil pada tanggal 6 Desember 2021 – 30 Desember 2021 dengan jumlah keseluruhan 26915 data.

4.2. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini sebanyak enam variabel. Berikut adalah definisi operasional variabel yang ditampilkan pada **Tabel 4.1**:

Tabel 4.1. Definisi Operasional Variabel

No.	Variabel	Definisi
1	<i>Category</i>	Pengelompokan produk <i>skincare</i> berdasarkan jenis <i>skincare</i> yang umum digunakan
2	<i>Sub Category</i>	Pengelompokan produk <i>skincare</i> yang lebih luas dari variabel <i>category</i>
3	<i>Brand</i>	Identitas produk yang dikeluarkan oleh suatu perusahaan
4	<i>Product Name</i>	Identitas yang digunakan untuk mengenali suatu produk
5	<i>Rating</i>	Penilaian yang diberikan oleh pengguna terhadap suatu produk
6	<i>Review</i>	Informasi mengenai suatu produk yang diberikan oleh pengguna

4.3. Metode Pengambilan Data

Pengambilan data dilakukan melalui *website* Femaledaily dengan cara *scraping* menggunakan aplikasi *ParseHub*. Data diambil pada tanggal 6

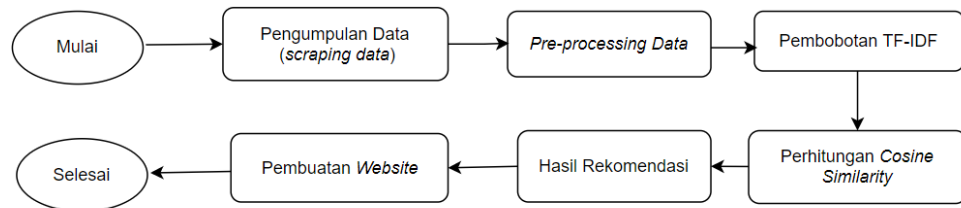
Desember 2021 – 30 Desember 2021. Data yang telah di *scrapping* disimpan di *Microsoft Excel* 2019. Data memuat empat kategori *skincare*, 22 sub-kategori *skincare*, dan sejumlah produk pada tiap sub-kategori dengan jumlah keseluruhan 26915 produk.

4.4. Metode Analisis Data

Metode analisis yang digunakan dalam penelitian ini yaitu metode *Content-Based Filtering* dengan menggunakan konsep perhitungan vektor dalam teks yang kemudian dilakukan pembobotan TF-IDF, dan selanjutnya dicari nilai kemiripan dokumen pembanding dengan dokumen uji menggunakan metode *Cosine Similarity* untuk menghasilkan rekomendasi produk yang mirip dengan produk yang diinginkan berdasarkan konten produk yaitu *brand* dan *review*.

4.5. Diagram Alir Penelitian

Berikut ini merupakan tahapan yang dilakukan pada penelitian ini dan digambarkan melalui diagram alir pada **Gambar 4.1**.



Gambar 4.1. *Flowchart* penelitian

Langkah – Langkah dari penelitian ini adalah:

1. Pengumpulan Data

Pengumpulan data dilakukan dengan cara *scrapping* pada *website* *FemaleDaily* menggunakan *software ParseHub*. Data yang diambil yaitu *rating* dan *review* produk.

2. *Pre-processing* Data

Pada tahap *pre-processing* data dilakukan *cleaning* data yaitu menghilangkan data-data yang tidak diperlukan. Pada tahap ini terdapat beberapa tahap yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*.

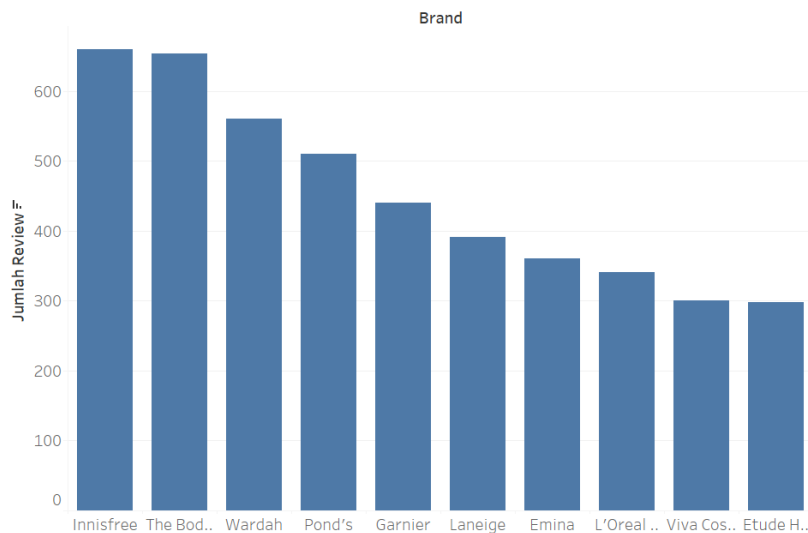
3. Pembobotan *Term Frequency-Invers Document Frequency* (TF-IDF)
Pembobotan *Term Frequency-Invers Document Frequency* (TF-IDF) dilakukan untuk menentukan seberapa jauh keterhubungan kata (*term*) terhadap dokumen dengan memberikan bobot setiap kata.
4. Perhitungan *Cosine Similarity*
Perhitungan *cosine similarity* digunakan untuk mencari nilai kemiripan antara *review* produk satu dengan yang lainnya dalam menghasilkan rekomendasi yang sesuai dengan spesifikasi.
5. Hasil Rekomendasi
Hasil rekomendasi produk memuat lima produk teratas yang direkomendasikan berdasarkan *review* pengguna.
6. Pembuatan *Website*
Pembuatan *website* digunakan agar pengguna dapat memperoleh hasil rekomendasi dengan lebih praktis.

BAB V

HASIL DAN PEMBAHASAN

5.1. Analisis Deskriptif

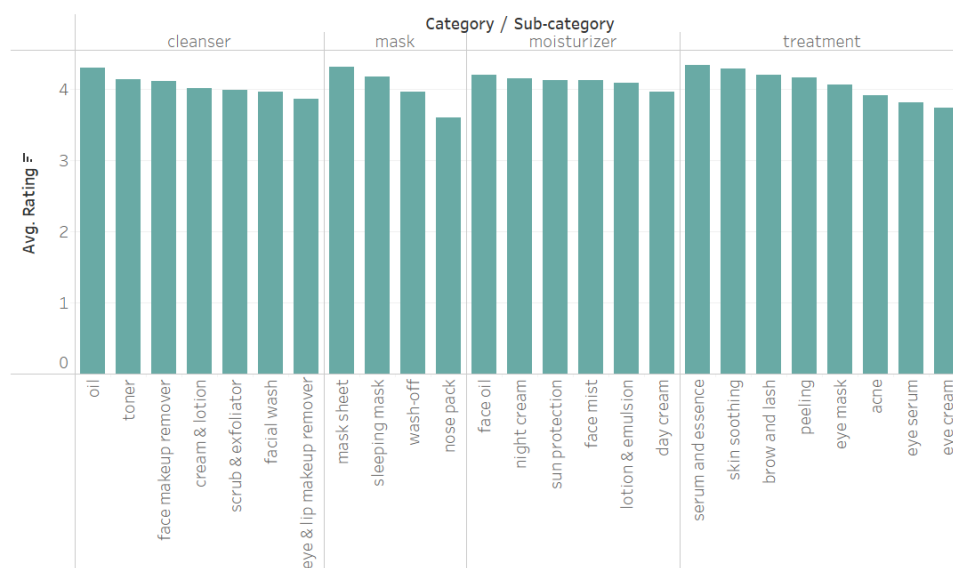
Analisis deskriptif digunakan untuk melihat gambaran data secara umum yang diperoleh dari hasil *scrapping review* produk *skincare* di Femaledaily. Berdasarkan **Gambar 5.1** dapat dilihat 10 *brand* atau merk *skincare* yang paling banyak direview oleh pengguna Femaledaily. Sumbu x (horizontal) menjelaskan nama *brand skincare* sedangkan sumbu y (vertikal) menjelaskan jumlah *review* pengguna *skincare* di Femaledaily. Dari total 502 *brand skincare* yang terdapat pada data, jumlah *review* terbanyak dimiliki oleh *brand* Innisfree yaitu sebanyak 660 *review* kemudian diikuti dengan *brand* The Body Shop dengan jumlah *review* 654, dan seterusnya. Banyaknya *review* yang dimiliki suatu *brand* dapat menandakan popularitas yang dimiliki *brand* tersebut cukup tinggi. Selain itu, dapat disebabkan oleh banyaknya jenis produk yang dikeluarkan oleh *brand* tersebut. Dari 10 *brand* dengan *review* terbanyak, hanya tiga *brand* yang berasal dari Indonesia yaitu Wardah, Emina, dan Viva Cosmetics. Hal ini menunjukkan banyaknya tren kecantikan yang berasal dari luar negeri.



