

**SISTEM REKOMENDASI PRODUK *SUNSCREEN*  
UNTUK PERLINDUNGAN DARI SINAR  
*ULTRAVIOLET* DENGAN METODE *CONTENT-  
BASED FILTERING***

(Studi Kasus : Produk *Suncreen* di Website *Sociolla*)

**TUGAS AKHIR**

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



Disusun Oleh:

Khoirul Ni'am

20611076

**PROGRAM STUDI STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2024**

**HALAMAN PERSETUJUAN PEMBIMBING  
TUGAS AKHIR**

Judul : Sistem Rekomendasi Produk *Sunscreen* untuk  
Perlindungan dari Sinar *Ultraviolet* dengan  
Metode *Content-Based Filtering* (Studi Kasus:  
Produk *Sunscreen* di *Website Sociolla*)

Nama Mahasiswa : Khoirul Ni'am

NIM : 20611076

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


Mengetahui

Ketua Prodi Statistika

Yogyakarta, 13 Mei 2024

Dosen Pembimbing

  
(Dr. Atina Andika, S.Si., M.Si.)

  
(Arum Handini Primandari, S.Rd.Si., M.Sc.)

# HALAMAN PENGESAHAN

## TUGAS AKHIR

**Sistem Rekomendasi Produk *Sunscreen* untuk Perlindungan dari Sinar  
*Ultraviolet* dengan Metode *Content-Based Filtering*  
(Studi Kasus : Produk *Sunscreen* di *Website Sociolla*)**

**Nama Mahasiswa : Khoirul Ni'am**

**NIM : 20611076**

**TUGAS AKHIR INI TELAH DIUJIKAN**

**PADA TANGGAL : 19 Juni 2024**

**Nama Penguji**

**Tanda Tangan**

1. Dina Tri Utari, S.Si., M.Sc.

2. Dr. techn., Rohmatul Fajriyah, S.Si., M.Si.

3. Arum Handini Primandari, S.Pd.Si., M.Sc.



Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



**Prof. Riyanto, S.Pd., M.Si., Ph.D.**



## KATA PENGANTAR



*Assalamu 'alaikum Wr. Wb*

*Alhamdulillah*, puji dan syukur dipanjatkan atas kehadiran Allah SWT karena telah memberikan kesempatan, kekuatan, kesehatan dan kemudahan dalam menyelesaikan tugas akhir ini. Shalawat serta salam tidak lupa dipanjatkan kepada junjungan besar kita, Nabi Muhammad SAW yang telah membawa umat islam dari zaman kebodohan hingga ke zaman yang terang benderang seperti sekarang.

Atas izin dan ridho Allah SWT, akhirnya penulis dapat menyelesaikan tugas akhir yang berjudul “Sistem Rekomendasi Produk *Sunscreen* untuk Perlindungan dari Sinar *Ultraviolet* dengan Metode *Content-Based Filtering*”. Sebagai persyaratan untuk mendapatkan gelar S1-Statistika sekaligus syarat kelulusan. Tugas akhir ini dapat diselesaikan dengan baik dan tepat waktu berkat dukungan serta arahan dari berbagai pihak, Oleh karena itu, penulis menyampaikan rasa terima kasih kepada:

1. Kedua orang tua yang selalu memberik banyak kontribusi, motivasi, dukungan dan semangat bagi penulis dalam setiap keadaan.
2. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia
3. Ibu Dr. Atina Ahdika, S.Si., M.Si. selaku Ketua Program Studi S1-Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
4. Ibu Arum Handini Primandari, S.Pd.Si., M.Sc. selaku Dosen Pembimbing tugas akhir yang telah membimbing dalam penelitian dan penulisan dari awal hingga penyusunan laporan tugas akhir ini.
5. Bapak Dr. Raden Bagus Fajriya Hakim, S.Si., M.Si. selaku Dosen Pembimbing Akademik yang telah banyak memberikan dukungan dan motivasi dari awal masuk kuliah hingga selesai.

6. Seluruh teman-teman Angkatan Statistka 2020 yang telah berjuang bersama dalam menyelesaikan tugas akhir ini.

Penulis menyadari bahwa tugas akhir ini yang masih jauh dari kata sempurna. Oleh karena itu, penulis memohon maaf yang sebesar-besarnya dan menghararapkan kritik serta saran seluas-luasnya yang bersifat membangun yang kemudian akan penulis jadikan sebagai evaluasi.

*Wassalamualaikum Wr.Wb*

Yogyakarta, 13 Mei 2024

A handwritten signature in black ink, consisting of several overlapping loops and a long horizontal stroke extending to the right.

(Khoirul Ni'am)

## 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.....	4
1.5. Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1. Penelitian Terdahulu.....	6
BAB III LANDASAN TEORI.....	11
3.1. Tabir Surya ( <i>Sunscreen</i> ).....	11
3.1.1. <i>Sun Protecting Factor</i> (SPF).....	11
3.1.2. <i>Protection Grade of UV-A</i> (PA).....	12
3.1.3. Jenis Kulit.....	12
3.1.4. Bahaya Paparan Sinar UV.....	13
3.1.5. Komposisi <i>Sunscreen</i> .....	13
3.2. <i>Sociolla</i> .....	13
3.3. Analisa Deskriptif.....	14
3.4. <i>Text Mining</i> .....	15
3.4.1. <i>Case Folding</i> .....	15
3.4.2. <i>Remove Punctuation</i> .....	16
3.4.3. <i>Remove Double White Space</i> .....	16
3.4.4. <i>Stopword Removal</i> .....	16
3.4.5. <i>Lemmatization</i> .....	16
3.4.6. <i>Tokenizing</i> .....	17
3.5. Pembobotan <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF) .....	17
3.6. <i>Cosine Similarity</i> .....	19
3.7. Sistem Rekomendasi.....	19
3.8. <i>Content Based Filtering</i> .....	20
3.9. <i>Phyton Anywhere</i> .....	22
BAB IV METODOLOGI PENELITIAN.....	23
4.1. Populasi dan Sampel Penelitian.....	23
4.2. Tempat dan Waktu Penelitian.....	23
4.3. Definisi Operasional Variabel Penelitian.....	23

4.4.	Alat dan Cara Organisir Data .....	24
4.5.	Metode Penelitian.....	24
BAB V PEMBAHASAN .....		27
5.1.	Analisis Deskriptif.....	27
5.2.	<i>Preprocessing Data</i> .....	29
5.2.1	<i>Case Folding</i> .....	30
5.2.2	<i>Remove Punctuation</i> .....	32
5.2.3	<i>Remove Double Space</i> .....	34
5.2.4	<i>Lemmatization</i> .....	35
5.2.5	<i>Tokenizing</i> .....	37
5.3.	Pembobotan <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	38
5.4.	Hasil Rekomendasi Berdasarkan <i>Cosine Similarity</i> .....	41
5.5.	Tampilan <i>Website</i> Sistem Rekomendasi <i>Sunscreen</i> .....	48
BAB VI PENUTUP.....		52
6.1.	Kesimpulan.....	52
6.2.	Saran.....	53
DAFTAR PUSTAKA.....		54
LAMPIRAN .....		62

## DAFTAR TABEL

<b>Tabel 2.1</b> Tabel Penelitian Sebelumnya .....	6
<b>Tabel 4. 1</b> Definisi Operasional Variabel .....	23
<b>Tabel 4. 2</b> Alat dan Organisir Data .....	24
<b>Tabel 5. 1</b> Contoh Dokumen Pada <i>Text Preprocessing</i> .....	30
<b>Tabel 5. 2</b> Hasil <i>Case Folding</i> .....	31
<b>Tabel 5. 3</b> Hasil <i>Remove Punctuation</i> .....	32
<b>Tabel 5. 4</b> Hasil <i>Remove Double Space</i> .....	34
<b>Tabel 5. 5</b> Hasil Proses <i>Lemmatization</i> .....	35
<b>Tabel 5. 6</b> Hasil Tahapan <i>Tokenizing</i> .....	37
<b>Tabel 5. 7</b> Hasil Pembobotan Nilai TF .....	39
<b>Tabel 5. 8</b> Hasil Pembobotan Nilai TF-IDF .....	41
<b>Tabel 5. 9</b> Hasil Rekomendasi dengan Similaritas Tertinggi .....	42
<b>Tabel 5. 10</b> 5 Rekomendasi Produk <i>Sunscreen</i> Teratas .....	43
<b>Tabel 5. 11</b> <i>Ingredients</i> Dokumen ke-170 dan 171 .....	43
<b>Tabel 5. 12</b> Perbandingan <i>Ingredients</i> Dokumen ke-170 dan 171 .....	45
<b>Tabel 5. 13</b> Perbandingan <i>Ingredients</i> Dokumen ke-170 dan 92 .....	47
<b>Tabel 5. 14</b> Highlight Pencarian Produk dan Hasil Rekomendasi .....	49



## DAFTAR GAMBAR

<b>Gambar 3. 1</b> <i>SPF Rating dan UV Protection</i> (Sumber: Health Gennie) .....	12
<b>Gambar 3. 2</b> Logo <i>Sociolla</i> (Sumber: <i>Sociolla.com</i> ) .....	14
<b>Gambar 3. 3</b> Metode Sistem Rekomendasi (Sumber: (Isinkaye, 2015)).....	20
<b>Gambar 3. 4</b> Ilustrasi Sistem Rekomendasi (Sumber: Dokumen Pribadi) .....	21
<b>Gambar 4. 1</b> Diagram Alir Penelitian (Sumber: Dokumen Pribadi) .....	25
<b>Gambar 5. 1</b> 10 <i>Brand Sunscreen</i> Teratas di <i>Sociolla</i> .....	27
<b>Gambar 5. 2</b> <i>Boxplot</i> Harga <i>Sunscreen</i> .....	28
<b>Gambar 5. 3</b> <i>Wordcloud</i> pada <i>Ingredients Sunscreen</i> .....	29
<b>Gambar 5. 4</b> Tampilan <i>Home</i> Website Sistem Rekomendasi .....	48
<b>Gambar 5. 5</b> Tampilan Hasil Rekomendasi dari <i>Website</i> .....	49

## DAFTAR LAMPIRAN

Lampiran 1 <i>Dataset</i> .....	62
Lampiran 2 Perhitungan <i>Cosine Similarity</i> .....	62
Lampiran 3 <i>Script</i> Analisis Sistem Rekomendasi .....	62
Lampiran 4 <i>Deploy</i> Sistem dan <i>Web</i> .....	62

## 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, 13 Juni 2024



SEWILAH RIBU RUPIAH  
1000  
TOL. 20  
METERAI  
TEMPEL  
1D966ALX160482241

Khoirul Ni'am

## INTISARI

### **Sistem Rekomendasi Produk *Sunscreen* untuk Perlindungan dari Sinar *Ultraviolet* dengan Metode *Content-Based Filtering***

(Studi Kasus : Produk *Sunscreen* di *Website Sociolla*)

Khoirul Ni'am

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

Pemanasan global yang disebabkan oleh iklim yang ekstrim seperti sekarang membuat kebanyakan masyarakat sekarang harus lebih sadar untuk melindungi kulit dari paparan sinar UV, terutama bagi masyarakat yang hidup di wilayah tropis dan terkena sinar matahari sepanjang tahun. Paparan sinar UV yang berlebihan dalam jangka panjang dapat menyebabkan berbagai masalah kulit seperti iritasi, *sunburn* hingga pemicu kanker kulit. Berdasarkan permasalahan tersebut produk *sunscreen* hadir sebagai solusi untuk menjadi perlindungan terluar kulit. Dalam pemilihan *sunscreen* sering kali masyarakat masih bingung untuk memilih produk yang cocok sesuai kebutuhan kulit maupun manfaat dari *ingredients* yang terkandung didalamnya. Sistem rekomendasi berkembang untuk memudahkan masyarakat dalam memberikan alternatif lain terhadap produk yang digunakan berdasarkan *ingredients* yang terkandung dalam produk yang pernah dipakai sebelumnya. Penelitian ini menerapkan metode sistem rekomendasi *content-based filtering* dengan memanfaatkan konsep TF-IDF sebagai pembobotan kata dan perhitungan *cosine similarity* untuk mendapatkan rekomendasi *sunscreen* berdasarkan *ingredients* yang dimiliki. Berdasarkan hasil analisis dengan menggunakan sampel produk *sunscreen ANESSA Trial Set Brightening Gel* diperoleh 5 rekomendasi *sunscreen* serupa yakni *ANESSA Trial Gold Gel*, *New Brightening UV Sunscreen Gel*, *Twinpack New Perfect Sunscreen Skincare Gel*, *New Perfect UV Sunscreen Skincare Gel* dan *ANESSA Perfect Gentle UV Kit* yang memiliki nilai *cosine similarity* berturut-turut sebesar 0.90, 0.86, 0.85, 0.85, 0.50.

**Kata Kunci** : Rekomendasi, Kemiripan, *Sunscreen*, *Content-Based Filtering*.

## ABSTRACT

### **Sunscreen Product Recommendation System for Protection from Ultraviolet Rays using the Content-Based Filtering Method** (Case Study : Sunscreen Products on the *Sociolla* Website)

Khoirul Ni'am

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

*Global warming caused by extreme climates like now means that most people now have to be more aware of protecting their skin from exposure to UV rays, especially for people who live in tropical areas and are exposed to sunlight all year round. Excessive exposure to UV rays in the long term can cause various skin problems such as irritation, sunburn and even trigger skin cancer. Based on these problems, sunscreen products are presented as a solution to provide the skin's outer protection. When choosing sunscreen, people are often confused about choosing a product that suits their skin's needs and the benefits of the ingredients it contains. For this reason, a recommendation system has been developed to make it easier for the public to provide other alternatives to the products they use based on the ingredients contained in products that have been used previously. This research will try to apply a content-based filtering recommendation system method by utilizing the TF-IDF concept as word weighting and cosine similarity calculations to get sunscreen recommendations based on the ingredients available. Based on the results of analysis using samples of the ANESSA Trial Set Brightening Gel sunscreen product, 5 similar sunscreen recommendations were obtained, namely ANESSA Trial Gold Gel, New Brightening UV Sunscreen Gel, Twinpack New Perfect Sunscreen Skincare Gel, New Perfect UV Sunscreen Skincare Gel and ANESSA Perfect Gentle UV Kit has cosine similarity values of 0.90, 0.86, 0.85, 0.85, 0.50 respectively.*

**Keywords:** *Recommendations, Similarities, Sunscreen, Content-Based Filtering*

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Tinggal di Indonesia yang termasuk negara beriklim tropis memberikan manfaat bagi masyarakatnya untuk mendapatkan paparan sinar matahari sepanjang tahun, hal tersebut dipengaruhi letak geografis negara Indonesia yang berada pada garis khatulistiwa. Sinar matahari menjadi salah satu sumber energi yang mempunyai banyak manfaat bagi kehidupan manusia. Dalam paparan sinar matahari sendiri terdapat berbagai macam sinar yang dapat dilihat dengan mata maupun yang kasat mata. Indikasi sinar matahari dapat dilihat dengan mata secara langsung adalah sinar matahari yang memiliki pancaran 400nm, sedangkan sinar matahari yang kasat mata biasanya memiliki panjang gelombang antara 10nm sampai 400nm. Pancaran sinar matahari yang tidak dapat dilihat biasanya disebut sinar *ultraviolet* (Isfardiyana & Safitri, 2014).

Sinar UV sejatinya memiliki manfaat untuk manusia diantaranya paparan sinar UV dapat dimanfaatkan untuk mensintesa vitamin D dan juga dapat digunakan untuk membunuh bakteri apabila digunakan pada waktu yang tepat seperti pagi hari. Dampak lain sinar UV juga dapat membahayakan manusia apabila terpapar dengan waktu yang lama, kondisi ini biasanya ketika matahari sudah naik dengan waktu lebih dari jam 9 pagi. Penggolongan sinar UV dapat dilihat dari panjang gelombang yang dihasilkan diantaranya adalah sinar UV-A, UV-B dan UV-C. Sinar UV-A memiliki panjang gelombang 320-400nm dampaknya bagi kulit adalah dapat mempengaruhi kolagen pada tubuh, sinar UV-B panjang gelombangnya adalah 280-320nm apabila terpapar terlalu lama oleh sinar ini akan menyebabkan kerusakan pada kulit, sedangkan sinar UV-C memiliki panjang gelombang kurang dari 280nm dimana sinar UV-C sangat berbahaya karena pemaparan dalam jangka pendek dapat menyebabkan kemerahan hingga peradangan seperti iritasi pada kulit (Mukaromah A.H., 2008).

Salah satu upaya yang dapat dilakukan untuk melindungi kulit dari paparan sinar UV adalah menggunakan proteksi pada kulit dengan *sunscreen* atau *sunblock* agar kulit dapat terlindungi dari paparan sinar UV secara langsung ketika

beraktivitas diluar ruangan. *Sunscreen* atau yang biasa disebut tabir surya merupakan salah satu produk yang dapat digunakan untuk melindungi kulit dari kerusakan akibat paparan sinar UV. Produknya berbentuk dalam krim, gel, *lotion*, atau semprotan yang dioleskan atau disemprotkan ke bagian kulit sebelum melakukan aktivitas diluar ruangan. Penggunaan *sunscreen* sangat diperlukan sebagai langkah pertama untuk melindungi kulit dari paparan sinar UV secara langsung. Kerusakan kulit akibat paparan sinar UV terjadi dikarenakan pembentukan radikal bebas (*Reactive Oxygen Species/ROS*). Radiasi yang disebabkan oleh paparan sinar UV secara langsung dapat mengaktifkan bahan kimia dalam tubuh yang memicu kerusakan kulit, *sunburn*, eritema, iritasi dan kulit menjadi kering bahkan apabila pemaparan berlangsung lama atau berlebihan dapat menimbulkan perubahan hingga degenerasi pada kulit serta kanker kulit (Wadoe, et al., 2019).

Dalam pengaplikasian *sunscreen* untuk kulit konsumen harus mengetahui komposisi yang terkandung dalam produk yang dipilih dan juga menyesuaikan kondisi kulit yang dimiliki. Menurut (Ramlah, et al., 2021), beberapa kandungan *sunscreen* yang bagus diantaranya adalah *Salicylic Acid*, *Hyaluronic Acid*, *Mandelic Acid*, *Beta Hydroxy Acid (BHA)*, *Glycolic Acid*, *Alpha Arbutin* dan *Ceramid*. Masing-masing bahan tersebut memiliki fungsi dan kelebihan masing-masing. Faktor lain yang dapat dijadikan pertimbangan pemilihan produk *sunscreen* adalah usia dikarenakan sekarang juga banyak produk *sunscreen* yang dikhususkan untuk anak-anak karena kondisi kulit mereka yang masih sensitif. Hal tersebut membuat kebanyakan produk khusus anak-anak biasanya terbebas dari zat berbahaya seperti *Oxybenzone*, *Avobenzone*, *Retinyl Palmitate*, *Homsalate* dan lain sebagainya yang masih dirasa sensitif untuk kulit anak-anak (Shuib & Muniandy, 2018).