Gambar 5.1. 10 *Brand* dengan frekuensi terbanyak

Pada **Gambar 5.2** terdapat jumlah rata-rata *rating* produk *skincare* berdasarkan kategori dan sub-kategori *skincare*. Garis bawah sumbu x (horizontal) menjelaskan sub-kategori *skincare* sedangkan garis atas sumbu x

(horizontal) menjelaskan kategori *skincare*. Kemudian sumbu y (vertikal) menjelaskan rata-rata *rating* produk *skincare* di Femaledaily. Rata-rata *rating* tertinggi dimiliki oleh produk dengan sub-kategori *oil*, *mask sheet*, dan *serum & essence* dengan nilai 4. Sedangkan rata-rata *rating* terendah yaitu produk *nose pack* dengan nilai lebih dari 3. Dapat dilihat bahwa rata-rata *rating* produk *skincare* cukup baik, yaitu lebih dari 3 dalam rentang nilai 0-4. Hal ini membuktikan bahwa kebanyakan produk *skincare* yang berada di pasaran saat ini mempunyai kualitas cukup baik dan tentunya sudah melalui penelitian tertentu sehingga dapat disukai oleh banyak orang.



Gambar 5.2. Rata-rata *rating* produk *skincare*

5.2. *Pre-processing Data*

Pre-processing data diperlukan untuk merapikan data agar terstruktur dengan cara menghilangkan beberapa bagian data yang tidak penting atau tidak diperlukan sehingga data yang didapat adalah data yang berkualitas dan dapat menghasilkan hasil analisis yang lebih baik serta akurat. Pada penelitian ini, peneliti menggunakan variabel *review* untuk dilakukan *pre-processing*. Terdapat beberapa langkah dalam *pre-processing* yaitu *case folding*, *remove punctuation*, *stopword removal*, *lemmatization*, dan *tokenizing*. Peneliti menggunakan variabel *review* pada dokumen ke-27 sebagai contoh untuk tahapan *pre-processing*.

5.2.1 Case Folding

Pada tahap *case folding* dilakukan perubahan huruf kapital menjadi huruf kecil. Hal ini bertujuan untuk menyeragamkan karakter pada data. Karakter yang dideteksi oleh sistem yaitu huruf a-z. Karakter selain huruf seperti tanda baca dan angka akan dihilangkan. Berikut ini adalah contoh dokumen hasil tahapan *case folding* yang dapat dilihat pada **Tabel 5.1**:

Tabel 5.1. Contoh dokumen pada tahap *case folding*

Sebelum	Sesudah
Recomended banget, aku pakai ini awalnya karena mau coba exfoliasi kulit tapi karena belum pernah aku pake ini dulu, bener" ngaruh, hidungku jadi haluss gitu ngurangi komedo di hidungku trus muka aku jadi lebih halus enggak berminyak juga dan gak bikin kering di muka. Selama aku pake ini mukaku juga gak muncul jerawat berlebih dan muka ku baik" aja malah tambah baik. kalo aku pake ini biasanya mukanya halus gitu komedonya berkurang juga, padahal sebelumnya di daerah hidung dan dekat mata tu keliatan banget komedonya dan dipegang kaya bruntus" gitu, trusmukaku lebih cerah kalo habis cuci muka, gak bikin berminyak tapi juga gak bikin kering, pokoknya bagus banget. Anjurannya dipakai 2 hari sekali di malam hari tapi juga boleh dipakai sehari sekali di malem hari juga dan biasanya aku pake di malem hari, kadang sehari sekali kadang 2 hari sekali.	recomended banget, aku pakai ini awalnya karena mau coba exfoliasi kulit tapi karena belum pernah aku pake ini dulu, bener" ngaruh, hidungku jadi haluss gitu ngurangi komedo di hidungku trus muka aku jadi lebih halus enggak berminyak juga dan gak bikin kering di muka. selama aku pake ini mukaku juga gak muncul jerawat berlebih dan muka ku baik" aja malah tambah baik. kalo aku pake ini biasanya mukanya halus gitu komedonya berkurang juga, padahal sebelumnya di daerah hidung dan dekat mata tu keliatan banget komedonya dan dipegang kaya bruntus" gitu, trusmukaku lebih cerah kalo habis cuci muka, gak bikin berminyak tapi juga gak bikin kering, pokoknya bagus banget. anjurannya dipakai hari sekali di malam hari tapi juga boleh dipakai sehari sekali di malem hari juga dan biasanya aku pake di malem hari, kadang sehari sekali kadang hari sekali.

5.2.2 Tokenizing

Pada tahap *tokenizing* dilakukan proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token atau dengan kata lain pemecahan kalimat menjadi kata untuk kemudian di analisa. Pada tahap ini juga dilakukan proses *remove whitespace* yaitu penghapusan spasi dan proses *remove punctuation* yaitu penghapusan simbol ataupun tanda baca seperti titik, koma, tanda tanya, tanda seru, dan lain sebagainya dikarenakan tanda baca tidak mempunyai arti/makna dalam analisis teks. Berikut ini adalah contoh dokumen hasil tahapan *tokenizing* yang dapat dilihat pada **Tabel 5.2**:

Tabel 5.2. Contoh dokumen pada tahap *tokenizing*

Sebelum	Sesudah
recomended banget, aku pakai ini awalnya karena mau coba exfoliasi kulit tapi karena belum pernah aku pake ini dulu, bener" ngaruh, hidungku jadi haluss gitu ngurangi komedo di hidungku trus muka aku jadi lebih halus enggak berminyak juga dan gak bikin kering di muka. selama aku pake ini mukaku juga gak muncul jerawat berlebih dan muka ku baik" aja malah tambah baik. kalo aku pake ini biasanya mukanya halus gitu komedonya berkurang juga, padahal sebelumnya di daerah hidung dan deket mata tu keliatan banget komedonya dan dipegang kaya bruntus" gitu, trusmukaku lebih cerahan kalo habis cuci muka, gak bikin berminyak tapi juga gak bikin kering, pokoknya bagus banget. anjurannya dipakai hari sekali di malam hari tapi juga boleh dipakai sehari sekali	'recomended', 'banget', 'aku', 'pakai', 'ini', 'awalnya', 'karena', 'mau', 'coba', 'exfoliasi', 'kulit', 'tapi', 'karena', 'belum', 'pernah', 'aku', 'pake', 'ini', 'dulu', 'bener', 'ngaruh', 'hidungku', 'jadi', 'haluss', 'gitu', 'ngurangi', 'komedo', 'di', 'hidungku', 'trus', 'muka', 'aku', 'jadi', 'lebih', 'halus', 'enggak', 'berminyak', 'juga', 'dan', 'gak', 'bikin', 'kering', 'di', 'muka', 'selama', 'aku', 'pake', 'ini', 'mukaku', 'juga', 'gak', 'muncul', 'jerawat', 'berlebih', 'dan', 'muka', 'ku', 'baik', 'aja', 'malah', 'tambah', 'baik', 'kalo', 'aku', 'pake', 'ini', 'biasanya', 'mukanya', 'halus', 'gitu', 'komedonya', 'berkurang', 'juga', 'padahal', 'sebelumnya', 'di', 'daerah', 'hidung', 'dan', 'deket', 'mata', 'tu', 'keliatan', 'banget', 'komedonya',

<p>di malem hari juga dan biasanya aku pake di malem hari, kadang sehari sekali kadang hari sekali.</p>	<p>‘dan’, ‘dipegang’, ‘kaya’, ‘bruntus’, ‘gitu’, ‘trusmukaku’, ‘lebih’, ‘cerahan’, ‘kalo’, ‘habis’, ‘cuci’, ‘muka’, ‘gak’, ‘bikin’, ‘berminyak’, ‘tapi’, ‘juga’, ‘gak’, ‘bikin’, ‘kering’, ‘pokoknya’, ‘bagus’, ‘bangett’, ‘anjurannya’, ‘dipakai’, ‘hari’, ‘sekali’, ‘di’, ‘malam’, ‘hari’, ‘tapi’, ‘juga’, ‘boleh’, ‘dipakai’, ‘sehari’, ‘sekali’, ‘di’, ‘malem’, ‘hari’, ‘juga’, ‘dan’, ‘biasanya’, ‘aku’, ‘pake’, ‘di’, ‘malem’, ‘hari’, ‘kadang’, ‘sehari’, ‘sekali’, ‘kadang’, ‘hari’, ‘sekali’</p>
---	---

5.2.3 Filtering

Pada tahap *filtering* dilakukan proses penghapusan kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Hal ini bisa disebut dengan *stopword removal*. Contoh *stopword* dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dan sebagainya. Hal ini bertujuan untuk fokus pada kata-kata yang penting untuk dilakukan analisis. Berikut ini adalah contoh dokumen hasil tahapan *filtering* yang dapat dilihat pada **Tabel 5.3**:

Tabel 5.3. Contoh dokumen pada tahap *filtering*

Sebelum	Sesudah
<p>‘recomended’, ‘bangett’, ‘aku’, ‘pakai’, ‘ini’, ‘awalnya’, ‘karena’, ‘mau’, ‘coba’, ‘exfoliasi’, ‘kulit’, ‘tapi’, ‘karena’, ‘belum’, ‘pernah’, ‘aku’, ‘pake’, ‘ini’, ‘dulu’, ‘bener’, ‘ngaruh’, ‘hidungku’, ‘jadi’, ‘haluss’, ‘gitu’, ‘ngurangi’, ‘komedo’, ‘di’, ‘hidungku’, ‘trus’, ‘muka’, ‘aku’, ‘jadi’, ‘lebih’, ‘halus’, ‘enggak’, ‘berminyak’, ‘juga’, ‘dan’, ‘gak’, ‘bikin’, ‘kering’, ‘di’, ‘muka’,</p>	<p>‘recomended’, ‘pakai’, ‘coba’, ‘exfoliasi’, ‘kulit’, ‘pake’, ‘ngaruh’, ‘hidungku’, ‘haluss’, ‘ngurangi’, ‘komedo’, ‘hidungku’, ‘muka’, ‘halus’, ‘berminyak’, ‘kering’, ‘muka’, ‘pake’, ‘mukaku’, ‘muncul’, ‘jerawat’, ‘berlebih’, ‘muka’, ‘pake’, ‘mukanya’, ‘halus’, ‘komedonya’, ‘berkurang’, ‘daerah’, ‘hidung’, ‘mata’, ‘keliatan’, ‘komedonya’, ‘dipegang’, ‘bruntus’,</p>

'selama', 'aku', 'pake', 'ini', 'mukaku', 'juga', 'gak', 'muncul', 'jerawat', 'berlebih', 'dan', 'muka', 'ku', 'baik', 'aja', 'malah', 'tambah', 'baik', 'kalo', 'aku', 'pake', 'ini', 'biasanya', 'mukanya', 'halus', 'gitu', 'komedonya', 'berkurang', 'juga', 'padahal', 'sebelumnya', 'di', 'daerah', 'hidung', 'dan', 'deket', 'mata', 'tu', 'keliatan', 'banget', 'komedonya', 'dan', 'dipegang', 'kaya', 'bruntus', 'gitu', 'trusmukaku', 'lebih', 'cerahan', 'kalo', 'habis', 'cuci', 'muka', 'gak', 'bikin', 'berminyak', 'tapi', 'juga', 'gak', 'bikin', 'kering', 'pokoknya', 'bagus', 'bangett', 'anjurannya', 'dipakai', 'hari', 'sekali', 'di', 'malam', 'hari', 'tapi', 'juga', 'boleh', 'dipakai', 'sehari', 'sekali', 'di', 'malem', 'hari', 'juga', 'dan', 'biasanya', 'aku', 'pake', 'di', 'malem', 'hari', 'kadang', 'sehari', 'sekali', 'kadang', 'hari', 'sekali'	'mukaku', 'cerahan', 'habis', 'cuci', 'muka', 'berminyak', 'kering', 'pokoknya', 'bagus', 'bangett', 'anjurannya', 'dipakai', 'malam', 'dipakai', 'malem', 'pake', 'malem'
--	--

5.2.4 *Stemming*

Pada tahap terakhir yaitu *stemming* dilakukan proses pencarian kata *root* atau kata dasar dari setiap kata hasil dari proses normalisasi. Pada tahap ini dilakukan proses pengambilan berbagai bentuk kata kedalam suatu representasi yang sama serta menghilangkan imbuhan (awalan dan akhiran). Misanya kata “mendengarkan”, “dengarkan”, “didengarkan”, “didengar” akan ditransformasi menjadi kata “dengar”. Hal ini bertujuan untuk mereduksi variasi kata menjadi akar kata (Arimbawa & ER, 2020). Berikut ini adalah contoh dokumen hasil tahapan *stemming* yang dapat dilihat pada **Tabel 5.4**:

Tabel 5.4. Contoh dokumen pada tahap *stemming*

Sebelum	Sesudah
'recomended', 'pakai', 'coba', 'exfoliasi', 'kulit', 'pake', 'ngaruh', 'hidungku', 'haluss', 'ngurangi', 'komedo', 'hidungku', 'muka', 'halus', 'berminyak', 'kering', 'muka', 'pake', 'mukaku', 'muncul', 'jerawat', 'berlebih', 'muka', 'pake', 'mukanya', 'halus', 'komedonya', 'berkurang', 'daerah', 'hidung', 'mata', 'keliatan', 'komedonya', 'dipegang', 'bruntus', 'mukaku', 'cerahan', 'habis', 'cuci', 'muka', 'berminyak', 'kering', 'pokoknya', 'bagus', 'bangett', 'anjurannya', 'dipakai', 'malam', 'dipakai', 'malem', 'pake', 'malem'	'rekomendasi', 'pakai', 'coba', 'exfoliasi', 'kulit', 'pakai', 'pengaruh', 'hidung', 'halus', 'kurang', 'komedo', 'hidung', 'muka', 'halus', 'minyak', 'kering', 'muka', 'pakai', 'muka', 'muncul', 'jerawat', 'lebih', 'muka', 'pakai', 'muka', 'halus', 'komedo', 'kurang', 'daerah', 'hidung', 'mata', 'lihat', 'komedo', 'pegang', 'bruntus', 'muka', 'cerah', 'habis', 'cuci', 'muka', 'minyak', 'kering', 'pokok', 'bagus', 'banget', 'anjuran', 'pakai', 'malam', 'pakai', 'malem', 'pakai', 'malem'

5.3. Pembobotan *Term Frequency-Invers Document Frequency* (TF-IDF)

Pembobotan TF-IDF merupakan suatu hubungan kata (*term*) yang berada pada dokumen yang akan diberikan suatu nilai bobot. Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu frekuensi kemunculan sebuah kata dalam sebuah dokumen yang disebut dengan *Term Frequency* (TF) dan *inverse* frekuensi dokumen yang mengandung kata yang disebut dengan *Inverse Document Frequency* (IDF). Tujuan dari tahap ini yaitu untuk melihat seberapa penting kata (*term*) tersebut di dalam dokumen. Langkah pertama yaitu menghitung *Term Frequency* (TF). Peneliti mengambil beberapa kata dari dokumen seperti kata "bersih", "cocok", "facial", "jerawat", "kering", dan "kulit" sebagai sampel hasil nilai dari TF yang dapat dilihat pada **Tabel 5.5**.

Tabel 5.5. Contoh hasil perhitungan TF

Dokumen	bersih	<i>facial</i>	jerawat	kering	...	kulit
1	0	0	1	0	...	0
2	1	0	0	0	...	1
3	0	0	0	0	...	0
4	1	2	1	1	...	1
5	0	0	2	1	...	2
6	0	0	0	1	...	2
7	0	0	1	0	...	0
8	0	3	0	1	...	0
9	1	0	2	0	...	1
10	0	0	0	1	...	0
11	1	0	0	1	...	0
12	0	0	0	0	...	0
13	0	0	2	0	...	1
14	0	0	0	0	...	0
15	0	2	1	0	...	0
16	1	1	2	0	...	1
17	1	0	0	0	...	0
18	0	0	1	1	...	0
19	0	1	2	0	...	0
20	0	0	2	1	...	0
:	:	:	:	:	...	:
20917	0	0	1	0	...	0

Setelah dilakukan perhitungan TF, langkah selanjutnya yaitu menghitung nilai *Inverse Document Frequency* (IDF). IDF merupakan nilai \ln dari jumlah seluruh dokumen dibagi dengan jumlah dokumen yang mengandung kata terpilih (df) yang tertera pada **Persamaan 3.2**. Sebagai contoh pada kata "*facial*", maka dalam menghitung nilai df akan menampilkan jumlah dokumen yang memuat kata "*facial*". Dari proses perhitungan df didapatkan banyaknya dokumen yang

mengandung kata “*facial*” sebanyak 637 sehingga nilai tersebut merupakan nilai DF. Berikut merupakan perhitungan IDF dari kata “*facial*” :

$$IDF_{facial} = \ln\left(\frac{D}{df_{facial}}\right)$$

$$IDF_{facial} = \ln\left(\frac{20915}{637}\right) = 3.4914$$

Tahap selanjutnya yaitu melakukan pembobotan TF-IDF dengan cara mengalikan nilai TF dengan nilai IDF yang tertera pada **Persamaan 3.3**. Diketahui nilai D yaitu jumlah dokumen, pada penelitian ini yakni berjumlah 20915 dokumen. Sebagai contoh, perhitungan manual untuk kata “*facial*” pada dokumen ke-4 adalah sebagai berikut:

$$W_{facial,4} = TF_{facial,4} \times \left(\ln\frac{D}{df_{facial}} + 1\right)$$

$$W = 2 \times \left(\ln\left(\frac{20915}{637}\right) + 1\right)$$

$$W = 2 \times (3.49145223 + 1) = 8.98290446$$

Sehingga untuk hasil perhitungan TF-IDF dapat dilihat pada **Tabel 5.6**.

Tabel 5.6. Contoh hasil pembobotan TF-IDF

Dokumen	bersih	<i>facial</i>	jerawat	kering	...	kulit
1	0.000000	0.000000	2.727693	0.000000	...	0.000000
2	3.432476	0.000000	0.000000	0.000000	...	1.984053
3	0.000000	0.000000	0.000000	0.000000	...	0.000000
4	3.432476	8.982904	2.727693	2.780446	...	1.984053
5	0.000000	0.000000	5.455386	2.780446	...	3.968107
6	0.000000	0.000000	0.000000	2.780446	...	3.968107
7	0.000000	0.000000	2.727693	0.000000	...	0.000000
8	0.000000	13.470081	0.000000	2.780446	...	0.000000
9	3.432476	0.000000	5.455386	0.000000	...	1.984053
10	0.000000	0.000000	0.000000	2.780446	...	0.000000
11	3.432476	0.000000	0.000000	2.780446	...	0.000000
12	0.000000	0.000000	0.000000	0.000000	...	0.000000
13	0.000000	0.000000	5.455386	0.000000	...	1.984053

Dokumen	bersih	facial	jerawat	kering	...	kulit
14	0.000000	0.000000	0.000000	0.000000	...	0.000000
15	0.000000	8.980054	2.727693	0.000000	...	0.000000
16	3.432476	4.490027	5.455386	0.000000	...	1.984053
17	3.432476	0.000000	0.000000	0.000000	...	0.000000
18	0.000000	0.000000	2.727693	2.780446	...	0.000000
19	0.000000	4.490027	5.455386	0.000000	...	0.000000
20	0.000000	0.000000	5.455386	2.780446	...	0.000000
:	:	:	:	:
20917	0.000000	0.000000	2.727693	0.000000	...	0.000000

Hasil pembobotan kata TF-IDF menunjukkan seberapa pentingnya kata dari sebuah dokumen terhadap seluruh dokumen. Sebuah kata dikatakan penting apabila banyak kemunculan kata di dalam dokumen tersebut tinggi. Semakin banyak suatu kata muncul di dalam suatu dokumen maka bobot hubungan antara suatu kata dan suatu dokumen akan semakin tinggi. Sebaliknya, jika kemunculan suatu kata di dalam suatu dokumen sedikit maka bobot hubungan antara suatu kata dan suatu dokumen akan semakin rendah.