Hal pertama yang menjadi pertimbangan dalam pemilihan *sunscreen* yaitu dengan mengetahui jenis kulit masing-masing, sehingga komposisi yang terkandung dalam produk dapat bekerja secara maksimal pada jenis kulit yang tepat. Terdapat empat jenis kulit wajah yang umum dimiliki manusia yaitu kulit normal, kulit kering, kulit berminyak, dan kulit kombinasi antara kering dan berminyak. Berdasarkan perbedaan kulit tersebut dapat menimbulkan berbagai masalah kulit wajah yang berbeda, seperti kulit kusam, berjerawat, flek,

berkomedo, dan lain sebagainya. Pada kulit kusam, *niacinamide* yang terkandung dalam *sunscreen* dapat menjadi solusi untuk mengatasi permasalahan kulit tersebut. Hal ini karena *niacinamide* dapat mengganggu transmisi melanosom dengan menghambat transfer melanosom dari melanosit ke keratinosit, sehingga dapat mencerahkan kulit yang kusam (Pratiwi, 2022).

Masih banyak sekali orang yang belum sadar pentingnya melindungi kulit dari sinar UV bahkan sering mengabaikan pengaplikasian *sunscreen* maupun *sunblock* ketika melakukan aktivitas diluar ruangan. Negara beriklim tropis seperti di Indonesia perlindungan dari sinar UV sangat penting untuk dikenalkan atau diedukasikan sejak dini karena dari level anak-anak hingga dewasa tidak terkecuali dapat terpapar radiasi tersebut. Sekarang juga sudah banyak produk-produk *sunscreen* yang beredar sesuai kebutuhan bahkan ada juga produk khusus anak-anak dengan beberapa kandungan yang tidak sensitif bagi kulit mereka. Hal ini tentunya menjadi pertimbangan bagi para orang tua maupun masyarakat untuk lebih sadar akan bahaya paparan sinar UV yang dapat menyerang siapa saja apalagi ketika musim kemarau dengan cuaca dan suhu yang sangat panas. Sehingga diperlukan adanya sistem rekomendasi produk-produk *sunscreen* bagi para konsumen yang masih bingung dalam menentukan keputusan pembelian (Mumtazah, et al., 2020).

Sistem rekomendasi dapat diartikan sebagai alat atau teknik yang digunakan untuk menyediakan saran mengenai suatu hal untuk dimanfaatkan bagi *user* yang sedang mengalami kebingungan dalam mengambil keputusan. Sistem rekomendasi sudah banyak diterapkan oleh berbagai perusahaan besar untuk merekomendasikan produk mereka kepada konsusmen yang membutuhkan bantuan penentuan keputusan pembelian dengan menggunakan beberapa metode atau algoritma terkait, konsumen dapat mendapatkan rekomendasi produk sesuai dengan kemiripan produk yang ingin dibeli. Contohnya dapat diterapkan pada rekomendasi film, musik, buku, berita maupun produk-produk berupa barang dan layanan jasa. Terdapat beberapa metode dalam sistem rekomendasi yaitu *collaborative filtering*, *content-based filtering*, *knowledge-based*, *contetnt based recommender* dan *hybrid filtering*. Metode *Content-based filtering* sendiri memanfaatkan informasi yang terdapat di



beberapa item atau data untuk dijadikan sebagai parameter yang menentukan rekomendasi yang sesuai untuk *user* (Nastiti, 2019).

Berdasarkan permasalahan tersebut, peneliti akan melakukan penelitian tentang sistem rekomendasi *sunscreen* berdasarkan komposisi produk dengan menggunakan pendekatan *content-based filtering*, dengan harapan penelitian ini dapat membantu masyarakat dalam menentukan *sunscreen* yang akan digunakan sesuai dengan komposisi dan fungsi yang dimiliki pada produk *sunscreen*.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang tersebut, maka rumusan masalah yang terkait dengan penelitian ini adalah sebagai berikut.

1. Bagaimana deskripsi dari data produk *sunscreen* berdasarkan produk yang dipasarkan pada *website Sociolla* ?
2. Bagaimana merekomendasikan produk *sunscreen* berdasarkan komposisinya menggunakan *Content-Based Filtering* ?
3. Bagaimana hasil rekomendasi produk *sunscreen* berdasarkan komposisinya yang diberikan oleh sistem rekomendasi menggunakan *Content-Based Filtering* ?

## **1.3. Batasan Masalah**

Dalam penelitian ini, penulis membatasi masalah sebagai berikut.

1. Data yang digunakan adalah sekunder hasil *scraping* melalui *website Sociolla* Indonesia pada tahun 2023, yang memiliki banyak katalog mengenai produk-produk perawatan kulit.
2. Data diolah dengan menggunakan bahasa pemrograman *Python* dengan metode analisis yang digunakan yaitu sistem rekomendasi dengan menggunakan pendekatan *content-based filtering*.

## **1.4. Tujuan Penelitian**

Berdasarkan rumusan masalah sebelumnya, penulis memiliki tujuan penelitian sebagai berikut.

1. Mengetahui deskripsi dari data produk *sunscreen* berdasarkan produk yang dipasarkan pada *website Sociolla*.

2. Mengetahui cara merekomendasikan produk *sunscreen* berdasarkan komposisinya menggunakan *Content-Based Filtering*.
3. Mengetahui hasil rekomendasi produk *sunscreen* berdasarkan komposisinya yang diberikan oleh sistem rekomendasi menggunakan *Content-Based Filtering*.

### **1.5. Manfaat Penelitian**

Manfaat yang diharapkan penulis dalam penelitian ini adalah sebagai berikut.

1. Dapat membantu masyarakat untuk lebih sadar akan bahaya dari pemaparan sinar UV secara langsung.
2. Memberikan rekomendasi produk *sunscreen* kepada masyarakat untuk melindungi kulit berdasarkan komposisi bahan penyusunnya.
3. Memberikan pengetahuan bagaimana cara kerja sistem rekomendasi dengan metode *content-based filtering*.
4. Diharapkan dapat menjadi referensi untuk penelitian selanjutnya maupun penelitian yang berhubungan dengan sistem rekomendasi.

## BAB II

### TINJAUAN PUSTAKA

Dalam penelitian yang penulis akan lakukan membutuhkan banyak sekali pustaka yang dijadikan sebagai acuan, referensi hingga arahan dalam menyelesaikan penulisan tugas akhir ini. Oleh karenanya, penulis mencari beberapa referensi dari jurnal yang sudah terakreditasi dan diakui secara nasional oleh para cendekia dibidangnya. Kebanyakan pustaka yang digunakan memiliki hubungan erat dari kesamaan metode yang penulis gunakan yakni dengan pendekatan *Content-Based Filtering*, dimana nantinya akan didapatkan sebuah data yang memiliki persamaan dalam nilai *cosine similarity* sebagai acuan untuk menjadikan data tersebut layak untuk direkomendasikan kepada *user*. Harapannya dapat memberikan warna baru dalam bidang ilmu pengetahuan terutamanya untuk topik sistem rekomendasi. Berikut merupakan beberapa pustaka yang digunakan penulis sebagai acuan pengerjaan tugas akhir:

#### 2.1. Penelitian Terdahulu

**Tabel 2.1** Tabel Penelitian Sebelumnya

Tahun	Nama	Judul	Hasil Penelitian
2023	Anak Agung Aditya Nugraha, Ngurah Agus Sanjaya ER	Penyusunan Sistem Rekomendasi Produk <i>Diecast</i> Mobil dengan Metode <i>Content-Based Filtering</i> (CBF)	Semakin berkembangnya teknologi sekarang membuat mobil <i>diecast</i> semakin populer dan menciptakan banyak permintaan, sehingga perlu adanya sistem rekomendasi yang membantu para <i>user</i> untuk menemukan produk yang sesuai. Penelitian ini akan berfokus menggunakan metode <i>CBF</i> dengan penggabungan Teknik TF-IDF dan <i>cosine similarity</i> . Dengan menganalisis fitur tekstual produk mobil <i>diecast</i> seperti <i>brand</i> , model dan spesifikasi. Sistem <i>CBF</i> akan menghasilkan hasil yang dipersonalisasi dalam memberikan rekomendasi yang akurat dan relevan yang dapat

Tahun	Nama	Judul	Hasil Penelitian
			meningkatkan pengalaman pengguna dan memfasilitasi eksplorasi pasar mobil <i>diecast</i> .
2021	Fatoni Batari Agung Larasati, Herny Februariyanti	Sistem Rekomendasi <i>Product Emina Cosmetics</i> dengan Menggunakan Metode <i>Content-Based Filtering</i>	<i>Emina Cosmetics</i> adalah produk komestik untuk pengguna remaja-dewasa dengan bertambahnya konsumen membuat produk yang diproduksi juga menjadi semakin beragam dengan berbagai macam variasi dan kegunaannya. Hal tersebut juga membuat beberapa konsumen menjadi bingung untuk memilih produk mana yang cocok untuk dibeli dan digunakan sehingga perlu adanya sistem rekomendasi yang mengarahkan para konsumen ini untuk menentukan produk yang sesuai dengan kebutuhan dan permasalahan kulit mereka. Penelitian ini digunakan metode <i>content-based filtering</i> dengan menghitung kemiripan antar produk menggunakan <i>cosine similarity</i> , Sehingga didapatkan hasil produk dengan similaritas tertinggi dengan nilai sebesar 0.7195.
2020	Rahman Indra Kesuma, Amirul Iqbal	Penerapan <i>Content-Boosted Collaborative Filtering</i> untuk Meningkatkan Kemampuan Sistem Rekomendasi Penyedia Jasa Acara Pernikahan	Perubahan pola hidup masyarakat <i>modern</i> yang semakin maju memerlukan beragam kebutuhan yang membuat munculnya berbagai jenis usaha maupun ide bisnis dibidang jasa terutamanya untuk menangani acara pernikahan. Banyak sekali penyedia jasa acara pernikahan yang bermunculan dengan berbagai keunggulan masing-masing sehingga menjadikan perlunya sistem rekomendasi yang dapat membantu pemilihan jasa tersebut sesuai dengan konsep

Tahun	Nama	Judul	Hasil Penelitian
			yang diinginkan. Dengan memanfaatkan metode <i>CBCF</i> untuk meningkatkan sistem rekomendasi dalam pemberian rekomendasi untuk jasa acara pernikahan. Penelitian ini diperoleh metode <i>CBCF</i> yang bisa mempertahankan kualitas akurasi rekomendasi untuk pengguna lama sebesar 85% dan pengguna baru sebesar 54%.
2021	Dwi Ayu Nur Safitri, Risa Helilintar, Lilia Sinta Wahyuniar	Sistem Rekomendasi <i>Skincare</i> Menggunakan Metode <i>Content-Based Filtering</i> dan Algoritma <i>Apriori</i>	Dalam pemilihan produk <i>skincare</i> masih banyak konsumen yang mengalami kebingungan mengenai produk dan <i>brand</i> yang cocok untuk digunakan dalam memenuhi kebutuhan kulitnya. Hal tersebut terjadi karena pengetahuan yang masih minim mengenai produk-produk <i>skincare</i> yang beredar di Indonesia. Penelitian ini merancang sebuah aplikasi yang dapat memberikan rekomendasi <i>skincare</i> kepada konsumen berdasarkan produk yang disukai sebelumnya, Dengan menggunakan metode <i>content-based filtering</i> dan algoritma <i>apriori</i> menghasilkan produk dengan <i>rating</i> tertinggi hingga terendah dan nilai <i>support</i> serta nilai minimum <i>confidence</i> untuk menentukan produk yang di rekomendasikan. Berdasarkan 40 data produk didapatkan hasil <i>rating</i> tertinggi sebesar 0.447 dan menggunakan minimum nilai <i>support</i> dan <i>confidence</i> sebesar 40% diperoleh nilai <i>confidence</i> sebesar 88.89%.
2022	Maria Sri Wulandari, Rahayu	Prediksi Jenis Kulit Wajah Menggunakan	Beberapa <i>user</i> memiliki masalah kebingungan dalam menentukan pembelian suatu

<b>Tahun</b>	<b>Nama</b>	<b>Judul</b>	<b>Hasil Penelitian</b>
	Noveandini dan Soegijanto	Metode <i>Certainty Factor</i> Sebagai Sistem Rekomendasi Produk Berbasis <i>Web</i>	produk, sehingga diberikan suatu sistem rekomendasi untuk memudahkan <i>user</i> dalam pemilihan aksesoris <i>smarthphone</i> . Sistem yang dibuat akan menghitung nilai <i>similarity</i> dan rating tiap <i>item</i> dengan persamaan <i>weighted average of deviation</i> . Hasilnya nanti akan diurutkan dari nilai terbesar dan terkecil sebelum nantinya direkomendasikan kepada <i>user</i> , hasil menunjukkan nilai rata-rata <i>MAE</i> 0.572039 tetapi untuk proses eksekusinya memerlukan waktu yang lama yakni 6.4 detik.
2020	Mariani Widia Putri, Achmad Muchayan, Made Kamusutara	Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode <i>Content-Based Filtering</i> dan TF-IDF	Sitem rekomendasi sekarang menjadi suatu hal yang banyak dibutuhkan masyarakat dikarenakan transaksi secara <i>online</i> yang semakin menjamur dikalangan masyarakat Indonesia. Penelitian ini penulis membangun sistem rekomendasi produk dengan pendekatan metode <i>CBF</i> dan TF-IDF dari model <i>Information Retrieval (IR)</i> . Tujuannya untuk meningkatkan <i>brand awareness</i> pelanggan dan meminimalisir gagalnya transaksi karena kurangnya informasi produk yang diterima pembeli. Data yang digunakan dalam penelitian ini ada sejumlah 258 produk yang terdiri dari 8 kategori dan 33 kata kunci pembentuk sesuai dengan <i>product knowledge</i> yang dimiliki. Didapatkan hasil nilai TF-IDF sebesar 13.854 dengan menampilkan rekomendasi produk terbaik pertama dan nilai keakuratan
<b>Tahun</b>	<b>Nama</b>	<b>Judul</b>	<b>Hasil Penelitian</b>

			sebesar 95,6% dalam memberikan rekomendasi produk pena.
--	--	--	---

Beberapa literatur yang digunakan penulis memiliki kesamaan pada metode yang digunakan tetapi untuk perbedaannya sendiri dapat dilihat pada studi kasus yang diteliti. Penelitian yang akan dilakukan penulis ini berfokus pada studi kasus data produk *sunscreen* yang ada pada *web Sociolla*. Data tersebut akan diolah dan dengan melihat dari komposisinya untuk nanti dapat direkomendasikan kepada *user* yang memiliki permasalahan kulit maupun kebutuhan sesuai dengan kandungan yang ada dalam produk *sunscreen*.

## **BAB III**

### **LANDASAN TEORI**

#### **3.1. Tabir Surya (*Sunscreen*)**

Tabir surya atau yang lebih familiar disebut dengan *sunscreen* adalah suatu zat atau material yang berbentuk seperti *lotion*, krim, *gel* atau *spray* yang diaplikasikan pada bagian kulit sebagai sebuah perlindungan terluar dari radiasi sinar UV. *Sunscreen* merupakan suatu produk yang sangat penting untuk digunakan oleh masyarakat yang hidup di wilayah tropis seperti di Indonesia, hal tersebut di dasari karena negara Indonesia mendapatkan pancaran sinar matahari hampir setiap hari dengan suhu yang tinggi bahkan dalam kategori ekstrem. Kenyataan yang terjadi masih sekali banyak orang yang mengabaikan penggunaan *sunscreen* ketika melakukan aktivitas diluar ruangan. Kebanyakan orang-orang tersebut masih belum mengerti untuk apa pentingnya perlindungan kulit dengan *sunscreen* serta pemilihan produk *sunscreen* yang sesuai dengan tipe kulit yang dimiliki. Kandungan komposisi penyusun produk *sunscreen* juga masih belum dilihat secara jeli oleh masyarakat Indonesia hal tersebut yang cenderung membuat kebanyakan konsumen hanya membeli berdasarkan sentimen produk bukan dari suatu seleksi yang sudah disesuaikan dengan kebutuhan kulit. Beberapa kandungan penting dalam *sunscreen* diantaranya adalah SPF (*Sun Protecting Factor*) sebagai salah satu indikator yang digunakan untuk memilih produk yang aman (Asmiati, et al., 2021).

##### **3.1.1. *Sun Protecting Factor* (SPF)**

Dalam pengertiannya *Sun Protecting Factor* (SPF) adalah sebuah indikator pengukuran yang digunakan untuk melihat keefektivitasan tingkat perlindungan suatu produk *sunscreen*. Nilai yang didapatkan untuk setiap SPF berada pada kisaran angka 2-60 yang digunakan untuk menunjukkan ketahanan produk dalam melindungi kulit dari paparan sinar UV. Semakin tinggi angka yang tertera pada setiap produk *sunscreen* maka perlindungan yang diberikan juga akan semakin efektif. SPF 15 memiliki manfaat dapat menyaring 93% paparan sinar UV-B, SPF 30 dapat menghalangi 97% sinar UV, sementara SPF 50 memiliki tingkat perlindungan tertinggi dengan indikasi 98% dapat menghalangi paparan sinar UV. Memahami setiap indikasi SPF pada produk *sunscreen* yang digunakan akan dapat



membantu para konsumen untuk memilih produk yang baik. Sangat penting memberikan edukasi mengenai SPF kepada setiap konsumen yang ingin membeli *sunscreen*, hal itu bertujuan agar konsumen lebih teliti dalam memilih produk dan tidak hanya melihat dari segi harga maupun merk tetapi fungsi serta keunggulan produk yang ditawarkan jauh lebih penting untuk dijadikan faktor pengambilan keputusan pembelian (Widolaras & Ikhsanto, 2022).