5.4. Cosine Similarity

Menghitung nilai *cosine similarity* dilakukan untuk membandingkan kemiripan antar dokumen. Nilai *cosine similarity* berada pada rentang 0-1. Jika nilai *cosine similarity* semakin mendekati 1, maka semakin mirip antara teks A dan B. Sebaliknya, jika nilai *cosine similarity* mendekati 0, maka semakin tidak mirip antara teks A dan B. Sebagai contoh, peneliti akan mencari nilai *cosine similarity* antara dokumen ke-14716 dengan dokumen ke-709 menggunakan **Persamaan 3.4** sebagai berikut:

$$\text{Cos } \alpha = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

$$\text{Cos } \alpha = \frac{\sum_{i=1}^n D14716_i \times D709_i}{\sqrt{\sum_{i=1}^n (D14716_i)^2} \times \sqrt{\sum_{i=1}^n (D709_i)^2}}$$

$$\begin{aligned} \cos \alpha &= \frac{(0,0 \times 0,0) + (0,0 \times 0,0) + \dots + (0,0 \times 0,0)}{\sqrt{(0,0 \times 0,0)^2 + \dots + (0,0 \times 0,0)^2} \times \sqrt{(0,0 \times 0,0)^2 + \dots + (0,0 \times 0,0)^2}} \\ \cos \alpha &= \frac{782,259997}{30,467557 \times 29.330369} = 0,875397 \end{aligned}$$

Nilai kemiripan dokumen ke-14716 terhadap dokumen ke-709 yaitu sebesar 87%. Disini peneliti akan mencari 5 dokumen yang paling mirip dengan dokumen ke-14716 dengan mencari nilai *cosine similarity* yang diurutkan dari nilai terbesar seperti pada **Tabel 5.7**.

Tabel 5.7. 5 Dokumen yang paling mirip dengan dokumen ke-14716

No.	Dokumen	Nilai <i>Cosine Similarity</i>
1	709	0,875379
2	5971	0,774920
3	19872	0,670935
4	15908	0,385445
5	18145	0,296287

5.5. Hasil Rekomendasi

Berdasarkan hasil nilai *cosine similarity*, diperoleh rekomendasi produk yang mempunyai nilai *cosine similarity* atau nilai kemiripan paling besar. Pada dokumen ke-14726 yaitu produk *serum* dan *essence* dari *brand* Raiku Beauty dengan nama “Antioxidant Serum”, akan ditampilkan 5 produk yang memiliki kemiripan dengan dokumen tersebut. Maka berdasarkan hasil dari **Tabel 5.7**, berikut merupakan 5 urutan teratas hasil rekomendasi produk yang tertera pada **Tabel 5.8**.

Tabel 5.8. 5 Rekomendasi produk “Antioxidant Serum”

No.	Dokumen	Nama Produk	Nilai <i>Cosine Similarity</i>
1	709	Antioxidant Cleansing Foam	0,875379
2	5971	Antioxidant Toner	0,774920
3	19872	Antioxidant Soothing Gel	0,670935
4	15908	Antioxidant Night Cream	0,385445
5	18145	Antioxidant Morning Cream	0,296287

Berdasarkan hasil dari **Tabel 5.8**, maka ketika *user* melihat produk “Antioxidant Serum”, akan muncul 5 rekomendasi produk *skincare* yang mempunyai kemiripan dengan produk tersebut berdasarkan perhitungan dengan metode *content-based filtering*. Kemiripan dalam hal ini diperoleh dari *review* pengguna *skincare*. Sebagai contoh, apabila pengguna produk “Antioxidant Serum” menuliskan *review* yang memuat kesan dari penggunaan produk ataupun kondisi kulit yang sesuai dengan produk, maka hasil rekomendasi produk akan mempunyai *review* pengguna yang mirip dengan produk “Antioxidant Serum”.

Tabel 5.9. *Review* produk “Antioxidant Serum” dan “Antioxidant Cleansing Foam”

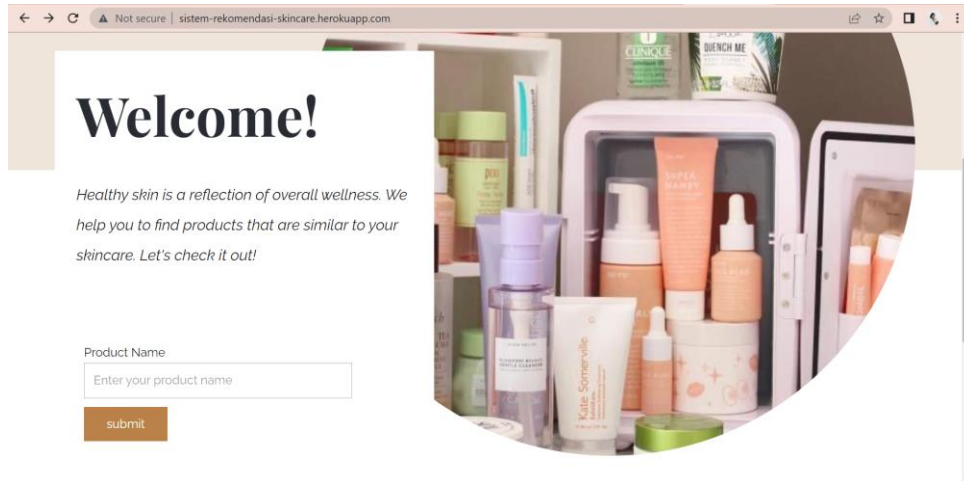
Antioxidant Serum	Antioxidant Cleansing Foam
gel transparan dan cepet meresap, suka ak sama produk yg gampang meresap begini. Series antioxidant ini beneran seperti namanya. Bisa menangkal radikal bebas karna mengandung Kakadu Plum Extract dan Hydroxyacetophenone yang tinggi antioxidant.	Formulanya enak banget. lembut gitu, jadi gak bikin kulit kering dan ketarik. Series antioxidant ini beneran seperti namanya. Bisa menangkal radikal bebas karna mengandung Kakadu Plum Extract dan Hydroxyacetophenone yang tinggi antioxidant.

Berdasarkan **Tabel 5.9**, dapat dilihat bahwa antara produk “Antioxidant Serum” dengan produk “Antioxidant Cleansing Foam” mempunyai kemiripan pada *review* penggunaannya, hal ini selaras dengan nilai *cosine similarity* yang cukup besar yaitu 87%.

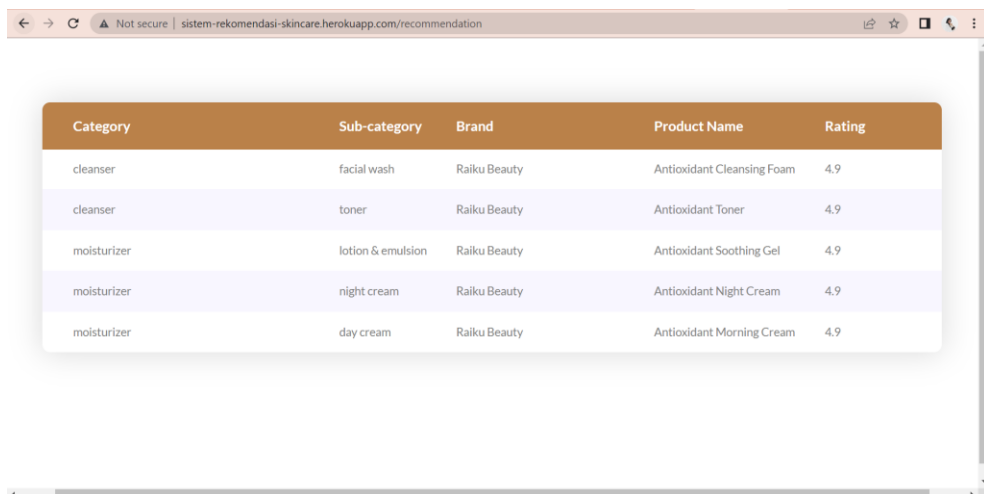
5.6. Tampilan Website

Pembuatan *website* dilakukan untuk mempermudah pengguna dalam mencari rekomendasi *skincare* yang dapat diakses di <http://sistem-rekomendasi-skincare.herokuapp.com/>. Pada halaman awal akan menampilkan kolom untuk meng-*input* nama produk *skincare*, kemudian klik *submit*. Maka *user* akan diarahkan ke halaman selanjutnya yang memuat hasil rekomendasi *skincare* yang sesuai dengan produk tersebut yaitu sebanyak lima produk *skincare*. Hasil

rekomendasi juga dilengkapi dengan informasi mengenai produk seperti kategori, sub-kategori, merek, dan *rating* seperti yang dapat dilihat pada **Gambar 5.3** dan **5.4** :



Gambar 5.3. Tampilan awal *website*

The image shows a web browser window with the URL 'sistem-rekomendasi-skincare.herokuapp.com/recommendation'. It displays a table with five columns: Category, Sub-category, Brand, Product Name, and Rating. The table contains five rows of product recommendations, all from the brand 'Raiku Beauty' with a rating of 4.9.

Category	Sub-category	Brand	Product Name	Rating
cleanser	facial wash	Raiku Beauty	Antioxidant Cleansing Foam	4.9
cleanser	toner	Raiku Beauty	Antioxidant Toner	4.9
moisturizer	lotion & emulsion	Raiku Beauty	Antioxidant Soothing Gel	4.9
moisturizer	night cream	Raiku Beauty	Antioxidant Night Cream	4.9
moisturizer	day cream	Raiku Beauty	Antioxidant Morning Cream	4.9

Gambar 5.4. Tampilan hasil rekomendasi

BAB VI

PENUTUP

6.1. Kesimpulan

Berdasarkan hasil dan pembahasan pada penelitian ini, maka dapat ditarik kesimpulan sebagai berikut:

1. Untuk merekomendasikan produk *skincare* berdasarkan *review* dari Female Daily menggunakan pendekatan *content-based filtering* dimulai dari melakukan *pre-processing* data untuk membersihkan data, kemudian melakukan pembobotan *Term Frequency-Invers Document Frequency* (TF-IDF) yang berfungsi untuk mengetahui seberapa penting sebuah kata dalam sekelompok kata atau dokumen, kemudian dari hasil pembobotan TF-IDF dilakukan perhitungan *cosine similarity* untuk mengetahui nilai kemiripan antara dua dokumen. Hasil rekomendasi produk *skincare* merupakan produk yang mempunyai nilai *cosine similarity* tertinggi.
2. Hasil rekomendasi yang didapatkan yaitu memberikan lima rekomendasi produk terhadap pengguna (*user*) sebagai contoh berdasarkan *review* pada produk “Antioxidant Serum” berturut-turut yaitu “Antioxidant Cleansing Foam”, “Antioxidant Toner”, “Antioxidant Soothing Gel”, “Antioxidant Night Cream”, dan “Antioxidant Morning Cream” dengan nilai *cosine similarity* berturut-turut adalah 0.875379, 0.772920, 0.670935, 0.385445, 0.296287.

6.2. Saran

Dari penelitian yang sudah dilakukan, maka peneliti dapat memberikan saran sebagai berikut:

1. Bagi penelitian selanjutnya, diharapkan dapat menambahkan *stopwords* yang tidak tercantum pada *library simager* agar hasil *filtering* kata dapat lebih baik.
2. Dapat menambahkan analisis sentimen pada *review* untuk mengklasifikasikan *review* positif dan negatif dari suatu produk.
3. Merekomendasikan produk tidak hanya berdasarkan *review*, namun dikombinasikan dengan faktor lainnya misalnya *rating*, profil pengguna, dan lain-lain.

4. Mengembangkan atau membandingkan dengan jenis metode rekomendasi lainnya seperti metode *Collaborative Filtering*, *Hybrid Recommendation System*, dan metode lainnya.

DAFTAR PUSTAKA

- Alkaff, M., Khatimi, H., & Eriady, A. (2020). Sistem Rekomendasi Buku Menggunakan Weighted Tree Similarity dan Content Based Filtering. *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 193-202.
- Amanda, M. (2021, 7 16). *HelloSehat*. Diambil kembali dari hellosehat: <https://hellosehat.com/sehat/merawat-kulit-wajah-sejak-dini/>.
- Aulia, G. N., & Patriya, E. (2019). Implementasi Lexicon Based dan Naive Bayes Pada Analisis Sentimen Pengguna Twitter Topik Pemilihan Presiden 2019. *Jurnal Ilmiah Informatika Komputer*, 4-5.
- Badriyah, T. (2018). Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori. *Konferensi Nasional Sistem Informasi*, 555.
- Echols, J. M., & Shadily, H. (2003). *Kamus Inggris Indonesia*. Jakarta: PT. Gramedia Pustaka Utama.
- Ferio, G., Intan, R., & Rostianingsih, S. (2018). Sistem Rekomendasi Mata Kuliah Pilihan Menggunakan Metode User Based Collaborative Filtering Berbasis Algoritma Adjusted Cosine Similarity. *Teknik Informatika Fakultas Teknologi Industri Universitas Kristen Petra*, 2.
- Fiarni, C., Maharani, H., & Calista, N. (2019). Product Recommendation System Design Using Cosine Similarity and Content-based Filtering Methods. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, 42-48.
- Girsang, A. S. (2020, November 17). Diambil kembali dari <https://mti.binus.ac.id/2020/11/17/sistem-rekomendasi-content-based/>
- Girsang, A. S. (2020, November 17). *MTI*. Diambil kembali dari Binus University: <https://mti.binus.ac.id/2020/11/17/sistem-rekomendasi-content-based/>
- Ikhsania, A. A. (2020, 12 31). *sehatq*. Diambil kembali dari sehatq.com: <https://www.sehatq.com/artikel/arti-skincare-serta-manfaat-dan-macam-macamnya-sudah-tahu>

- Imbar, R. V., Adelia, Ayub, M., & Rehatta, A. (2014). Implementasi Cosine Similarity dan Algoritma Smith-Waterman untuk Mendeteksi Kemiripan Teks. *Jurnal Informatika*, 31-42.
- Kalakota, R., & Whinston, A. B. (1997). *Electronic Commerce*. Harlow, England: Addison-Wesley.
- Konguasa, R. R. (2018). *Rancang Bangun Aplikasi Rekomendasi Wisata Museum di Jakarta Dengan Metode Item Based Collaborative Filtering Berbasis Android*. Universitas Multimedia Nusantara.
- Larasati, F. B., & H. F. (2021). Sistem Rekomendasi Product Emina Cosmetics Dengan Menggunakan Metode Content-Based Filtering. *MISI (Jurnal Manajemen Informatika dan Sistem Informasi)*, 2614-3739.
- Luqyana, W. A., Cholissudin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4704-4713.
- Masruri, F., & Mahmudy, W. F. (2007). Personalisasi Web E-Commerce Menggunakan Recommender System dengan Metode Item-Based Collaborative Filtering. *Kursor*, 3.
- Mondi, R. H., Wijayanto, A., & Winarno. (2019). Recommendation System With Content-based Filtering Method For Culinary Tourism in Mangan Application. *ITSMART: Jurnal Ilmiah Teknologi dan Informasi*, 66.
- Na'imah, S. (2021, 07 16). *hellosehat*. Diambil kembali dari [hellosehat.com: https://hellosehat.com/penyakit-kulit/perawatan-kulit/produk-skincare/](https://hellosehat.com/penyakit-kulit/perawatan-kulit/produk-skincare/)
- Nayoan, A. (2021, September 5). Diambil kembali dari <https://www.niagahoster.co.id/blog/apa-itu-ecommerce/?amp>
- Perwitasari, N. H. (2021, 10 29). *tirto.id*. Diambil kembali dari [tirto.id: https://tirto.id/mengenal-arti-skincare-dan-tahapan-merawat-kulit-eimA](https://tirto.id/mengenal-arti-skincare-dan-tahapan-merawat-kulit-eimA)
- Prasetyo, E. (2012). *Data mining konsep dan aplikasi menggunakan matlab*. Yogyakarta: Andi.
- Putra, A. I., & Santika, R. R. (2020). Implementasi Machine Learning dalam Penentuan Rekomendasi Musik dengan Metode Content-Based Filtering. *EDUMATIC (Jurnal Pendidikan Informatika)*, 121-130.