**Gambar 3. 1** SPF Rating dan UV Protection (Sumber: Health Gennie)

### 3.1.2 Protection Grade of UV-A (PA)

PA (*Protection Grade of UV-A*) adalah suatu sistem indikator yang dijadikan penilain untuk pengembangan tingkat perlindungan produk *sunscreen* dari sinar UV-A. Sinar UV-A sendiri memiliki dampak negatif bagi kulit manusia karena dapat menyebabkan kerusakan kulit, penuaan dini hingga resiko kanker kulit. Sistem penilaian PA biasanya terdapat penambahan simbol (+) yang bervariasi dari 1-4 simbol (+). Arti simbol tersebut adalah semakin banyak simbol (+) yang tertera pada produk maka tingkat perlindungan sinar UV-A yang didapatkan juga semakin tinggi (Rahmawati & Solandjari, 2019).

### 3.1.3 Jenis Kulit

Kulit adalah bagian terluar dari tubuh manusia yang memiliki fungsi untuk melindungi dari berbagai ancaman yang berpotensi melukai tubuh bagian dalam. Pada umumnya jenis kulit dibagi menjadi empat kategori, kulit kering, kulit normal, kulit berminyak dan kulit kombinasi. Kulit normal dapat diartikan dengan kondisi

kulit yang seimbang dan sehat serta memiliki pori-pori yang tidak terlalu besar. Kulit kering cenderung tidak bisa mempertahankan kelembapan sehingga menjadikannya terlihat kering dengan struktur kasar sampai bisa mengelupas. Kulit berminyak ditandai dengan kondisi kelebihan produksi minyak pada tubuh dan kulit kombinasi dari kulit berminyak dan kulit kering atau normal (Kalangi, 2013).

#### **3.1.4 Bahaya Paparan Sinar UV**

Sinar UV yang terpancar dan terkena kulit dalam jangka waktu yang panjang memiliki dampak buruk. Dampak negatif yang dapat diakibatkan akibat pemaparan sinar UV secara berlebihan diantaranya adalah sinar UV dapat membuat kulit terbakar yang disebabkan paparan sinar UV-B, eritema yang ditandai dengan kemerahan pada kulit hingga bengkak, memicu pertumbuhan sel kanker karena terjadinya kerusakan fotokimia pada DNA dari sel-sel yang berada di dalam tubuh, kehilangan elastisitas akibat paparan sinar UV-A yang menembus bagian demis, kerutan pada kulit yang terjadi karena efek samping dari elastisitas kulit yang menurun (Isfardiyana & Safitri, 2014)

#### **3.1.5 Komposisi Sunscreen**

Komposisi *sunscreen* merupakan bahan-bahan yang terkandung dalam suatu produk yang mengisi dan membentuk produk tersebut hingga memiliki fungsi dan bisa dipasarkan. Produk *sunscreen* memiliki jenis-jenis bahan dengan keunggulan masing-masing. Diantaranya ada *Zinc Oxide* dan *Titanium Dioxide* yang cocok digunakan untuk melindungi dari sinar UV-A, *Octisalate* dan *Octocrylene* yang umumnya digunakan untuk melindungi dari sinar UV-B, vitamin E yang dapat membantu melindungi kulit dan menjaga kelembaban, *Niacinamide* (vitamin B3) yang berfungsi untuk mengurangi peradangan pada kulit dan *Hyaluronic Acid* yang dapat menjaga kelembaban kulit (Gwak, et al., 2021).

### **3.2. Sociolla**

*Sociolla* merupakan sebuah *e-commerce* yang memiliki nama usaha PT. Social Bella, dimana perusahaan ini bergerak untuk menyediakan berbagai macam produk kecantikan dan kesehatan mulai dari *sunscreen*, *bodycare*, *skincare* hingga pilihan produk-produk *make up*. *Sociolla* secara resmi berdiri dan beroperasi di Indonesia pada tahun 2015 oleh Christopher Madiam, Chrisanti Indiana dan John Marco Rasjid dengan pusat operasional di Kembangan, Jakarta Barat. Permulaan

perusahaan ini dapat berdiri karena melihat peluang melalui berkembangnya *tren* kecantikan di Indonesia dengan pengalaman belanja secara *online* maupun *offline*. Maka dari itu *Sociolla* hadir dan memberikan penawaran dengan berkolaborasi dengan berbagai distribusi nasional dan pemilik merk resmi di Indonesia untuk menghadirkan lebih dari 5000 pilihan produk yang original dan bersertifikasi BPOM (Badan Pengawasan Obat dan Makanan).

*Sociolla* memiliki salah satu fitur andalan yakni *SOCO* dan *Beauty Journal* yang menawarkan pengalaman belanja menarik dengan menambahkan keterangan *review* produk untuk dijadikan salah satu pertimbangan konsumen. Dengan fitur tersebut konsumen diharapkan dapat lebih paham karena dapat melihat pengalaman pemakaian oleh konsumen sebelumnya, disisi lain hal ini menjadi hal menarik karena produk-produk yang di *review* juga akan memiliki rating bintang 1-5 sebagai indikator tingkat kualitas produk yang di *review*. *Website Sociolla* juga terdapat produk-produk dengan berbagai pilihan dari produk dalam negeri dan luar negeri yang dipasarkan via *e-commerce* dan toko *offline* yang dapat diakses langsung oleh para konsumennya diseluruh Indonesia (Rizky & Noor, 2020).

*sociolla*

**Gambar 3. 2** Logo *Sociolla* (Sumber: *Sociolla.com*)

### **3.3. Analisa Deskriptif**

Metode analisa deskriptif pada dasarnya merupakan bagian dari ilmu statistik yang mempelajari bagaimana cara pengumpulan, peringkasan, penyederhanaan dan penyajian data mentah untuk diubah menjadi suatu bentuk informasi yang mudah dipahami dan bermanfaat bagi orang lain. Statistika deskriptif juga memiliki arti mendeskripsikan suatu data untuk memberikan gambaran umum dari suatu penelitian maupun objek yang sedang diteliti. Penggunaan metode analisis deskriptif dapat dicirikan dengan menampilkan informasi berupa grafik, tabel dan ukuran statistik seperti *mean*, variansi, korelasi dan presentase dari suatu data (Goos & Meintrup., 2015).

Penelitian ini sendiri akan menggunakan analisis deskriptif berupa penggambaran data dengan menggunakan bantuan berupa visualisasi grafik *barchart*. Tujuannya untuk menjelaskan mengenai beberapa produk-produk yang berhasil di *input* dari *web Sociolla* dan juga penggolongan produk *sunscreen* berdasarkan merk atau *brand*. Komposisi yang sering muncul dalam produk *sunscreen* dapat dilihat dengan menambahkan visualisasi dengan *wordcloud*.

### **3.4. Text Mining**

*Text mining* atau yang dikenal juga sebagai analisis kata atau dokumen untuk mengumpulkan informasi melalui identifikasi dan eksplorasi pola pada sumber data berupa data tekstual. Pengaplikasiannya *text mining* memanfaatkan dokumen sebagai elemen dasar dan kumpulan dokumen menjadi sebuah *corpus*. Tujuan penggunaan *text mining* dalam analisis data adalah untuk mendapatkan informasi dengan melakukan analisis teks, klasifikasi, visualisasi, *clustering*, *machine learning* dan *data mining*. Perbedaan antara *text mining* dan *data mining* terletak pada sumber data yang digunakan, dimana *data mining* mengekstraksi pola dari data sedangkan *text mining* mengesktraksi pola dari data tekstual (Hidayatullah & Ma'arif, 2016).

Metode *text mining* terdapat tahapan *text preprocessing* yang merupakan tahapan pengolahan data dengan cara mengubah data yang semula tidak terstruktur menjadi satuan data yang rapi. Hasil dari *text preprocessing* dapat dimanfaatkan untuk algoritma *text mining* untuk nantinya dapat dianalisis dengan menampilkan hasil akhir berupa informasi dari data tekstual (Nugraha & Sebastian, 2018). Beberapa tahapan yang harus dilakukan *text preprocessing* dalam penelitian ini adalah *case folding*, *remove punctuation*, *remove double white space*, *stopword removal*, *lemmatization* dan *tokenizing*.

#### **3.4.1 Case Folding**

*Case folding* merupakan proses analisis data tekstual dengan mengubah huruf dalam teks menjadi format yang konsisten dengan mengkonversi seluruh huruf besar menjadi kecil maupun sebaliknya. Sebagai contoh karakter huruf 'A', 'B', 'C' diubah menjadi 'a', 'b', 'c'. Tahapan *case folding* ini bertujuan untuk menstandarisasi karakter pada keseluruhan data tekstual (Alita & Rahman, 2020).

### **3.4.2 Remove Punctuation**

*Remove punctuation* adalah tahapan yang ada pada proses *text preprocessing* dengan tujuan menghilangkan tanda baca seperti tanda titik, koma, seru, tanya dan lain sebagainya. Hal tersebut dilakukan dengan tujuan karena tanda baca tidak dibutuhkan selama analisis dan supaya data tekstual yang digunakan dapat lebih terbaca dengan detail (Henni Natalia Sijabat, 2023).

### **3.4.3 Remove Double White Space**

*Remove double white space* adalah proses penghapusan atau menggantikan dua spasi yang terdapat pada data hasil *scraping* untuk menjadikan dua spasi tersebut menjadi spasi tunggal. Hal tersebut perlu dilakukan selama proses *text preprocessing* untuk menjaga konsistensi pada data tekstual agar tetap rapi dan mudah untuk analisis (Fikriyah & Sibaroni, 2022).

### **3.4.4 Stopword Removal**

*Stopword removal* merupakan proses penghilangan kata-kata yang tidak memiliki peranan penting pada data tekstual yang digunakan. Beberapa contoh kata yang sering dihilangkan adalah “*are*”, “*she*”, “*i*”, “*the*”, “*you*”, “*are*”, “*with*”, “*is*” dan lain sebagainya. *Stopword* juga dapat dijumpai dalam bahasa Indonesia contohnya terdapat pada kata “*yang*”, “*di*”, “*ke*”, “*dari*”, “*ini*”, “*oleh*”, “*saya*”, dan lainnya. Penghapusan *stopword* berguna dalam membersihkan teks, mengurangi dimensi data dan meningkatkan efisiensi dalam *text preprocessing* (Kaur & Preetpal, 2018).

### **3.4.5 Lemmatization**

*Lemmatization* merupakan proses *text preprocessing* untuk mengubah data tekstual menjadi bentuk dasar dari sebuah kata. Dalam pengertian lainnya *lemma* artinya adalah bentuk kata dasar dari suatu kata yang menggambarkan bentuk kata tersebut dalam kamus. *Lemma* memiliki memiliki fungsi serupa dengan *stemming*, hanya saja hasil *output* yang ditampilkan lebih baik karena alur *lemma* lebih kompleks dengan melibatkan kamus dalam dalam pencarian kata dasar (*root*) (Bargamanis & Goldwater, 2018).

### 3.4.6 *Tokenizing*

Tahapan akhir dalam proses *text processing* data adalah *tokenizing*, *tokenizing* sendiri memiliki pengertian proses pembagian teks menjadi suatu unit-unit terkecil atau sekumpulan kata yang disebut *token*. Adanya *tokenizing* diharapkan peneliti dapat memperoleh potongan kata yang memiliki makna penting sebagai suatu entitas yang memiliki arti dalam penyusunan matriks dokumen pada saat *text preprocessing* berlangsung. Pengertian lainnya, sebelum data tekstual diproses lebih lanjut maka data tersebut harus disegmentasi terlebih dahulu dari yang semula kalimat menjadi penggalan kata-kata yang disebut sebagai *token* (Abbe, et al., 2015).

### 3.5. Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF)

Tahapan setelah data berhasil melewati proses *text preprocessing* adalah menambahkan nilai pembobotan dengan *Terms Frequency-Inverse Document Frequency* (TF-IDF). Sebelum melakukan pembobotan syarat utama yaitu data yang digunakan harus berbentuk numerik terlebih dahulu. Pengertian *Terms Frequency-Inverse Document Frequency* (TF-IDF) merupakan metode analisis yang berguna untuk mengevaluasi seberapa penting sebuah kata di dalam sebuah dokumen atau sekelompok kalimat. Dalam hal dokumen tunggal setiap kalimat yang ada dianggap sebagai sebuah dokumen terpisah (Putri, et al., 2020).

Berdasarkan beberapa pendekatan penelitian sebelumnya, algoritma TF-IDF dikenal memiliki kelebihan sangat efisien dalam menghitung tingkat kemiripan antar dokumen dan kata kunci yang tertera, karena metode TF-IDF hanya berfokus pada perhitungan kemiripan satu item dengan yang lain. Dalam pengertian lainnya TF-IDF juga digunakan untuk menentukan tingkat korelasi kata (*terms*) terhadap suatu dokumen dengan pemberian bobot pada setiap kata. TF-IDF sendiri berdiri dengan cara menggabungkan dua konsep berupa frekuensi kemunculan kata dalam dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut. Berikut merupakan rumus perhitungan TF-IDF (Herwijayanti, et al., 2018), Perhitungan *Terms Frequency* (TF) dengan persamaan berikut ini.

$$TF = TF_{ij} \quad (3.1)$$

Keterangan:

TF : *Terms Fruquency*

$TF_{ij}$  : Banyak kemunculan term  $T_j$  dalam dokumen ke-i.

TF adalah *terms frequency* dan  $TF_{ij}$  merupakan banyaknya kemunculan term  $T_j$  dalam dokumen  $D_i$ . Maka nilai *terms frequency* dapat diperoleh dengan menghitung banyaknya term  $T_j$  dalam dokumen  $D_j$  (Moh. Afif Rofiqi, et al., Desember 2019)

Perhitungan *Inverse Document Frequency* (IDF) dengan rumus berikut.

$$IDF_t = \ln\left(\frac{D}{DF_t}\right) \quad (3.2)$$

Keterangan:

$IDF_t$  : *Inverse Document Frequency* (kata ke-t)

D : Jumlah Dokumen

$DF_t$  : Jumlah frekuensi dokumen yang mengandung term (kata ke-t)

IDF memiliki fungsi untuk menangani kata yang sering muncul pada setiap dokumen yang dianggap sebagai kata yang tidak memiliki makna. IDF juga dibuat untuk mengurangi pengaruh kata-kata yang terlalu sering muncul dalam sebuah dokumen, dengan kata lain semakin sering sebuah kata muncul pada suatu dokumen maka nilai pembobotnya juga akan semakin kecil juga (Adi, 2010). Perhitungan *Term Frequency-Inverse Document Frequency* (TF-IDF) menggunakan rumus sebagai berikut.

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

$$W_{i,j} = TF_{i,j} \times \log\left(\frac{D}{DF_j}\right) \quad (3.4)$$

Keterangan:

$W_{ij}$  : Bobot TF-IDF (dokumen ke-i terhadap kata ke-j)

$TF_{ij}$  : Frekuensi kemunculan kata ke-i dalam dokumen

$IDF_j$  : *Inverse Document Frequency* (kata ke-j)

D : Jumlah Dokumen

$DF_j$  : Jumlah frekuensi dokumen yang mengandung *term* (kata ke-j)

Pada **persamaan 3.4**, berapapun nilai yang diperoleh dari  $TF_{ij}$  apabila hasil dari  $D = DF_j$  diperoleh hasil 0 untuk perhitungan IDF, maka perlu ditambahkan nilai 1 pada sisi IDF dan formula perhitungan perhitungannya. Sehingga rumusnya akan berubah menjadi berikut.

$$W_{i,j} = TF_{i,j} \times \left( \ln \left( \frac{D}{DF_j} \right) + 1 \right) \quad (3.5)$$

### 3.6. Cosine Similarity

*Cosine similarity* merupakan suatu teknik analisis yang berguna untuk mengukur sejauh mana dua dokumen atau teks memiliki kemiripan. Pengukuran ini dilakukan dengan cara menghitung nilai *cosinus* antara dua vektor dalam ruang vektor. Metode ini memperhitungkan kesamaan arah vektor yang dapat diketahui dengan melihat kedekatan antar sudut yang mengartikan kemiripan kedua dokumen tersebut. Teknik analisa ini sering diterapkan dalam konteks pemrosesan teks dengan vektor yang merepresentasikan nilai numerik dari suatu dokumen atau data (Larasati & Februariyanti, 2021). Oleh karenanya, didapatkan rumus untuk perhitungan nilai *cosine similarity* sebagai berikut.

$$\cos \alpha = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n (A_i B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (3.6)$$

Keterangan:

A : Vektor A, yang dibandingkan kemiripannya

B : Vektor B, yang dibandingkan kemiripannya

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

$\|A\|$  : Panjang vektor A

$\|B\|$  : Panjang vektor B

$A_i$  : Bobot *term* dalam blok  $A_i$

$B_i$  : Bobot *term* dalam blok  $B_i$

$i$  : Jumlah *term* dalam kalimat

$n$  : jumlah vektor

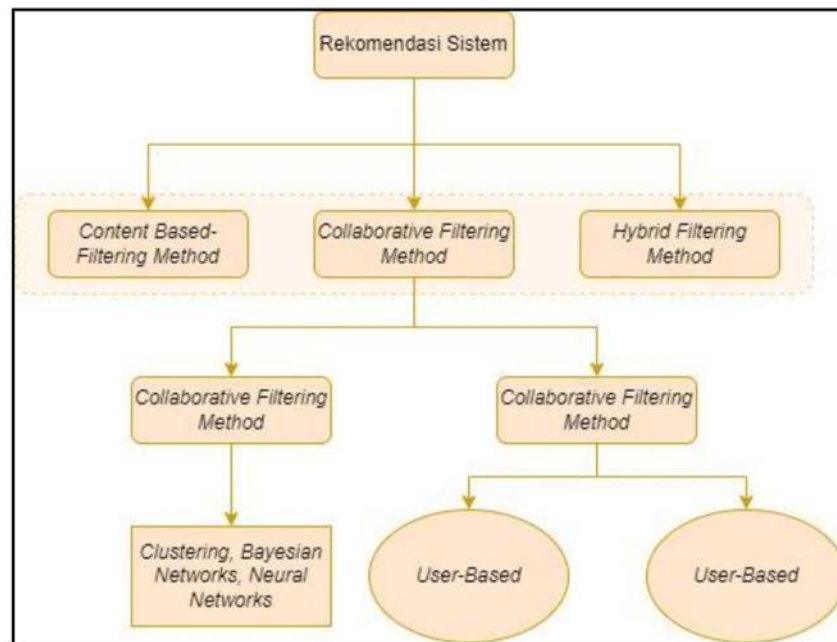
### 3.7. Sistem Rekomendasi

Sistem rekomendasi adalah suatu metode yang berhasil dikembangkan berdasarkan perkembangan ilmu *artificial intelligence* dibidang *machine learning*. Sistem ini memiliki cara kerja untuk membuat suatu rangkaian program yang nantinya bisa memberikan rekomendasi produk atau layanan jasa kepada *user* sesuai dengan kebutuhan. Sistem ini juga bisa dibuat dengan mengambil data berdasarkan kebiasaan *user* sebelumnya untuk nantinya dapat menghasilkan *output*



berupa informasi atau produk yang sesuai dengan kemiripan penggunaan sebelumnya (Nugraha & Sanjaya, 2023).

Tujuan dibuatnya sistem rekomendasi yakni dapat memudahkan sekaligus membantu dalam pengambilan keputusan bagi orang-orang yang masih merasa bingung untuk menentukan produk atau layanan mana yang harus digunakan. Beberapa hal yang bisa diterapkan dalam pengaplikasian sistem rekomendasi diantaranya adalah rekomendasi buku, film, handphone, laptop, barbershop, *coffee shop* dan lain sebagainya. Hal-hal tersebut dapat dibuat dalam suatu sistem rekomendasi dengan mengandalkan kemiripan pada data, sehingga nantinya akan menghasilkan rekomendasi sesuai dengan skor tertinggi sesuai dengan kemiripan dari data pengguna sebelumnya (Huda, et al., 2022).



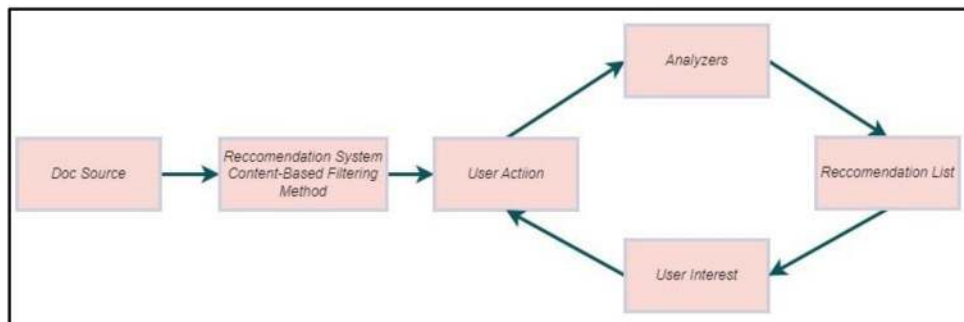
**Gambar 3. 3** Metode Sistem Rekomendasi (Sumber: (Isinkaye, 2015))

### 3.8. *Content Based Filtering*

*Content-based filtering* merupakan suatu metode pengembangan sistem rekomendasi yang digunakan untuk memberikan rekomendasi produk atau layanan jasa kepada pengguna berdasarkan karakteristik atau kesamaan *content* sebelumnya. Contoh penerapannya dilakukan pada suatu *item* dokumen atau data teks dengan atribut pembentuknya berupa kata-kata yang memiliki kesamaan untuk nantinya dapat diberikan pembobot sesuai dengan produk atau layanan jasa yang digunakan *user* sebelumnya (Putri, et al., 2020). Kelebihan penggunaan metode

*content-based filtering* adalah pengguna akan mendapatkan alasan mengapa suatu *item* dapat dijadikan rekomendasi berdasarkan konteks relevansi dan representatif. Kelemahan yang dimiliki adalah hanya mampu merekomendasikan suatu *item* berdasarkan kemiripan kata kunci. *Content-based filtering* masih belum bisa memberikan rekomendasi berdasarkan permasalahan yang lebih kompleks seperti rekomendasi berdasarkan jenis atribut pada data (Nastiti, 2019).

Metode *content-based filtering* biasanya digabungkan dengan konsep perhitungan TF-IDF untuk menentukan kepentingan relatif dari sebuah dokumen, artikel, atau item berita yang digunakan dalam penentuan rekomendasi. Setelah perhitungan TF-IDF maka peneliti dapat melanjutkan dengan pemberian *cosine similarity* untuk menentukan *item* yang memiliki kemiripan dengan produk yang sebelumnya sudah menjadi preferensi pengguna. Berikut ilustrasi cara kerja sistem rekomendasi berdasarkan metode *content-based filtering*



**Gambar 3. 4** Ilustrasi Sistem Rekomendasi (Sumber: Dokumen Pribadi)

Berdasarkan **Gambar 3.4** dapat dijelaskan sebagai berikut.

1. *Doc source* merupakan bagian awal dari sistem rekomendasi yang memuat sumber dokumen. Dalam penelitian ini sumber dokumen yang digunakan diambil dari *website Sociolla* dengan memunculkan variabel deskripsi (komposisi) pada produk.
2. Dokumen yang digunakan sistem akan memproses dokumen tersebut dengan metode *content based-filtering* untuk membuat model rekomendasi berdasarkan kemiripan antar deskripsi (komposisi) pada produk berdasarkan relevansi dan preferensi *user*.
3. Setelah itu akan ada aksi atau tindakan dari *user* untuk memasukkan kata kunci dari produk yang di inginkan maka sistem akan bekerja dengan mempelajari pola-pola yang ada.

4. Kemudian ketika sistem sudah melakukan analisis, maka sistem akan memunculkan *recommendation list* 5 produk teratas berdasarkan kemiripan deskripsi (komposisi) produk yang diperoleh sebelumnya.
5. Dan *user interest* merupakan bagian ketertarikan pengguna berdasarkan hasil rekomendasi yang diberikan oleh sistem apakah sudah sesuai atau belum. Jika belum maka pengguna dapat memasukkan kata kunci yang lain dan sistem akan memproses kembali pada langkah ketiga.

### **3.9 *Phyton Anywhere***

*Phyton anywhere* adalah sebuah platform berbasis *web* yang digunakan untuk menulis, menjalankan, membangun dan meng-*hosting* kode *python* dalam lingkungan *cloud*. Alat ini sangat berguna bagi para programmer yang ingin membuat *website* sendiri, karena pada *pythonanywhere* terdapat fitur *trial hosting* yang memungkinkan pengguna meng-*hosting website* secara gratis. Kelemahan penggunaan *trial hosting* terdapat pada waktu akses *website* yang hanya bisa diakses secara luas dalam kurun waktu 3 bulan, jika ingin lebih dari itu sebagai pengguna harus membayar agar akses yang diberikan juga semakin lama (Fandango, 2017).

## BAB IV

### METODOLOGI PENELITIAN

#### 4.1. Populasi dan Sampel Penelitian

Populasi yang digunakan untuk melakukan penelitian ini adalah seluruh produk yang ada dalam *website sociolla*, sedangkan untuk sampel data yang dianalisis adalah produk *sunscreen* yang berjumlah 243 produk. Data tersebut berdasarkan dengan topik penelitian yang digunakan dan didapatkan dengan cara *scrapping* menggunakan bantuan *webscrapper*.

#### 4.2. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Universitas Islam Indonesia dari tanggal 30 September 2023 sampai dengan 30 Oktober 2023. Pengambilan datanya sendiri dilakukan pada tanggal 10 Oktober 2023 yang di *scraping* dari *website sociolla*.

#### 4.3. Definisi Operasional Variabel Penelitian

Dalam penelitian ini didapatkan 4 variabel. Berikut merupakan definisi setiap variabel yang digunakan:

**Tabel 4. 1** Definisi Operasional Variabel

No.	Nama Variabel	Definisi Operasional
1.	<i>Brand</i>	Suatu identitas dagang berupa merk yang digunakan perusahaan untuk meningkatkan citra pemasaran mereka.
2.	Nama Produk	Sebuah identitas yang menggambarkan kegunaan dan fungsi dari suatu produk yang diperjual-belikan.
3.	Harga Produk	Sebuah label yang dipasangkan kepada produk sebagai nilai jual dipasaran.
4.	Komposisi Produk	Kandungan yang terdapat dalam produk sebagai bahan-bahan penyusun dan menciptakan produk tersebut.

#### 4.4. Alat dan Cara Organisir Data

Beberapa alat yang digunakan dalam penelitian ini adalah sebuah *software* untuk mengumpulkan dan menganalisis data. Berikut adalah alat-alat yang digunakan penulis.

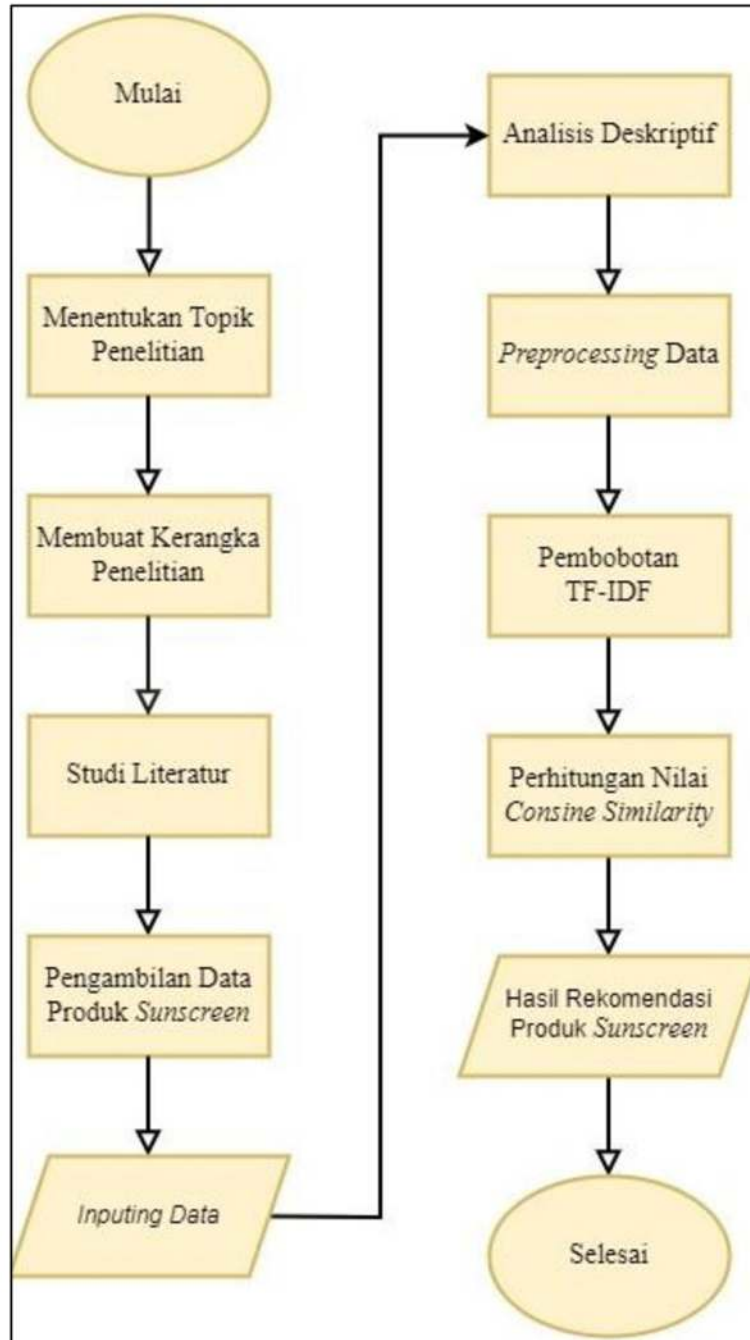
**Tabel 4. 2** Alat dan Organisir Data

No.	Nama <i>Software</i>	Fungsi
1.	<i>Webscraper</i>	Untuk pengambilan data
2.	<i>Microsoft Excel 2016</i>	Untuk menyimpan data
3.	<i>Google Data</i>	Untuk menganalisis data
4.	<i>Visual Studio Code</i>	Untuk membuat <i>website</i>

#### 4.5. Metode Penelitian

Metode yang digunakan untuk menyelesaikan penulisan proposal tugas akhir ini adalah:

1. Menggunakan metode *library research* (studi literatur), yaitu metode penelitian yang dilakukan dengan cara mencari referensi atau literatur pustaka melalui buku, jurnal, hasil laporan penelitian hingga catatan penelitian sebelumnya dengan ketentuan topik maupun metode yang sama.
2. Menggunakan analisis statistika deskriptif dengan cara mengumpulkan, meringkas, menyederhanakan dan menyajikan data dalam bentuk visualisasi yang mudah untuk dipahami. Tujuannya agar memberikan gambaran umum mengenai *dataset* yang digunakan dalam penelitian.
3. Menggunakan metode *Content-Based Filtering* dengan penambahan perhitungan pada data *vector* dalam teks yang ada pada variabel komposisi produk yang nantinya akan diberi pembobotan nilai TF-IDF. Hasil pembobotan akan menghasilkan kemiripan suatu teks yang mewakili setiap produk untuk nantinya dapat di rekomendasikan kepada *user*. Untuk nilai kemiripan dari setiap produk dapat dinilai dari hasil *cosine similarity* untuk menghasilkan rekomendasi produk sesuai dengan produk sebelumnya.



**Gambar 4. 1** Diagram Alir Penelitian (Sumber: Dokumen Pribadi)

Berdasarkan **Gambar 4.1** diatas didapatkan untuk alur penelitian yang dilakukan penulis adalah seperti berikut:

1. Mulai.
2. Menentukan topik dan menyusun kerangka penelitian.
3. Mencari studi literatur berdasarkan kesamaan topik maupun metode penelitian yang digunakan.

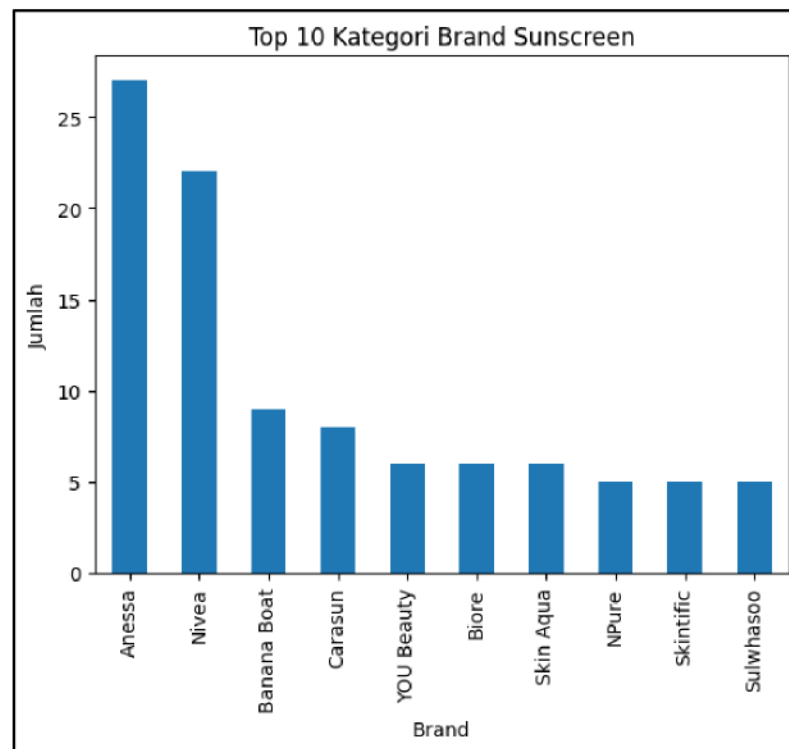
4. Mengidentifikasi topik penelitian dan merumuskan permasalahan yang ingin dibahas dalam penelitian.
5. Melakukan pengambilan data produk *sunscreen* dengan metode *scraping*.
6. Melanjutkan pada proses *text preprocessing* dengan tujuan membersihkan data untuk diolah ke tahap analisis selanjutnya.
7. Menambahkan pembobotan kata menggunakan metode *Term Frequency Inverse Document Frequency* (TF-IDF).
8. Melakukan perhitungan dengan metode *Cosine Similarity* untuk mendapatkan hasil rekomendasi produk *sunscreen* berdasarkan kemiripan komposisi produk yang dimiliki.
9. Membuat *website* sistem rekomendasi sebagai alat pencarian yang digunakan oleh *user*.
10. Mendapatkan hasil rekomendasi 5 produk *sunscreen* berdasarkan kemiripan *ingredients* dari produk sebelumnya.
11. Selesai.

## BAB V

### PEMBAHASAN

#### 5.1. Analisis Deskriptif

Analisis deskriptif merupakan ilmu dari bidang statistika yang digunakan untuk mengolah, meringkas, menginterpretasikan suatu data mentah ke dalam bentuk informasi yang mudah untuk dipahami. Penerapan analisis deskriptif bertujuan untuk memberikan gambaran umum atau mendeskripsikan mengenai data yang didapatkan dari hasil penelitian. Analisis deskriptif dalam penulisan ini menggunakan data produk *sunscreen* yang di dapatkan dengan cara *scrapping*, sehingga untuk yang pertama diperoleh informasi mengenai 10 teratas *brand sunscreen* yang diperjual-belikan pada *website Sociolla*.

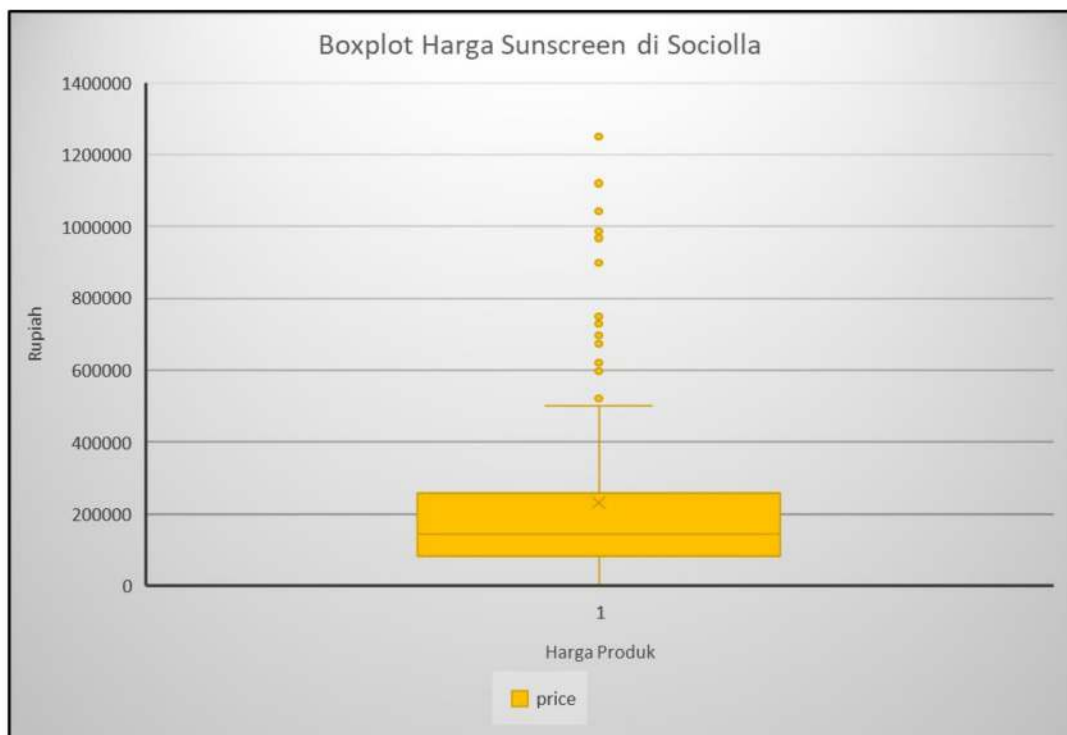


**Gambar 5. 1** 10 *Brand Sunscreen* Teratas di *Sociolla*

Dari data yang digunakan dalam penelitian ini didapatkan analisis deskriptif yang pertama adalah *barchart* yang berisi informasi mengenai 10 *brand* teratas untuk produk *sunscreen* yang ada di *Sociolla*. Berdasarkan **Gambar 5.1** dapat diperoleh informasi dalam *website Sociolla* produk *sunscreen* yang mendominasi pasaran adalah brand dari *Anessa* dan *Nivea* yang masing-masing memiliki 27



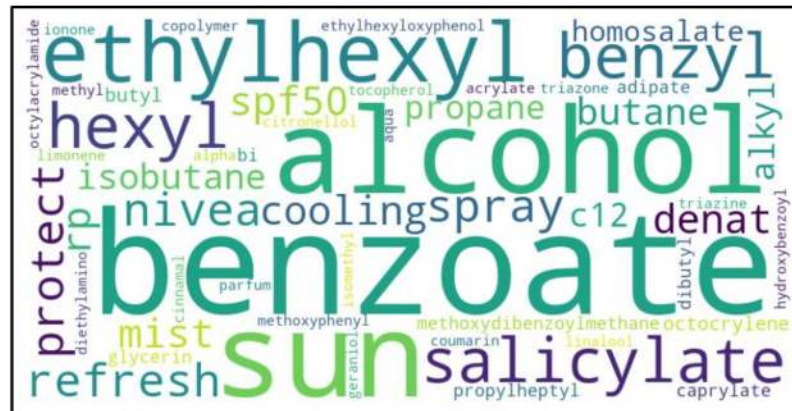
produk dan 22 produk yang diperjual-belikan. Kedua brand tersebut memiliki varian produk *sunscreen* yang beragam dengan penjualan yang menjanjikan sehingga tidak heran produknya mudah dijumpai di *Sociolla*. Selain itu ada juga brand-brand lain yang tidak kalah populer seperti *Banana Boat*, *Carasun*, *Skin Aqua*, *Biore*, *YOU Beauty*, *Skintific* dan lain sebagainya. Pilihan yang beragam tersebut dapat membuat konsumen menjadi lebih mudah mendapatkan produk *sunscreen* yang di inginkan karena dari *website Sociolla* sendiri sudah menyediakan berbagai macam brand terkenal. Berikut merupakan hasil visualisasi untuk harga *sunscreen* yang ada di *website Sociolla*.



**Gambar 5. 2** *Boxplot* Harga *Sunscreen*

Berdasarkan **Gambar 5.2** diperoleh visualisasi *boxplot* berbentuk *right skew* dikarenakan nilai rata-rata lebih besar dari nilai median yang mengartikan data tidak berdistribusi normal. Untuk informasi yang terkandung pada *boxplot* terdapat rata-rata harga produk *sunscreen* yang diperjual-belikan di *Sociolla* sebesar Rp 231.225 dengan *range* harga paling rendah berada dikisaran sebesar Rp 27.158 dari produk *Flawless Bright Brightening Day Cream Spf18* dengan brand Pond's hingga harga termahal berada di angka Rp 1.250.000 pada produk *UV Wise Brightening Multi Protector No. 1 Creamy Glow* dari brand Sulwhasoo yang berasal dari Korea serta

untuk nilai tengahnya berada di angka Rp 143.650; pada visualisasi tersebut juga terdapat beberapa harga yang masuk dalam kategori *outlier* dengan berjumlah 13 item. Selain berdasarkan harga salah satu hal yang menjadi pertimbangan pengambilan keputusan dalam pembelian produk *sunscreen* adalah kandungan bahan penyusun dari produk tersebut. Apabila *ingredients* yang ada pada produk sudah sesuai dengan kebutuhan kulit maka hasil yang didapatkan juga akan jauh lebih maksimal. Untuk itu berikut merupakan visualisasi *ingredients* yang paling banyak ditemukan dalam produk *sunscreen* yang ada *Sociolla*.



**Gambar 5. 3** Wordcloud pada *Ingredients Sunscreen*

Pada **Gambar 5.3** memuat informasi berupa kandungan bahan penyusun atau *ingredients* yang paling sering digunakan pada produk *sunscreen* yang ada dalam *Sociolla*. Beberapa kandungan yang sering digunakan sebagai *ingredients* pada produk *sunscreeen* diantaranya yakni *alcohol* yang berfungsi sebagai pelarut atau pengencer, *benzoate* yang berfungsi untuk bahan pengawet sehingga meminimalisir pertumbuhan mikroorganisme yang menyebabkan kontaminasi pada produk, *ethylhexyl* yang berperan dalam perlindungan dari sinar UV-A maupun UV-B dan ada *salicylate* yang memiliki fungsi utama sebagai penyerapan terhadap radiasi dari sinar UV.

## 5.2. *Preprocessing Data*

Tahapan *text preprocessing* sendiri memiliki pengertian sebagai tahapan awal yang dilakukan dengan cara membersihkan atau menghapus data yang dirasa tidak memiliki nilai tidak penting dan terindikasi adanya duplikat pada data tersebut. Data mentah yang mengandung bagian kata yang tidak terstruktur perlu terlebih dahulu dibersihkan agar menghasilkan analisis yang diperoleh menjadi lebih akurat.

Dalam melakukan *text preprocessing* ada beberapa tahap yang harus dilakukan diantaranya adalah *Case Folding*, *Remove Punctuation*, *Stopword Removal*, *Lemmatization* dan *Tokenizing*. Penelitian ini akan berfokus pada variabel *ingredients* yang terdapat pada produk *sunscreen* yang diperjual-belikan pada *website Sociolla*, untuk itu berikut merupakan contoh variabel *ingredients* yang terdapat pada dokumen ke-81 yang akan digunakan pada tahapan *text preprocessing*.

**Tabel 5. 1** Contoh Dokumen Pada *Text Preprocessing*

<b>Ingredients Produk</b>
Water (Aqua), Alcohols, Ethylhexyl Methoxycinnamate, Dimethicone, Ethylhexyl Salicylate, Homosalate, PEG/PPG-9/2 Dimethyl Ether, Bis-Ethylhexyloxyphenol Methoxyphenyl Triazine, Diethylamino Hydroxybenzoyl Hexyl Benzoate, Diisopropyl Sebacate, Ethylhexyl Triazone, Silica, Titanium Dioxide, Caprylyl Methicone, Glycerin, PPG-17, Acrylamides/Dmapa Acrylates/Methoxy PEG Methacrylate Copolymer, Phenoxyethanol, Dextrin Palmitate/Ethylhexanoate, Trifluoropropyldimethyl/Trimethylsiloxysilicate, Hydrated Silica, PEG-60 Hydrogenated Castor Oil, C30-45 Alkyl Methicone, C30-45 Olefin, Triethoxycaprylylsilane, Carbomer, Potassium Hydroxide, Agar, Xanthan Gum, Acrylates/C10-30 Alkyl Acrylates Crosspolymer, Fragrance (Parfum), Distearyltrimonium Chloride, Butylene Glycol, Dextrin Palmitate, PEG/PPG-14/7 Dimethyl Ether, Stearic Acid, PEG-6, BHT, Saccharide Isomerate, Dipotassium Glycyrrhizate, Camellia Sinensis Leaf Extract, Potentilla Erecta Root Extract, Citric Acid, Sodium Acetylated Hyaluronate, Sodium Citrate, Synthetic Fluorphlogopite, Triethoxysilyethyl Polydimethylsiloxyethyl Dimethicone, Tocopherol, Soluble Collagen

Berdasarkan **Tabel 5.1** memuat contoh dokumen yang digunakan pada tahapan *text preprocessing* yang berisikan informasi mengenai *ingredients* dari produk *sunscreen Anessa Perfect UV Gentle Kit*

### **5.2.1 Case Folding**

Proses *case folding* merupakan tahapan awal dalam *text preprocessing* yang memiliki tujuan untuk mengubah huruf kapital pada data menjadi huruf kecil keseluruhan. *Case folding* perlu dilakukan karena adanya *inconsistency* pada data

sehingga harus dilakukan penyamaratakan struktur kata menjadi huruf kecil semua. Hal tersebut digunakan agar memudahkan adanya pencarian, pemrosesan dan analisis data supaya hasil yang didapatkan lebih akurat. Berikut merupakan hasil dari *case folding* untuk contoh dokumen ke-81 pada *ingredients* produk *sunscreen Anessa Perfect UV Gentle Kit*.

**Tabel 5. 2** Hasil *Case Folding*

Sebelum	Sesudah
Water (Aqua), Alcohol, Ethylhexyl Methoxycinnamate, Dimethicone, Ethylhexyl Salicylate, Homosalate, PEG/PPG-9/2 Dimethyl Ether, Bis-Ethylhexyloxyphenol Methoxyphenyl Triazine, Diethylamino Hydroxybenzoyl Hexyl Benzoate, Diisopropyl Sebacate, Ethylhexyl Triazone, Silica, Titanium Dioxide, Caprylyl Methicone, Glycerin, PPG-17, Acrylamides/Dmapa Acrylates/Methoxy PEG Methacrylate Copolymer, Phenoxyethanol, Dextrin Palmitate/Ethylhexanoate, Trifluoropropyldimethyl/ Trimethylsiloxysilicate, Hydrated Silica, PEG-60 Hydrogenated Castor Oil, C30-45 Alkyl Methicone, C30-45 Olefin, Triethoxycaprylylsilane, Carbomer, Potassium Hydroxide, Agar, Xanthan Gum, Acrylates/C10-30 Alkyl Acrylate Crosspolymer, Fragrance (Parfum), Distearyltrimonium Chloride, Butylene Glycol, Dextrin Palmitate, PEG/PPG-	water (aqua), alcohols, ethylhexyl methoxycinnamate, dimethicone, ethylhexyl salicylate, homosalate, peg/ppg-9/2 dimethyl ether, bis-ethylhexyloxyphenol methoxyphenyl triazine, diethylamino hydroxybenzoyl hexyl benzoate, diisopropyl sebacate, ethylhexyl triazone, silica, titanium dioxide, caprylyl methicone, glycerin, ppg-17, acrylamides/dmapa acrylates/methoxy peg methacrylate copolymer, phenoxyethanol, dextrin palmitate/ethylhexanoate, trifluoropropyldimethyl/ trimethylsiloxysilicate, hydrated silica, peg-60 hydrogenated castor oil, c30-45 alkyl methicone, c30-45 olefin, triethoxycaprylylsilane, carbomer, potassium hydroxide, agar, xanthan gum, acrylates/c10-30 alkyl acrylates crosspolymer, fragrance (parfum), distearyldimonium chloride, butylene glycol, dextrin palmitate, peg/ppg-14/7 dimethyl ether, stearic acid, peg-6, bht, saccharide isomerate, dipotassium

Sebelum	Sesudah
14/7 Dimethyl Ether, Stearic Acid, PEG-6, BHT, Saccharide Isomerate, Dipotassium Glycyrrhizate, Camellia Sinensis Leaf Extract, Potentilla Erecta Root Extract, Citric Acid, Sodium Acetylated Hyaluronate, Sodium Citrate, Synthetic Fluorophlogopite, Triethoxysilylethyl Polydimethylsiloxylethyl Dimethicone, Tocopherol, Soluble Collagen	glycyrrhizate, camellia sinensis leaf extract, potentilla erecta root extract, citric acid, sodium acetylated hyaluronate, sodium citrate, synthetic fluorophlogopite, triethoxysilylethyl polydimethylsiloxylethyl dimethicone, tocopherol, soluble collagen

Berdasarkan **Tabel 5.2** dapat diketahui bahwa perubahan struktur kata yang semula huruf kapital telah diubah menjadi huruf kecil setelah melewati tahapan *case folding*, perubahan tersebut ditandai dengan huruf bergaris bawah dan berwarna kuning

### 5.2.2 Remove Punctuation

*Remove Punctuation* merupakan tahapan *text preprocessing* yang digunakan untuk menghilangkan tanda baca atau spasi berlebih pada data teks yang akan dianalisis, tanda baca atau spasi pada data teks yang dianalisis tidak memiliki makna atau nilai penting pada hasil analisis sehingga perlu dihilangkan agar hasil yang didapatkan semakin akurat. Beberapa tanda baca yang dihapus diantaranya adalah tanda koma (,), tanda seru (!), tanda tanya (?), garis miring (/) dan lain sebagainya. Berikut merupakan hasil dari tahapan *remove punctuation* dari dokumen ke-81 pada *ingredients* produk *sunscreen Anessa Perfect UV Gentle Kit*.

**Tabel 5.3** Hasil *Remove Punctuation*

Sebelum	Sesudah
water (aqua), alcohols, ethylhexyl methoxycinnamate, dimethicone, ethylhexyl salicylate, homosalate, peg/ppg 9/2 dimethyl ether, bis-ethylhexyloxyphenol methoxyphenyl	water aqua alcohols ethylhexyl methoxycinnamate dimethicone ethylhexyl salicylate homosalate peg ppg 9 2 dimethyl ether bis ethylhexyloxyphenol methoxyphenyl

Sebelum	Sesudah
triazine, diethylamino hydroxybenzoyl hexyl benzoate, diisopropyl sebacate, ethylhexyl triazone, silica, titanium dioxide, caprylyl methicone, glycerin, ppg-17, acrylamides, dmapa acrylates/ methoxy peg methacrylate copolymer, phenoxyethanol, dextrin palmitate/ ethylhexanoate, trifluoropropyldimethyl/ trimethylsiloxysilicate, hydrated silica, peg-60 hydrogenated castor oil, c30-45 alkyl methicone, c30-45 olefin, triethoxycaprylylsilane, carbomer, potassium hydroxide, agar, xanthan gum, acrylates/c10-30 alkyl acrylates crosspolymer, fragrance (parfum), distearyldimonium chloride, butylene glycol, dextrin palmitate, peg/ppg-14/7 dimethyl ether, stearic acid, peg-6, bht, saccharide isomerate, dipotassium glycyrrhizate, camellia sinensis leaf extract, potentilla erecta root extract, citric acid, sodium acetylated hyaluronate, sodium citrate, synthetic fluorphlogopite, triethoxysilylethyl polydimethylsiloxylethyl dimethicone, tocopherol, soluble collagen	triazine diethylamino hydroxybenzoyl hexyl benzoate diisopropyl sebacate ethylhexyl triazone silica titanium dioxide caprylyl methicone glycerin ppg 17 acrylamides dmapa acrylates methoxy peg methacrylate copolymer phenoxyethanol dextrin palmitate ethylhexanoate trifluoropropyldimethyl trimethylsiloxysilicate hydrated silica peg 60 hydrogenated castor oil c30 45 alkyl methicone c30 45 olefin triethoxycaprylylsilane carbomer potassium hydroxide agar xanthan gum acrylates c10 30 alkyl acrylates crosspolymer fragrance parfum distearyldimonium chloride butylene glycol dextrin palmitate peg ppg 14 7 dimethyl ether stearic acid peg 6 bht saccharide isomerate dipotassium glycyrrhizate camellia sinensis leaf extract potentilla erecta root extract citric acid sodium acetylated hyaluronate sodium citrate synthetic fluorphlogopite triethoxysilylethyl polydimethylsiloxylethyl dimethicone tocopherol soluble collagen

Berdasarkan **Tabel 5.3** dapat diketahui pada perubahan yang ditandai dengan tanda baca bergaris bawah berwarna kuning bahwa beberapa tanda baca seperti titik, koma dan garis miring telah dihilangkan dari struktur kata yang dianalisis.

### 5.2.3 Remove Double Space

*Remove double space* adalah tahapan *text preprocessing* yang digunakan untuk menghilangkan adanya spasi berganda yang terdapat pada data teks yang dianalisis. Dalam tahapan *remove double white space* memiliki peranan untuk merapikan struktur data teks yang digunakan supaya memunculkan hasil analisis yang lebih akurat. Berikut merupakan hasil dokumen ke-81 yang sudah melalui tahapan *remove double white space*.

**Tabel 5. 4** Hasil *Remove Double Space*

Sebelum	Sesudah
water aqua alcohols ethylhexyl methoxycinnamate dimethicone ethylhexyl salicylate homosalate peg ppg 9 2 dimethyl ether bis ethylhexyloxyphenol methoxyphenyl triazine diethylamino hydroxybenzoyl hexyl benzoate diisopropyl sebacate ethylhexyl triazone silica titanium dioxide caprylyl methicone glycerin ppg 17 acrylamides dmapa acrylates methoxy peg methacrylate copolymer phenoxyethanol dextrin palmitate ethylhexanoate trifluoropropyldimethyl trimethylsiloxysilicate hydrated silica peg 60 hydrogenated castor oil c30 45 alkyl methicone c30 45 olefin triethoxycaprylylsilane carbomer potassium hydroxide agar xanthan gum acrylates c10 30 alkyl acrylates crosspolymer fragrance parfum distearyldimonium chloride butylene glycol dextrin palmitate peg ppg 14 7	water aqua alcohols ethylhexyl methoxycinnamate dimethicone ethylhexyl salicylate homosalate peg ppg 9 2 dimethyl ether bis ethylhexyloxyphenol methoxyphenyl triazine diethylamino hydroxybenzoyl hexyl benzoate diisopropyl sebacate ethylhexyl triazone silica titanium dioxide caprylyl methicone glycerin ppg 17 acrylamides dmapa acrylates methoxy peg methacrylate copolymer phenoxyethanol dextrin palmitate ethylhexanoate trifluoropropyldimethyl trimethylsiloxysilicate hydrated silica peg 60 hydrogenated castor oil c30 45 alkyl methicone c30 45 olefin triethoxycaprylylsilane carbomer potassium hydroxide agar xanthan gum acrylates c10 30 alkyl acrylates crosspolymer fragrance parfum distearyldimonium chloride butylene glycol dextrin palmitate peg ppg 14 7

Sebelum	Sesudah
dimethyl ether stearic acid peg 6 bht saccharide isomerate dipotassium glycyrrhizate camellia sinensis leaf extract potentilla erecta root extract citric acid sodium acetylated hyaluronate sodium citrate synthetic fluorphlogopite triethoxysilylethyl polydimethylsiloxylethyl dimethicone tocopherol soluble collagen	dimethyl ether stearic acid peg 6 bht saccharide isomerate dipotassium glycyrrhizate camellia sinensis leaf extract potentilla erecta root extract citric acid sodium acetylated hyaluronate sodium citrate synthetic fluorphlogopite triethoxysilylethyl polydimethylsiloxylethyl dimethicone tocopherol soluble collagen

Berdasarkan hasil **Tabel 5.4** diperoleh hasil dari dokumen ke-81 yang merupakan *ingredients* dari produk *sunscreen Anessa Perfect UV Gentle Kit* telah dihilangkan *double space* yang terdapat pada data. Sehingga terdapat perbedaan sebelum dan sesudah melewati tahapan *double space removal*.

#### 5.2.4 Lemmatization

*Lemmatization* merupakan suatu proses analisis *text preprocessing* yang ditujukan untuk menemukan bentuk dasar dalam data teks yang digunakan pada kalimat. Dokumen yang digunakan dalam analisis biasanya memiliki kumpulan kata kerja, aktif, pasif sebagai penyusun sebuah kalimat. Maka dalam proses *lemmatization* kata tersebut nantinya akan dikembalikan dalam bentuk kata dasarnya agar dapat mengurangi jumlah variasi kata dalam dokumen atau data teks yang digunakan serta analisis akan mendapatkan hasil yang lebih akurat. Pada dokumen yang digunakan pada analisis ini menggunakan *ingredients* yang tertulis dalam Bahasa Inggris, berikut merupakan hasil dari proses *lemmatization*.

**Tabel 5.5** Hasil Proses *Lemmatization*

Sebelum	Sesudah
water aqua alcohols ethylhexyl methoxycinnamate dimethicone ethylhexyl salicylate homosalate peg ppg 9 2 dimethyl ether bis ethylhexyloxyphenol methoxyphenyl	water aqua alcohol ethylhexyl methoxycinnamate dimethicone ethylhexyl salicylate homosalate peg ppg 9 2 dimethyl ether bi ethylhexyloxyphenol methoxyphenyl



Sebelum	Sesudah
triazine diethylamino hydroxybenzoyl hexyl benzoate diisopropyl sebacate ethylhexyl triazone silica titanium dioxide caprylyl methicone glycerin ppg 17 acrylamides dmapa acrylates methoxy peg methacrylate copolymer phenoxyethanol dextrin palmitate ethylhexanoate trifluoropropyldimethyl trimethylsiloxysilicate hydrated silica peg 60 hydrogenated castor oil c30 45 alkyl methicone c30 45 olefin triethoxycaprylylsilane carbomer potassium hydroxide agar xanthan gum acrylates c10 30 alkyl acrylates crosspolymer fragrance parfum distearyldimonium chloride butylene glycol dextrin palmitate peg ppg 14 7 dimethyl ether stearic acid peg 6 bht saccharide isomerate dipotassium glycyrrhizate camellia sinensis leaf extract potentilla erecta root extract citric acid sodium acetylated hyaluronate sodium citrate synthetic fluorphlogopite triethoxysilylethyl polydimethylsiloxylethyl dimethicone tocopherol soluble collagen	triazine diethylamino hydroxybenzoyl hexyl benzoate diisopropyl sebacate ethylhexyl triazone silica titanium dioxide caprylyl methicone glycerin ppg 17 <b>acrylamide</b> dmapa <b>acrylate</b> methoxy peg methacrylate copolymer phenoxyethanol dextrin palmitate ethylhexanoate trifluoropropyldimethyl trimethylsiloxysilicate hydrated silica peg 60 hydrogenated castor oil c30 45 alkyl methicone c30 45 olefin triethoxycaprylylsilane carbomer potassium hydroxide agar xanthan gum <b>acrylate</b> c10 30 alkyl <b>acrylate</b> crosspolymer fragrance parfum distearyldimonium chloride butylene glycol dextrin palmitate peg ppg 14 7 dimethyl ether stearic acid peg 6 bht saccharide isomerate dipotassium glycyrrhizate camellia sinensis leaf extract potentilla erecta root extract citric acid sodium acetylated hyaluronate sodium citrate synthetic fluorphlogopite triethoxysilylethyl polydimethylsiloxylethyl dimethicone tocopherol soluble collagen

Hasil yang ditunjukkan pada **Tabel 5.5** proses *lemmatization* untuk dokumen ke-81 yang memuat informasi *ingredients* dari produk *sunscreen Anessa Perfect UV Gentle Kit*, menunjukkan perubahan kata pada *alcohols* menjadi *alcohol*, *bis* menjadi *bi*, *acrylamides* menjadi *acrylamide* dan *acrylates* menjadi *acrylate*.

Perubahan tersebut dapat dilihat untuk setiap kata yang ditandai dengan warna kuning.

### 5.2.5 Tokenizing

*Tokenizing* adalah bagian dari proses *text preprocessing* yang digunakan untuk membagi atau memisahkan suatu kalimat pada data teks menjadi potongan kata dalam unit terkecil yang disebut token. Token yang sudah dibagi dapat berupa kata, simbol, frasa, tanda baca dan tanda penghubung. Tujuan dilakukannya proses *tokenizing* agar setiap kata yang sudah dibagi ke dalam unit terkecil menjadi suatu entitas yang memiliki nilai ketika nantinya diberikan pembobotan pada tahapan analisis selanjutnya, berikut merupakan hasil yang diperoleh dari tahapan *tokenizing*.

**Tabel 5. 6** Hasil Tahapan *Tokenizing*

Sebelum	Sesudah
water aqua alcohol ethylhexyl methoxycinnamate dimethicone ethylhexyl salicylate homosalate peg ppg 9 2 dimethyl ether bi ethylhexyloxyphenol methoxyphenyl triazine diethylamino hydroxybenzoyl hexyl benzoate diisopropyl sebacate ethylhexyl triazone silica titanium dioxide caprylyl methicone glycerin ppg 17 acrylamide dmapa acrylate methoxy peg methacrylate copolymer phenoxyethanol dextrin palmitate ethylhexanoate trifluoropropyldimethyl trimethylsiloxysilicate hydrated silica peg 60 hydrogenated castor oil c30 45 alkyl methicone c30 45 olefin triethoxycaprylylsilane carbomer potassium hydroxide agar xanthan gum	[ 'water', 'aqua', 'alcohol', 'ethylhexyl', 'methoxycinnamate', 'dimethicone', 'ethylhexyl', 'salicylate', 'homosalate', 'peg', 'ppg', '9', '2', 'dimethyl', 'ether', 'bi', 'ethylhexyloxyphenol', 'methoxyphenyl', 'triazine', 'diethylamino', 'hydroxybenzoyl', 'hexyl', 'benzoate', 'diisopropyl', 'sebacate', 'ethylhexyl', 'triazone', 'silica', 'titanium', 'dioxide', 'caprylyl', 'methicone', 'glycerin', 'ppg', '17', 'acrylamide', 'dmapa', 'acrylate', 'methoxy', 'peg', 'methacrylate', 'copolymer', 'phenoxyethanol', 'dextrin', 'palmitate', 'ethylhexanoate', 'trifluoropropyldimethyl', 'trimethylsiloxysilicate', 'hydrated', 'silica', 'peg', '60', 'hydrogenated',

Sebelum	Sesudah
acrylate c10 30 alkyl acrylate crosspolymer fragrance parfum distearyldimonium chloride butylene glycol dextrin palmitate peg ppg 14 7 dimethyl ether stearic acid peg 6 bht saccharide isomerate dipotassium glycyrrhizate camellia sinensis leaf extract potentilla erecta root extract citric acid sodium acetylated hyaluronate sodium citrate synthetic fluorphlogopite triethoxysilylethyl polydimethylsiloxylethyl dimethicone tocopherol soluble collagen	'castor', 'oil', 'c30', '45', 'alkyl', 'methicone', 'c30', '45', 'olefin', 'triethoxycaprylylsilane', 'carbomer', 'potassium', 'hydroxide', 'agar', 'xanthan', 'gum', 'acrylate', 'c10', '30', 'alkyl', 'acrylate', 'crosspolymer', 'fragrance', 'parfum', 'distearyldimonium', 'chloride', 'butylene', 'glycol', 'dextrin', 'palmitate', 'peg', 'ppg', '14', '7', 'dimethyl', 'ether', 'stearic', 'acid', 'peg', '6', 'bht', 'saccharide', 'isomerate', 'dipotassium', 'glycyrrhizate', 'camellia', 'sinensis', 'leaf', 'extract', 'potentilla', 'erecta', 'root', 'extract', 'citric', 'acid', 'sodium', 'acetylated', 'hyaluronate', 'sodium', 'citrate', 'synthetic', 'fluorphlogopite', 'triethoxysilylethyl', 'polydimethylsiloxylethyl', 'dimethicone', 'tocopherol', 'soluble', 'collagen' ]

Berdasarkan **Tabel 5.6** dari proses *tokenizing* pada dokumen ke-81 untuk *ingredients* dari produk *sunscreen Anessa Perfect UV Gentle Kit* telah dibagi atau dipisahkan dalam bentuk unit terkecil yang disebut token. Hasil pembagian tersebut ditandai dengan setiap token ditandai dengan setiap potongan memiliki tanda petik dan koma sebagai pemisah.

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

Tahapan analisis setelah dilakukan proses *text preprocessing* adalah pembobotan kata menggunakan metode *Term Frequency-Inverse Document* (TF-IDF). Pembobotan ini merupakan gabungan dari metode *Term Frequency* (TF) dan

*Inverse Document Frequency* (IDF), Pengertian *Term Frequency* sendiri merupakan metode pembobotan sesuai dengan frekuensi munculnya kata dalam suatu dokumen sedangkan *Inverse Document Frequency* adalah banyaknya dokumen pada suatu kalimat yang muncul. Pembobotan disini berfungsi untuk mengubah data teks menjadi data numerik dengan nilai pembobot agar dapat mengukur hubungan sebuah kata dalam konteks dokumen yang dianalisis.

Proses pembobotan yang pertama menggunakan metode *Term Frequency* dengan ketentuan *ngram* = 2,2 atau gabungan *ingredients* yang tersusun dari 2 suku kata. Hal tersebut didasari karena melihat karakteristik dari data tekstual berupa nama *ingredients* yang banyak tersusun dari 2 senyawa kimia, sehingga dengan menggunakan *ngram*=2,2 maka akan lebih merepresentasikan *ingredients* tersebut. Berikut merupakan beberapa sampel kata berdasarkan dokumen yang dianalisis yaitu “*acid calcium*”, “*acne oil*”, “*alcohol butylene*”, “*alkyl benzoate*”, dan “*allantoin ceramide*”. Berikut merupakan hasil pembobotan dengan TF dari sampel kata tersebut

**Tabel 5. 7** Hasil Pembobotan Nilai TF

Dokumen	<i>acid calcium</i>	<i>acne oil</i>	<i>alcohol butylene</i>	<i>alkyl benzoate</i>	...	<i>allantoin ceramide</i>
1	0	0	0	0	...	0
2	0	0	0	0	...	0
3	0	0	0	0	...	0
4	0	0	0	0	...	0
5	0	0	0	0	...	0
6	0	0	0	0	...	0
7	0	0	0	0	...	0
8	0	0	0	0	...	0
9	0	0	0	0	...	0
10	0	0	0	0	...	0
11	0	0	0	0	...	0
12	0	0	0	0	...	0
13	0	0	0	0	...	0
14	0	0	0	0	...	0

Dokumen	<i>acid</i>	<i>acne</i>	<i>alcohol</i>	<i>alkyl</i>	...	<i>allantoin</i>
	<i>calcium</i>	<i>oil</i>	<i>butylene</i>	<i>benzoate</i>		<i>ceramide</i>
15	0	0	0	1	...	0
16	0	0	0	1	...	0
17	0	0	0	0	...	0
18	0	0	0	0	...	0
19	0	0	0	0	...	0
20	0	0	0	1	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
243	0	0	0	0	...	0
DF	1	3	2	51	...	1
IDF	5.497	4.398	4.804	2.561	...	5.497

Berdasarkan hasil pembobotan dengan metode *Term Frequency* pada **Tabel 5.7** diperoleh matriks berukuran  $243 \times 7050$ . Nilai TF pada **Tabel 5.7** bernilai 0 maka pada suatu dokumen tidak terdapat *term* atau kata tertentu, sedangkan apabila nilai  $TF \geq 0$  artinya pada suatu dokumen terdapat 1, 2, 3, ...,  $n$  *term*. Hal tersebut dapat dilihat pada contoh nilai TF dokumen ke-15 dimana untuk sampel kata “*alkyl benzoate*” mengalami frekuensi kemunculan sebanyak 1 kali karena nilai TF-nya sendiri adalah 1.

Tahapan setelah mendapatkan nilai TF pada masing-masing sampel kata adalah dilakukan perhitungan nilai IDF, Perhitungan ini dapat dihitung menggunakan rumus sebagai berikut.

$$IDF_{alkyl\ benzoate} = \ln\left(\frac{D}{DF_{alkyl\ benzoate}}\right) + 1$$

$$IDF_{alkyl\ benzoate} = \ln\left(\frac{243}{51}\right) + 1 = 2.561$$

Nilai D pada rumus diatas merupakan nilai yang mewakili banyaknya jumlah dokumen atau data sebanyak 243 dokumen. Untuk nilai  $DF_{alkyl\ benzoate}$  merupakan nilai yang mewakili jumlah kemunculan kata “*alkyl benzoate*”.

Berdasarkan hasil perhitungan nilai TF-IDF diatas dapat diperoleh kesimpulan bahwa sampel kata “*alkyl benzoate*” pada dokumen ke-15 mempunyai nilai

pembobotan TF-IDF 2.561. Berikut merupakan hasil perhitungan TF-IDF untuk seluruh sampel kata yang digunakan.

**Tabel 5. 8** Hasil Pembobotan Nilai TF-IDF

Dokumen	<i>acid</i> <i>calcium</i>	<i>acne</i> <i>oil</i>	<i>alcohol</i> <i>butylene</i>	<i>alkyl</i> <i>benzoate</i>	...	<i>allantoin</i> <i>ceramide</i>
1	0	0	0	0	...	0
2	0	0	0	0	...	0
3	0	0	0	0	...	0
4	0	0	0	0	...	0
5	0	0	0	0	...	0
6	0	0	0	0	...	0
7	0	0	0	0	...	0
8	0	0	0	0	...	0
9	0	0	0	0	...	0
10	0	0	0	0	...	0
11	0	0	0	0	...	0
12	0	0	0	0	...	0
13	0	0	0	0	...	0
14	0	0	0	0	...	0
15	0	0	0	2.561	...	0
16	0	0	0	2.561	...	0
17	0	0	0	0	...	0
18	0	0	0	0	...	0
19	0	0	0	0	...	0
20	0	0	0	2.561	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
243	0	0	0	0	...	0

#### 5.4. Hasil Rekomendasi Berdasarkan *Cosine Similarity*

Proses selanjutnya setelah dilakukannya pembobotan kata dengan metode TF-IDF maka peneliti melakukan perhitungan *cosine similarity* yang digunakan untuk membandingkan similaritas antar dokumen. Nilai *cosine similarity* sendiri

berada pada rentang 0 hingga 1, dimana suatu dokumen dikatakan mirip atau memiliki similaritas yang tinggi apabila nilainya mendekati 1 begitu juga sebaliknya apabila nilai dokumen yang didapatkan mendekati 0 maka tingkat similaritas dokumen tersebut juga semakin rendah. Nilai yang diperoleh sebagai acuan kemiripan untuk setiap dokumen yang sudah di analisis yang selanjutnya akan diurutkan berdasarkan 5 dokumen teratas yang memiliki nilai similaritas tertinggi sebagai hasil dari rekomendasi suatu produk *sunscreen*. Berdasarkan hasil *cosine similarity* peneliti menggunakan sampel dari dokumen ke-171 yang merupakan produk *ANNESSA Trial Set Brightening Gel*, berikut merupakan hasil 5 dokumen teratas yang memiliki hasil nilai *cosine similraty* tertinggi

**Tabel 5. 9** Hasil Rekomendasi dengan Similaritas Tertinggi

No.	Dokumen	Nilai <i>Cosine Similarity</i>
1	170	0.902268
2	92	0.866628
3	224	0.855433
4	200	0.852363
5	81	0.505011

Berikut merupakan contoh perhitungan manual dalam menentukan *cosine similarity* dengan menggunakan sampel dokumen ke-171 terhadap dokumen ke-170. Berikut merupakan vektor untuk dokumen ke-170 dan dokumen ke-171:

$$\overrightarrow{D_{170}} = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \dots, 0]$$

$$\overrightarrow{D_{171}} = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2,8, 0, 0, 0, 0, \dots, 0]$$

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

$$\cos \alpha = \frac{\sum_{i=1}^n (D171_i \cdot D170_i)}{\sqrt{\sum_{i=1}^n (D171_i)^2} \times \sqrt{\sum_{i=1}^n (D170_i)^2}}$$

$$\frac{(0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + \dots + (0 \times 0)}{\sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + \dots + 0^2} \times \sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + \dots + 0^2}}$$

$$\cos \alpha = \frac{1646.428}{\sqrt{(6.535)^2} \times \sqrt{(6.493)^2}}$$

$$\cos \alpha = \frac{1646.428}{\sqrt{42.708} \times \sqrt{42.1638}} = 0.90226$$

Berdasarkan hasil perhitungan manual yang dilakukan pada sampel dokumen ke-171 dengan dokumen ke-170 menghasilkan nilai kemiripan sebesar 90%. Sehingga hasil rekomendasi untuk produk *sunscreen* dokumen ke-171 yakni *ANNESSA Trial Set Brightening Gel* mendapatkan 5 hasil rekomendasi *sunscreen* dengan nilai *cosine similarity* paling tinggi sebagai berikut.

**Tabel 5. 10** 5 Rekomendasi Produk *Sunscreen* Teratas

No.	Dokumen	Nama <i>Sunscreen</i>	Nilai <i>Cosine Similarity</i>
1	170	<i>ANESSA Trial Set Gold Gel</i>	0.902268
2	92	<i>New Brightening UV Sunscreen Gel</i>	0.866628
3	224	<i>Twinpack New Perfect UV Sunscreen Skincare Gel</i>	0.855433
4	200	<i>New Perfect UV Sunscreen Gel</i>	0.852363
5	81	<i>Anessa Perfect Gentle UV Kit</i>	0.505011

Dari hasil rekomendasi pada **Tabel 5.10** untuk kata kunci yang sudah dimasukkan pengguna yakni *ANNESSA Trial Set Brightening Gel* diberikan hasil rekomendasi dengan nilai *cosine similarity* tertinggi pada produk *sunscreen ANESSA Trial Set Gold Gel*, Hasil tersebut didasarkan dengan kemiripan dokumen pada kedua produk yang dilihat berdasarkan kandungan *ingredients* masing-masing produk. Nantinya akan digunakan untuk memvalidasi hasil rekomendasi yang diberikan oleh sistem maka peneliti akan membandingkan *ingredients* yang terkandung pada masing-masing produk yakni pada dokumen ke-170 dan 171, berikut adalah *ingredients* kedua produk *sunscreen* tersebut.

**Tabel 5. 11** *Ingredients* Dokumen ke-170 dan 171

<i>ANESSA Trial Set Brightening Gel</i>	<i>ANESSA Trial Set Gold Gel</i>
Water (Aqua), Alcohols, Ethylhexyl Methoxycinnamate, Dimethicone, Ethylhexyl Salicylate, Homosalate,	Water (Aqua), Alcohol, Dimethicone, Ethylhexyl Methoxycinnamate, Ethylhexyl Salicylate, Bis-



ANESSA Trial Set Brightening Gel	ANESSA Trial Set Gold Gel
<p>PEG/PPG-9/2 Dimethyl Ether, Bis-Ethylhexyloxyphenol Methoxyphenyl Triazine, Diethylamino Hydroxybenzoyl Hexyl Benzoate, Diisopropyl Sebacate, Ethylhexyl Triazone, Silica, Titanium Dioxide, Caprylyl Methicone, Glycerin, PPG-17, Acrylamides/Dmapa Acrylates/Methoxy PEG Methacrylate Copolymer, Phenoxyethanol, Dextrin Palmitate/Ethylhexanoate, Trifluoropropyldimethyl/Trimethylsiloxysilicate, Hydrated Silica, PEG-60 Hydrogenated Castor Oil, C30-45 Alkyl Methicone, C30-45 Olefin, Triethoxycaprylylsilane, Carbomer, Potassium Hydroxide, Agar, Xanthan Gum, Acrylates/C10-30 Alkyl Acrylates Crosspolymer, (Parfum), Distearyltrimonium Chloride, Butylene Glycol, Dextrin Palmitate, PEG/PPG-14/7 Dimethyl Ether, Stearic Acid, PEG-6, BHT, Saccharide Isomerate, Dipotassium Glycyrrhizate, Camellia Sinensis Leaf Extract, Potentilla Erecta Root Extract, Citric Acid, Sodium Acetylated Hyaluronate, Sodium Citrate, Synthetic Fluorophlogopite, Triethoxysilylethyl</p>	<p>Ethylhexyloxyphenol Methoxyphenyl Triazine, Diisopropyl Sebacate, Silica, Ethylhexyl Triazone, Diethylamino Hydroxybenzoyl Hexyl Benzoate, Titanium Dioxide, Glycerin, Caprylyl Methicone, Ppg-17, Acrylamides/Dmapa Acrylates/Methoxy Peg Methacrylate Copolymer, Phenoxyethanol, Dextrin Palmitate/Ethylhexanoate, Peg-60 Hydrogenated Castor Oil, C30-45 Alkyl Methicone, Hydrated Silica, C30-45 Olefin, Triethoxycaprylylsilane, Carbomer, Potassium Hydroxide, Agar, Xanthan Gum, Acrylates/C10-30 Alkyl Acrylate Crosspolymer, (Parfum), Distearyltrimonium Chloride, Butylene Glycol, Stearic Acid, Dextrin Palmitate, Peg/Ppg-14/7 Dimethyl Ether, Bht, Dipotassium Glycyrrhizate, Leontopodium Alpinum Extract, Camellia Sinensis Leaf Extract, Potentilla Erecta Root Extract, Talc, Sodium Acetylated Hyaluronate, Soluble Collagen.</p>

ANESSA Trial Set Brightening Gel	ANESSA Trial Set Gold Gel
Polydimethylsiloxylethyl Dimethicone, Tocopherol, Soluble Collagen	

Berdasarkan **Tabel 5.11** telah dipaparkan kandungan *ingredients* antara produk *sunscreen ANESSA Trial Set Brightening Gel* dengan produk *ANESSA Trial Set Gold Gel* yang memiliki tingkat kemiripan sebesar 90%. Lebih jelasnya letak kemiripan *ingredients* pada kedua produk tersebut, berikut perbandingan *ingredients* yang terkandung pada masing-masing produk.

**Tabel 5. 12** Perbandingan *Ingredients* Dokumen ke-170 dan 171

ANESSA Trial Set Brightening Gel	ANESSA Trial Set Gold Gel
Water (Aqua), Alcohols, Ethylhexyl Methoxycinnamate, Dimethicone, Ethylhexyl Salicylate, Homosalate, PEG/PPG-9/2 Dimethyl Ether, Bis-Ethylhexyloxyphenol Methoxyphenyl Triazine, Diethylamino Hydroxybenzoyl Hexyl Benzoate, Diisopropyl Sebacate, Ethylhexyl Triazone, Silica, Titanium Dioxide, Caprylyl Methicone, Glycerin, PPG-17, Acrylamides/Dmapa Acrylates/Methoxy PEG Methacrylate Copolymer, Phenoxyethanol, Dextrin Palmitate/Ethylhexanoate, Trifluoropropyldimethyl/Trimethylsiloxysilicate, Hydrated Silica, PEG-60 Hydrogenated Castor Oil, C30-45 Alkyl Methicone, C30-45 Olefin, Triethoxycaprylylsilane, Carbomer, Potassium Hydroxide, Agar, Xanthan Gum, Acrylates/C10-30	Water (Aqua), Alcohol, Dimethicone, Ethylhexyl Methoxycinnamate, Ethylhexyl Salicylate, Bis-Ethylhexyloxyphenol Methoxyphenyl Triazine, Diisopropyl Sebacate, Silica, Ethylhexyl Triazone, Diethylamino Hydroxybenzoyl Hexyl Benzoate, Titanium Dioxide, Glycerin, Caprylyl Methicone, Ppg-17, Acrylamides/Dmapa Acrylates/Methoxy Peg Methacrylate Copolymer, Phenoxyethanol, Dextrin Palmitate/Ethylhexanoate, Peg-60 Hydrogenated Castor Oil, C30-45 Alkyl Methicone, Hydrated Silica, C30-45 Olefin, Triethoxycaprylylsilane, Carbomer, Potassium Hydroxide, Agar, Xanthan Gum, Acrylates/C10-30 Alkyl Acrylate Crosspolymer, Fragrance (Parfum), Distearyltrimonium

ANESSA Trial Set Brightening Gel	ANESSA Trial Set Gold Gel
Alkyl Acrylates Crosspolymer, Fragrance (Parfum), Distearyltrimonium Chloride, Butylene Glycol, Dextrin Palmitate, PEG/PPG-14/7 Dimethyl Ether, Stearic Acid, PEG-6, BHT, Saccharide Isomerate, Dipotassium Glycyrrhizate, Camellia Sinensis Leaf Extract, Potentilla Erecta Root Extract, Citric Acid, Sodium Acetylated Hyaluronate, Sodium Citrate, Synthetic Fluorophlogopite, Triethoxysilylethyl Polydimethylsiloxylethyl Dimethicone, Tocopherol, Soluble Collagen	Chloride, Butylene Glycol, Stearic Acid, Dextrin Palmitate, Peg/Ppg-14/7 Dimethyl Ether, Bht, Dipotassium Glycyrrhizate, Leontopodium Alpinum Extract, Camellia Sinensis Leaf Extract, Potentilla Erecta Root Extract, Talc, Sodium Acetylated Hyaluronate, Soluble Collagen.

Berdasarkan **Tabel 5.12** diperoleh hasil perbandingan *ingredients* yang terkandung pada *sunscreen ANESSA Trial Set Brightening Gel* dan *ANESSA Trial Set Gold Gel* dimana letak kemiripan kedua produk *sunscreen* tersebut ada pada bagian *ingredients* yang sudah diberikan tanda teks berwarna kuning. Terlihat banyaknya kemiripan pada *ingredients* keduanya maka hasil tingkat similaritas kedua produk *sunscreen* tersebut tinggi yakni sebesar 90%, jadi alternatif pilihan produk untuk *sunscreen ANESSA Trial Set Brightening Gel* apabila produk tersebut tidak tersedia maka pengguna dapat membeli produk serupa yang memiliki mayoritas kandungan *ingredients* yang sama yakni *ANESSA Trial Set Gold Gel*.

Hasil tersebut berguna untuk memvalidasi hasil rekomendasi yang diberikan oleh *website* sistem rekomendasi, peneliti juga akan membandingkan kembali untuk produk kedua *New Brightening UV Sunscreen Gel* yang mendapatkan tingkat kemiripan 86% dari produk *sunscreen* yang menjadi kata kunci yakni *ANESSA Trial Set Brightening Gel*. Berikut adalah perbandingan *ingredients* penyusun kedua produk tersebut.

**Tabel 5. 13** Perbandingan *Ingredients* Dokumen ke-170 dan 92

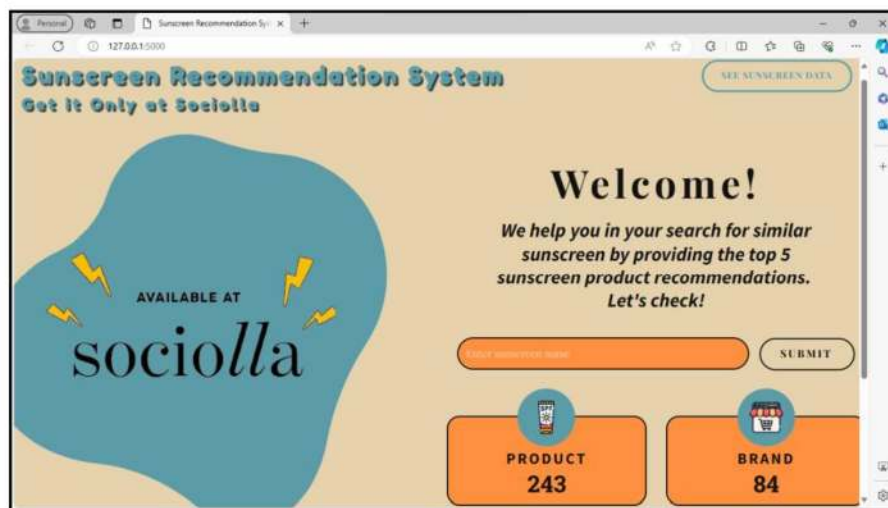
ANESSA Trial Set Brightening Gel	ANESSA Trial Set Gold Gel
<p>Water (Aqua), Alcohols, Ethylhexyl Methoxycinnamate, Dimethicone, Ethylhexyl Salicylate, Homosalate, PEG/PPG-9/2 Dimethyl Ether, Bis-Ethylhexyloxyphenol Methoxyphenyl Triazine, Diethylamino Hydroxybenzoyl Hexyl Benzoate, Diisopropyl Sebacate, Ethylhexyl Triazone, Silica, Titanium Dioxide, Caprylyl Methicone, Glycerin, PPG-17, Acrylamides/Dmapa Acrylates/Methoxy PEG Methacrylate Copolymer, Phenoxyethanol, Dextrin Palmitate/Ethylhexanoate, Trifluoropropyldimethyl/Trimethylsiloxysilicate, Hydrated Silica, PEG-60 Hydrogenated Castor Oil, C30-45 Alkyl Methicone, C30-45 Olefin, Triethoxycaprylylsilane, Carbomer, Potassium Hydroxide, Agar, Xanthan Gum, Acrylates/C10-30 Alkyl Acrylate Crosspolymer, Fragrance (Parfum), Distearyltrimonium Chloride, Butylene Glycol, Dextrin Palmitate, PEG/PPG-14/7 Dimethyl Ether, Stearic Acid, PEG-6, BHT, Saccharide Isomerate, Dipotassium Glycyrrhizate, Camellia Sinensis Leaf Extract, Potentilla Erecta Root Extract, Citric Acid, Sodium</p>	<p>Water (Aqua), Alcohol, Dimethicone, Ethylhexyl Methoxycinnamate, Ethylhexyl Salicylate, Bis-Ethylhexyloxyphenol Methoxyphenyl Triazine, Diisopropyl Sebacate, Silica, Ethylhexyl Triazone, Diethylamino Hydroxybenzoyl Hexyl Benzoate, Titanium Dioxide, Glycerin, Caprylyl Methicone, Ppg-17, Acrylamides/Dmapa Acrylates/Methoxy Peg Methacrylate Copolymer, Phenoxyethanol, Dextrin Palmitate/Ethylhexanoate, Peg-60 Hydrogenated Castor Oil, C30-45 Alkyl Methicone, Hydrated Silica, C30-45 Olefin, Triethoxycaprylylsilane, Carbomer, Potassium Hydroxide, Agar, Xanthan Gum, Acrylates/C10-30 Alkyl Acrylate Crosspolymer, Fragrance (Parfum), Distearyltrimonium Chloride, Butylene Glycol, Stearic Acid, Dextrin Palmitate, Peg/Ppg-14/7 Dimethyl Ether, Bht, Dipotassium Glycyrrhizate, Leontopodium Alpinum Extract, Camellia Sinensis Leaf Extract, Potentilla Erecta Root Extract, Talc, Sodium Acetylated Hyaluronate, Soluble Collagen</p>

ANESSA Trial Set Brightening Gel	ANESSA Trial Set Gold Gel
Acetylated Hyaluronate, Sodium Citrate, Synthetic Fluorophlogopite, Triethoxysilylethyl Polydimethylsiloxylethyl Dimethicone, Tocopherol, Soluble Collagen	

Berdasarkan **Tabel 5.13** diperoleh hasil perbandingan *ingredients* yang terkandung pada *sunscreen ANESSA Trial Set Brightening Gel* dan *UV Brightening Sunscreen Gel* dimana letak kemiripan kedua produk *sunscreen* tersebut ada pada bagian *ingredients* yang sudah dilabeli tulisan teks bewarna kuning.

### 5.5. Tampilan Website Sistem Rekomendasi Sunscreen

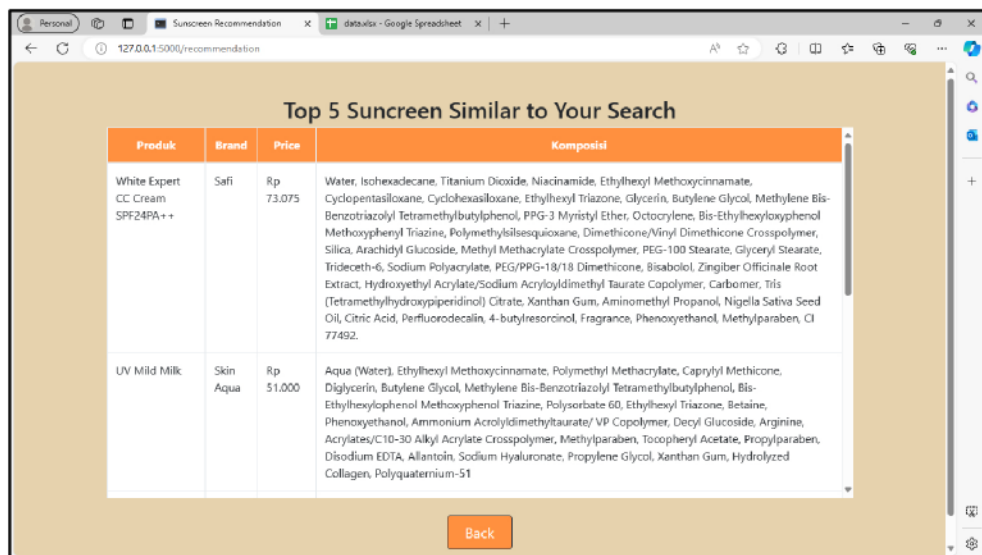
Tahapan terakhir dalam penelitian yang dilakukan adalah dengan membuat suatu sistem rekomendasi yang dapat diakses pengguna untuk mencari rekomendasi produk *sunscreen* yang dapat dibeli melalui *website Sociolla* secara langsung. Sistem rekomendasi yang dibuat peneliti sendiri berupa sebuah *website* yang sudah di *deploy* menggunakan *framework flask phyton* yang nantinya akan di *hosting* dengan *PhytonAnywhere* agar dapat diakses secara luas. Berikut merupakan tampilan *website* yang sudah dibuat sebagai sarana pengaplikasian sistem rekomendasi *sunscreen*.



**Gambar 5.4** Tampilan *Home* Website Sistem Rekomendasi

Berdasarkan **Gambar 5.4** telah memuat tampilan berupa menu *home* yang sudah ditampilkan oleh *website* sistem rekomendasi *sunscreen*. Penggunaan

website tersebut dapat dimulai dengan mengklik *dataset* pada menu “SEE DATA SUNSCREEN” yang berfungsi untuk melihat data produk *sunscreen* yang nantinya bisa dimasukkan dalam kolom *search* agar bisa memperoleh hasil produk yang memiliki kemiripan dengan produk yang dicari. Setiap kata kunci yang dimasukkan harus sesuai dengan nama produk yang ada dalam *dataset* seperti “UV Mild Milk” bukan “Uv mild milk”, karena apabila tidak sesuai maka *website* tidak akan memunculkan hasil. Jika sudah memasukkan kata kunci maka pengguna tinggal mengklik *icon submit* dan akan memunculkan hasil rekomendasi oleh sistem, untuk lebih detailnya dapat diakses melalui <https://khoirulniam.pythonanywhere.com/>. Sebagai contoh dengan menggunakan kata kunci “UV Whitening Milk” diperoleh hasil rekomendasi 5 produk *sunscreen* sebagai berikut.



**Gambar 5. 5** Tampilan Hasil Rekomendasi dari Website Berdasarkan **Gambar 5.5** dengan kata kunci “UV Whitening Milk” diperoleh 2 produk *sunscreen* yang direkomendasikan oleh sistem adalah *White Expert CC Cream SPF24PA++* dan *UV Mild Milk*, untuk memvalidasi hasil rekomendasi dari sistem berikut kecocokan produk yang dicari dan hasil rekomendasi

**Tabel 5. 14** Highlight Pencarian Produk dan Hasil Rekomendasi

Produk	Ingredients
UV Whitening Milk (Kata Kunci)	Aqua (Water), Ethylhexyl Methoxycinnamate, Butylene Glycol, Methylene Bis-Benzotriazolyl Tetramethylbutylphenol, Polymethylsilsesquioxane, Cyclopentasiloxane, Tranexamic Acid, Diethylamino

	Hydroxybenzoyl Hexyl Benzoate, Polysorbate 60, Cetearyl Alcohol, Silica, Bis-Ethylhexyloxyphenol Methoxyphenyl Triazine, Decyl Glucoside, Titanium Dioxide, Phenoxyethanol, Triethanolamine, Acrylates/C10-30 Alkyl Acrylate Crosspolymer, Hexylglycerin, Disodium EDTA, Propylene Glycol, Xanthan Gum, Sodium Acetylated Hyaluronate (AcHA)
White Expert CC Cream SPF24PA++ (Hasil)	Water, Isohexadecane, Titanium Dioxide, Niacinamide, Ethylhexyl Methoxycinnamate, Cyclopentasiloxane, Cyclohexasiloxane, Ethylhexyl Triazone, Glycerin, Butylene Glycol, Methylene Bis-Benzotriazolyl Tetramethylbutylphenol, PPG-3 Myristyl Ether, Octocrylene, Bis-Ethylhexyloxyphenol Methoxyphenyl Triazine, Polymethylsilsesquioxane, Dimethicone/Vinyl Dimethicone Crosspolymer, Silica, Arachidyl Glucoside, Methyl Methacrylate Crosspolymer, PEG-100 Stearate, Glyceryl Stearate, Trideceth-6, Sodium Polyacrylate, PEG/PPG-18/18 Dimethicone, Bisabolol, Zingiber Officinale Root Extract, Hydroxyethyl Acrylate/Sodium Acryloyldimethyl Taurate Copolymer, Carbomer, Tris (Tetramethylhydroxypiperidinol) Citrate, Xanthan Gum, Aminomethyl Propanol, Nigella Sativa Seed Oil, Citric Acid, Perfluorodecalin, 4-butylresorcinol, Fragrance, Phenoxyethanol, Methylparaben, CI 77492.
UV Mild Milk (Hasil)	Aqua (Water), Ethylhexyl Methoxycinnamate, Polymethyl Methacrylate, Caprylyl Methicone, Diglycerin, Butylene Glycol, Methylene Bis-Benzotriazolyl Tetramethylbutylphenol, Bis-Ethylhexyloxyphenol Methoxyphenyl Triazine, Polysorbate 60, Ethylhexyl Triazone, Betaine, Phenoxyethanol, Ammonium Acryloyldimethyltaurate/ VP Copolymer, Decyl Glucoside, Arginine, Acrylates/C10-30 Alkyl Acrylate Crosspolymer,

	Methylparaben, Tocopheryl Acetate, Propylparaben, Disodium EDTA, Allantoin, Sodium Hyaluronate, Propylene Glycol, Xanthan Gum, Hydrolyzed Collagen, Polyquaternium-51
--	---



## BAB VI

### PENUTUP

#### 6.1. Kesimpulan

Berdasarkan pengerjaan analisis dan pembahasan dari penelitian mengenai sistem rekomendasi *sunscreen* yang diambil dari studi kasus *website Sociolla* dapat diperoleh kesimpulan sebagai berikut.

1. Berdasarkan analisis deskriptif, pada penelitian ini diperoleh dalam *website sociolla* terdapat 2 brand besar yang mendominasi pasaran yakni brand dari *ANESSA* dan *NIVEA*. Ditemukan juga harga produk terendah dengan harga Rp 27.158 terdapat pada produk *Flawless Bright Brightening Day Cream Spf18* dengan brand *Pond's* hingga harga termahal berada di angka Rp 1.250.000 pada produk *UV Wise Brightening Multi Protector No. 1 Creamy Glow* dari brand *Sulwhasoo* yang berasal dari Korea. *Ingredients* yang banyak dijumpai diantaranya adalah *alcohol, benzoate, ethylhexyl* dan *salicylate*.
2. Dalam tahapan untuk memberikan rekomendasi produk *sunscreen* berdasarkan *ingredients* penyusun produk menggunakan metode *content-based filtering* dimulai dengan melakukan *text preprocessing* data yang terdiri dari *case folding, remove punctuation, remove double space, stopword removal, lemmatization* dan *tokenizing*, kemudian dilanjutkan dengan pembobotan kata dengan TF-IDF dan nantinya hasil pembobotan akan menjadi dasar perhitungan *cosine similarity* untuk mencari nilai kemiripan antar dokumen. Hasil akhir yang didapatkan berupa 5 rekomendasi produk *sunscreen* yang memiliki nilai kemiripan tertinggi sesuai kandungan *ingredients* penyusun produk. Langkah pembuatan *website* rekomendasi *sunscreen* berdasarkan model yang sudah diperoleh dengan *web framework* berbahasa *python* yaitu *Flask* yang kemudian akan di *hosting* menggunakan *PythonAnywhere* agar dapat digunakan pengguna secara luas
3. Dengan menggunakan perhitungan *cosine similarity* pada penelitian ini menggunakan sampel produk *ANESSA Trial Set Brightening Gel* ,

sehingga diperoleh 5 dokumen dengan tingkat kemiripan tertinggi yakni produk *ANESSA Trial Set Gold Gel* dengan nilai kemiripan sebesar 90%, produk *New Perfect UV Sunscreen Skincare Spray 60gr* dengan nilai kemiripan sebesar 86%, produk *Twinpack New Perfect UV Sunscreen Skincare Gel* dengan nilai kemiripan 85%, *New Perfect UV Sunscreen Skincare Gel* dengan nilai kemiripan atau 85% dan produk *Anessa Perfect Gentle UV Kit* dengan nilai kemiripan sebesar 50%.

## 6.2. Saran

Berdasarkan penelitian dan analisis yang telah dilakukan, diberikan beberapa saran yang dapat menunjang perbaikan penelitian selanjutnya sebagai berikut.

1. Dalam tahapan *text processing* apabila tidak menemukan perbedaan hasil analisis sebelum dan sesudah, maka dapat dihindari contohnya pada penelitian ini dibagian *stopword removal*.
2. Dapat mengembangkan metode penelitian sistem rekomendasi yang digunakan seperti metode *collaborative filtering*, *hybrid filtering* dan *demographic*.
3. Dalam *database* dapat dicatumkan *link* pembelian produk agar memudahkan pengguna untuk mendapatkan produk yang di inginkan
4. Untuk data *sunscreen* yang digunakan penelitian selanjutnya dapat menambahkan variabel seperti kecocokan sesuai dengan jenis kulit, kategori usia pengguna (dewasa/anak-anak).
5. Penelitian selanjutnya dapat melibatkan kolaborasi dengan jurusan kimia atau farmasi untuk menunjang kesesuaian hasil dan pengembangan penelitian yang dijalankan.
6. Untuk penggunaan kata kunci dapat dibuat lebih umum tidak spesifik sesuai dengan *database* sehingga harapannya kata kunci yang bisa dicari dalam *website* tidak terbatas.

## DAFTAR PUSTAKA

- Abbe, et al., A. (2015). Text mining applications in psychiatry: a systematic literature review. *International Journal of Methods in Psychiatric Research*, Volume 25, Issue No. 2, Page 86-100.
- Adi, P. S. (2010). Sistem Rekomendasi Nilai Mata Kuliah Menggunakan Metode Content-Based Filtering. *Seminar Nasional Informatika* (hal. semnasIF 2010). Yogyakarta: UPN "Veteran" Yogyakarta.
- Alita, D., & Rahman, A. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi (Ilmu Komputer Unila)*, Vol. 8 No. 2.
- Asmiati, et al., E. (2021). Edukasi Pentingnya Penggunaan Sunscreen pada Kalangan Remaja di SMA Islam Sabilillah Malang. *Jurnal Pengabdian UNDIKMA*, Vol. 2, No. 2, 189-194.
- Bargamanis, T., & Goldwater, S. (2018). Context Sensitive Neural Lemmatization with Lematus. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1, Pages 1391-1400.
- Fandango, A. (2017). Python Data Analysis. *Data Manipulation and Complex Data Analysis with Python*, Packt Publishing.
- Fikriyah, S. N., & Sibaroni, Y. (2022). Identify User Behavior based on Tweet Type on Twitter Platform using Mean Shift Clustering. *Jurnal Media Informatika Budidarma*, Volume 6, Nomor 3, Page 1396-1403.
- Goos, P., & Meintrup., D. (2015). *Statistics with JMP : Graphs, Descriptive Statistics and Probability*. United Kingdom: WILEY.
- Gwak, et al., M. A. (2021). Hyaluronic acid/tannic acid hydrogel sunscreen with excellent anti-UV, antioxidant, and cooling effects. *International Journal of Biological Macromolecules*, Pages 918-924.
- Henni Natalia Sijabat, S. T. (2023). An Error Analysis Of Using Punctuation In Writing Narrative Text Of Tenth Grade Students At Sma Gajah Mada Medan . *INNOVATIVE: Journal of Social Science Research*, Volume 3 Nomor 5 Page 2560-2573.

- Herwijayanti, et al., B. (2018). Klasifikasi Berita Online dengan Menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2, No. 1, hal. 306-312.
- Hidayatullah, A. F., & Ma'arif, M. R. (2016). Penerapan Text Mining dalam Klasifikasi Judul Skripsi. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, Yogyakarta.
- Huda, et al., A. A. (2022). Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita. *Building of Informatics Technology and Science (BITS)*, Volume 4, No. 3, Pages 1679-1686.
- Isfardiyana & Safitri . (2014). Pentingnya Melindungi Kulit dari Sinar Ultraviolet dan Cara Melindungi Kulit dengan Sunblock Buatan Sendiri. *Jurnal Inovasi Kewirausahaan*, Halaman 126-133.
- Isfardiyana, S. H., & Safitri, S. R. (2014). Pentingnya Melindungi Kulit dari Sinar Ultraviolet dan Cara Melindungi Kulit dengan Sunblock Buatan Sendiri. *Jurnal Inovasi dan Kewirausahaan*, Volume 3, Halaman 126-133.
- Isinkaye, F. O. (2015). Recommendation Systems: Principles, Methods and Evaluation. *Egyptian Informatics Journal*, Volume 16, Issue 3, Pages 261-273.
- Kalangi, S. J. (2013). Histofisiologi Kulit. *Jurnal Biomedik (JBM)*, Volume 5, Nomor 3, hlm. S12-20.
- Kaur, J., & Preetpal, K. B. (2018). A systematic review on stopwords removal algorithms. *International Journal on Future Revolution in Computer Science & Communication Engineering 4.4*, 207-210.
- Larasati, F. B., & Februariyanti, H. (2021). Sistem Rekomendasi Product Emina Cosmetics Dengan Menggunakan Metode Content - Based Filtering. *Jurnal Manajemen Informatika dan Sistem Informasi*, Volume 4, Nomor 1.
- Moh. Afif Rofiqi, et al. (Desember 2019). *ILKOMNIKA: Journal of Computer Science Applied Informatics*, Vol. 2, No. 2.
- Mukaromah A.H., M. E. (2008). Perbedaan Degradasi ION Cr(VI) Dengan Penyinaran UV dan Tanpa Penyinaran UV. *Jurnal Ilmu Kesehatan*, Volume 1, No. 1.

- Mumtazah, et al., E. F. (2020). Pengetahuan Mengenai Sunscreen Dan Bahaya Paparan Sinar Matahari Serta Perilaku Mahasiswa Teknik Sipil Terhadap Penggunaan Sunscreen. *Jurnal Farmasi Komunitas*, Volume 7, Nomor 2.
- Nastiti. (2019). Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan. *Jurnal Teknologi Informasi dan Komunikasi*, Vol. 8, No. 1.
- Nastiti, P. (2019). Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan. *TEKNIKA*, Volume 8, Nomor 1, 1-10.
- Nugraha, A. A., & Sanjaya, N. A. (2023). Penyusunan Penyusunan Sistem Rekomendasi Produk Diecast Sistem Rekomendasi Produk Diecast. *JNATIA (Jurnal Nasional Teknologi Informasi dan Aplikasinya)*, Volume 1, Nomor 3.
- Nugraha, A. K., & Sebastian, D. (2018). Pembentukan Dataset Topik Kata Bahasa Indonesia pada Twitter Menggunakan TF-IDF & Cosine Similarity. *Jurnal Teknik Informatika dan Sistem Informasi*, Volume 3, Nomor 3, 376-386.
- Pratiwi, N. P. (2022). Putri Pratiwi, N. A. B. I. L. A. (2022). Formulasi Dan Uji Aktivitas Antioksidan Krim Niacinamide Dengan Metode Dpph (2, 2-Diphenyl-1-Picrylhydrazyl) Menggunakan Spektrofotometri Uv-Vis. *Doctoral Dissertation*, Sekolah Tinggi Ilmu Kesehatan Muhammadiyah Ciamis.
- Putri, et al., M. W. (2020). Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF. *Journal of Information Technology and Computer Science*, Volume 5, No. 3.
- Rahmawati, K., & Solandjari, W. (2019). Pemanfaatan Ekstrak Kulit Pisang Kepok Kuning (*Musa paradisiaca* L) Sebagai Krim Tabir Surya. *Akademisi Analisis Farmasi dan Makanan Putra Indonesia*, Malang.
- Ramlah, et al., S. (2021). Sistem Pakar Penentuan Komposisi Skincare Berdasarkan Karakteristik Jenis Kulit Menggunakan Metode Certainty Factor. *Buletin Sistem Informasi dan Teknologi Islam*, Vol. 2, No. 1, Pages 36-42.
- Rizky, R. N., & Noor, I. (2020). *Pengaruh Harga, Pendapatan, Dan Halal Awareness Terhadap Keputusan Pembelian Produk Kosmetik Impor Pada*

- Konsumen Pengguna E-Commerce Sociolla (Studi Pada Mahasiswa Di Universitas Brawijaya)*. Malang: Fakultas Ekonomi dan Bisnis Universitas Brawijaya.
- Shuib, N. S., & Muniandy, D. (2018). Sun Protection Awareness Among Schoolchildren In Malaysia. *Proceedings of the Research and Development Seminar Nuklear Malaysia*, Universiti Kebangsaan Malaysia (UKM).
- Wadoe, et al., M. (2019). Penggunaan dan Pengetahuan Sunscreen Pada Mahasiswa UNAIR. *Jurnal Farmasi Komunitas*, Volume 6, No. 1.
- Widolaras, R., & Ikhsanto, M. N. (2022). Sistem Pendukung Keputusan Pemilihan Tabir Surya Wajah untuk Kulit Berminyak Menggunakan Metode Analytical Hierarchy Process. *Jurnal Teknologi Informatika dan Komputer MH. Thamrin*, Volume 8, No. 2.
- Adeline Abbe, C. G. (2015). Text mining applications in psychiatry: a systematic literature review. *International Journal of Methods in Psychiatric Research*, Volume 25, Issue No. 2, Page 86-100.
- Adi, P. S. (2010). Sistem Rekomendasi Nilai Mata Kuliah Menggunakan Metode Content-Based Filtering. *Seminar Nasional Informatika* (hal. semnasIF 2010). Yogyakarta: UPN "Veteran" Yogyakarta.
- Ahmad Fathan Hidayatullah, M. R. (2016). Penerapan Text Mining dalam Klasifikasi Judul Skripsi. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, Yogyakarta.
- Anak Agung Aditya Nugraha, N. A. (2023). Penyusunan Penyusunan Sistem Rekomendasi Produk Diecast Sistem Rekomendasi Produk Diecast . *JNATIA (Jurnal Nasional Teknologi Informasi dan Aplikasinya)*, Volume 1, Nomor 3.
- Arif Akbarul Huda, R. F. (2022). Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita. *Building of Informatics Technology and Science (BITS)*, Volume 4, No. 3, Pages 1679-1686.
- Bening Herwijayanti, D. E. (Januari 2018). Klasifikasi Berita Online dengan Menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal*

- Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2, No. 1, hal. 306-312.
- Debby Alita, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi (Ilmu Komputer Unila)*, Vol. 8 No. 2.
- Dwi Ayu Nur Safitri, R. H. (2021). Sistem Rekomendasi Skincare Menggunakan Metode Content-Based Filtering dan Algoritma Apriori. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, Vol. 5, No. 2, Pages 242-248.
- Edlia Fadilah Mumtazah, et al. (2020). Pengetahuan Mengenai Sunscreen Dan Bahaya Paparan Sinar Matahari Serta Perilaku Mahasiswa Teknik Sipil Terhadap Penggunaan Sunscreen. *Jurnal Farmasi Komunitas*, Volume 7, Nomor 2.
- Elva Asmiati, N. A. (2021). Edukasi Pentingnya Penggunaan Sunscreen pada Kalangan Remaja di SMA Islam Sabilillah Malang. *Jurnal Pengabdian UNDIKMA*, Vol. 2, No. 2, 189-194.
- Fatoni Batari Agung Larasati, H. F. (2021). Sistem Rekomendasi Product Emina Cosmetics Dengan Menggunakan Metode Content - Based Filtering. *Jurnal Manajemen Informatika dan Sistem Informasi*, Volume 4, Nomor 1.
- Henni Natalia Sijabat, S. T. (2023). An Error Analysis Of Using Punctuation In Writing Narrative Text Of Tenth Grade Students At Sma Gajah Mada Medan . *INNOVATIVE: Journal of Social Science Research*, Volume 3 Nomor 5 Page 2560-2573.
- Isinkaye, F. O. (2015). Recommendation Systems: Principles, Methods and Evaluation. *Egyptian Informatics Journal*, Volume 16, Issue 3, Pages 261-273.
- Jashanjot Kaur, K. B. (2018). A systematic review on stopword removal algorithms. *International Journal on Future Revolution in Computer Science & Communication Engineering 4.4*, 207-210.
- Kalangi, S. J. (2013). Histofisiologi Kulit. *Jurnal Biomedik (JBM)*, Volume 5, Nomor 3, hlm. S12-20.

- Kiki Rahmawati, W. S. (2019). Pemanfaatan Ekstrak Kulit Pisang Kepok Kuning (*Musa paradisiaca* L) Sebagai Krim Tabir Surya. *Akademisi Analis Farmasi dan Makanan Putra Indonesia*, Malang.
- Kristian Adi Nugraha, D. S. (2018). Pembentukan Dataset Topik Kata Bahasa Indonesia pada