- Putri, M. W., Muchayan, A., & Kamisutara, M. (2020). Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF. *JOINTECS*, 229-236.
- Rizaty, M. A. (2021, 10 5). *databoks*. Diambil kembali dari databoks.katadata: <https://databoks.katadata.co.id/datapublish/2021/10/05/industri-kosmetik-tumbuh-559-persen-ini-merek-perawatan-tubuh-terlaris-pada-agustus-2021>.
- Safitri, D. A., Helilintar, R., & Wahyuniar, L. S. (2021, Juli 24). Rekomendasi Skincare Menggunakan Metode Content-Based Filtering dan Algoritma Apriori. *Seminar Nasional Inovasi Teknologi*, hal. 242-248.
- Salsabil, Z. (2019). Efektivitas Website Femaledaily.com Dalam Memenuhi Kebutuhan Informasi Pengguna. *Jurnal Ilmu Perpustakaan*, 200.
- Samuel, R., Natan, R., Fitria, & Syafiqoh, U. (2018). Penerapan Cosine Similarity dan K-Nearest Neighbor (K-NN) pada Klasifikasi dan Pencarian Buku. *Journal of Big Data Analytic and Artificial Intelligence*, 10.
- Schafer. (2007). *Collaborative Filtering*. Jakarta: Buku Kedokteran Jakarta.
- Sebastia, L., Garcia, I., Onaindia, E., & Guzman, C. (2009). A tourist recommendation and planning application. *International Journal on Artificial Intelligence Tools*, 717-738.
- Setiaji, A. R. (2014). Analisis Sentimen Terhadap Acara Televisi Berdasarkan Opini Publik. *Universitas Komputer Indonesia*, 4.
- Swari, R. C. (2021, 10 15). *hello sehat*. Diambil kembali dari hellosehat: <https://hellosehat.com/penyakit-kulit/struktur-kulit-manusia/>.
- Turban, E., King, D., Lee, J. K., Liang, T.-P., & Turban, D. C. (2002). *Electronic Commerce: : A Managerial and Social Networks Perspective*. New York: Prentice Hall.
- Wijaya, A. E., & Alfian, D. (2018). Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering dan Content-Based Filtering. *Jurnal Computech dan Bisnis*, 11-27.
- Yusuf, M., & Cherid, A. (2020). Implementasi Algoritma Cosine Similarity dan Metode TF-IDF Berbasis PHP Untuk Menghasilkan Rekomendasi Seminar. *Jurnal Fakultas Ilmu Komputer*, 10-11.

LAMPIRAN

Lampiran 1 Data Review Produk *skincare* di Femaledaily

No	Category	Sub-category	Brand	Product Name	Rating	Review
1	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	Huhuhu ini sabun muka terbaik yang pernah aku coba, gatau udah botol ke berapa, cetaphil ini juga salah satu yang buat jerawatku hilang. Memang sebagus itu sih untuk harga yang cukup mahal. Tapi pas awal make bingung banget kok gk keset, karena menurutku sabun muka yang oketu yang keset, eh rupanya ini malah ngebantu ilangin jerawat, udah make dari april 2019 gk pernah ganti, sampe sekarang

2	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	busa produk ini hampir ngga ada, bagus tapi ngga sewaw itu di kulit aku. ngga buat iritasi juga, tp g ngasih perubahan yang signifikan juga. buat yang suka cuci muka yang bener bener ngasih efek bersih, menurut aku ini agak kurang karna agak lengket dikulit
3	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	ini sabun muka yang bener bener gentle baguss bangett apalagi kalau lagi breakout pakai sabun cuci muka ini enak bangett walaupun gk terlalu ada ngefek di muka aku tapi dia bener bener baguss kalau lagi breakout suka bangett dari dulu selalu pakai ini dan memang langsung bikin jatuh cinta

4	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1 facial wash yg udah aq pake lebih dari 5 th....kulitku super sensitif, kadang oily,kadang kering tp Alhamdulillah pas nyoba ini cocok bgt2, kulit jd bersih, jerawat kecil2nya jg hilang, dari awal emank nyari facial wash yg ga ada busanya dan ini pas bgt, cuz klo dikulitku yg super sensitif ini ga bisa pake yg ada busanya/ yg ada scrubnya pokokknya di aq ga cocok, dan klo pake cetaphil ini suka berada diruangan yg sering ber AC disarankan setelahnya dipakein moisturizing agar wajahnya bisa lembab, itu aja sie menurutku..mgkn aq terus pake selamanya x yak cuz cocok bgt...ga mau pindah ke merk lain deh hihi... recommended...ðŸ• ðŸ•
---	----------	-------------	----------	----------------------	--

5	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	Aku kira bakalan cocok ternyata ngga, Pertama kulit kemerahan ku tidak mereda sama sekali, apalagi jerawat ya tidak berpengaruh sama sekali. Jidatku yang sebelumnya mulus jadi full bruntusan, jerawat tambah jadi banyak Merah ² kaya habis di tampar. mungkin kalo kulit kering bakalan bagus sih soalnya cukup melembabkan
:	:	:	:	:	:	:
20914	moisturizer	face oil	Re-Gen	Re-Gen Oil	4.1	Beli ini krna direkomendasiin mba ² kimia farmanya.. Aku pake seminggu mukaku langsng breakout muncul beberapa jerawat batu (baru sembuh pas dibawa ke SPKK). Masih sisa banyak pdhl TT baunya jg wangi enak, tadinya mau dipake buat ngilangin bekas jrwt

Data selengkapnya dapat diakses di :

<http://bit.ly/Skripsi-FarahAyu>

Lampiran 2 Script dan output sistem rekomendasi

```
In [1]: from simager.preprocess import TextPreprocess
import pandas as pd
import numpy as np
import warnings
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
from nltk_id.stopword import StopWord
from sklearn.feature_extraction.text import CountVectorizer
warnings.filterwarnings('ignore')
```

```
In [2]: # Define Methods For Text Preprocessing

method = [
    "rm_hashtag", # "ini adalah contoh #hashtag" --> 'ini adalah contoh '
    "rm_mention", # "ini adalah contoh @mention" --> 'ini adalah contoh '
    "rm_nonascii", # "ini adalah contoh .(朱惠英) d\xe5" --> 'ini adalah contoh .() da'
    "rm_emoticons", # "ini adalah contoh 😍" --> 'ini adalah contoh '
    "rm_html", # "'<html><body>ini adalah contoh</body></html>'" --> 'ini adalah contoh '
    "rm_url", # "ini adalah contoh url http://google.com" --> 'ini adalah contoh url '
    "sparate_str_num", # "ini adalah angka3" --> 'ini adalah angka 3'
    "pad_punct", # "ini adalah contoh.saya" --> 'ini adalah contoh. saya'
    "rm_punct", # "$ ini *adalah contoh" --> ' ini adalah contoh'
    "rm_repeat_char", # "ini adalah contohhhhhhhhhhh" --> 'ini adalah contohh'
    "rm_repeat_word", # "ini ini ini ini ini ini adalah contoh" --> 'ini ini adalah contoh'
    "rm_num", # "ini adalah contoh 123" --> 'ini adalah contoh '
    "rm_whitespace", # " ini adalah contoh " --> 'ini adalah contoh'
    "normalize", # "yg dg" --> 'yang dengan'
    "stopwords" # "ini adalah contoh stopword" --> 'stopword'
]

cleaner = TextPreprocess(methods=method)
```

```
In [3]: # Load Dataset

data = pd.read_csv('TEST.csv')
data.head()
```

```
Out[3]:
```

	Category	Sub-category	Brand	Product Name	Rating	Review
0	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	Huhuhu ini sabun muka terbaik yang pernah aku ...
1	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	busa produk ini hampir ngga ada, bagus tapi ng...
2	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	ini sabun muka yang bener bener gentle baguss ...
3	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	facial wash yg udah aq pake lebih dari 5 th.....
4	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	Aku kira bakalan cocok ternyata ngga, Pertama ...

```
In [4]: # Apply Simager To Dataset

cleantext = []
stopword = StopWord()

for i in range(len(data)):
    data['Review'][i] = " ".join([str(entry).lower() for entry in data['Review'][i]])
    data['Review'][i] = stopword.remove_stopword(data['Review'][i])

for i in range(len(data)):
    x = cleaner(data['Review'][i])
    cleantext.append(x)
```

```
In [5]: # Fitur Yang Hendak Direkomendasikan

def combined_features(row):
    return row['Category']+" "+row['Sub-category']+" "+row['Brand']

data["combined_features"] = data.apply(combined_features, axis =1)
data.head()
```

```
Out[5]:
```

	Category	Sub-category	Brand	Product Name	Rating	Review	Bersih	combined_features
0	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	huhuhu sabun muka terbaik coba , gatau botol, ...	huhu sabun muka terbaik coba enggak tau botol ...	cleanser facial wash Cetaphil
1	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	busa produk ngga, bagus ngga sewaw kulit. ngga...	busa produk bagus sewaw kulit lritasi men...	cleanser facial wash Cetaphil
2	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	sabun muka bener bener gentle baguss banget b...	sabun muka gentle bagus banget breakout paka...	cleanser facial wash Cetaphil
3	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	facial wash aq pake 5 th... kulitku super sen...	facial wash pakai kulitku super sensitif ol...	cleanser facial wash Cetaphil
4	cleanser	facial wash	Cetaphil	Gentle Skin Cleanser	4.1	cocok ngga, kulit kemerahan ku mereda, jerawat...	cocok kulit kemerahan ku mereda jerawat berpe...	cleanser facial wash Cetaphil

In [6]: # TF (Review)

```
vect = CountVectorizer()
vect.fit(data['Bersih'])
dtm_tf = vect.transform(data['Bersih'])
dtm_tf = pd.DataFrame(dtm_tf.toarray(), columns = vect.get_feature_names())
dtm_tf

#dtm_tf.to_csv('tf.csv')
```

Out[6]:

	aa	aagar	aagghh	aah	aaja	aak	aakk	aaku	aand	aandd	...	zonenya	zoom	zoonk	zorro	zoya	zoyaa	zuper	zuzur	zuzurly	zwijsal
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
...
20910	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
20911	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
20912	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
20913	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

In [7]: # Fit To TF IDF (Review)

```
vect1 = TfidfVectorizer(norm = None)
vect1.fit(data['Bersih'])
dtm_tf_idf = vect1.transform(data['Bersih'])
dtm_tf_idf = pd.DataFrame(dtm_tf_idf.toarray(), columns = vect1.get_feature_names())
dtm_tf_idf.head()

#####dtm_tf_idf.to_csv('tf-idf.csv')#####
```

Out[7]:

	aa	aagar	aagghh	aah	aaja	aak	aakk	aaku	aand	aandd	...	zonenya	zoom	zoonk	zorro	zoya	zoyaa	zuper	zuzur	zuzurly	zwijsal
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	0.0	0.0	0.0	0.0	0.0	0.0
1	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	0.0	0.0	0.0	0.0	0.0
2	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	0.0	0.0	0.0	0.0	0.0
3	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	0.0	0.0	0.0	0.0	0.0
4	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	0.0	0.0	0.0	0.0	0.0

5 rows x 28584 columns

In [8]: # Get Cosine Similarity (Review)

```
cosine_sim = cosine_similarity(dtm_tf_idf)
co_sim=pd.DataFrame(cosine_sim)
co_sim.head()

###co_sim.to_csv('cosine.csv')
```

Out[8]:

	0	1	2	3	4	5	6	7	8	9	...	20905	20906	20907	20908	20909
0	1.000000	0.017518	0.226007	0.067796	0.020942	0.042580	0.008720	0.084564	0.047277	0.142036	...	0.000000	0.027838	0.011967	0.017382	0.000000
1	0.017518	1.000000	0.121353	0.023482	0.031211	0.026113	0.000000	0.039613	0.051118	0.061897	...	0.065014	0.015607	0.042470	0.022891	0.007923
2	0.226007	0.121353	1.000000	0.087700	0.026886	0.031789	0.140072	0.000000	0.060550	0.127154	...	0.066891	0.052732	0.045471	0.082395	0.021030
3	0.067796	0.023482	0.087700	1.000000	0.052174	0.032231	0.114503	0.233547	0.111441	0.128356	...	0.042488	0.034068	0.062735	0.072365	0.028940
4	0.020942	0.031211	0.026886	0.052174	1.000000	0.053206	0.024967	0.012747	0.055270	0.022558	...	0.017412	0.014329	0.038438	0.017146	0.026430

5 rows x 20915 columns

In [9]: # Recommendation Function (Review)

```
def recommendation_review (product):
    indexprod = data[data['Product Name'] == product].index.values[0]
    similar_review = list(enumerate(cosine_sim[indexprod]))
    sorted_similar_review = sorted(similar_review, key=lambda x:x[1], reverse=True)
    aa = []
    for i in range(1,6) :
        aa.append(sorted_similar_review[i][0])
    return data.iloc[aa,:5]
```

```
In [10]: # Get Recommendation
product = "Antioxidant Serum"
print(f'REKOMENDASI PRODUCT {product.upper()} BERDASARKAN REVIEW ADALAH : ')
display(recomendation_review(product))
```

REKOMENDASI PRODUCT ANTIOXIDANT SERUM BERDASARKAN REVIEW ADALAH :

	Category	Sub-category	Brand	Product Name	Rating
707	cleanser	facial wash	Raiku Beauty	Antioxidant Cleansing Foam	4.9
5969	cleanser	toner	Raiku Beauty	Antioxidant Toner	4.9
19870	moisturizer	lotion & emulsion	Raiku Beauty	Antioxidant Soothing Gel	4.9
15906	moisturizer	night cream	Raiku Beauty	Antioxidant Night Cream	4.9
18143	moisturizer	day cream	Raiku Beauty	Antioxidant Morning Cream	4.9

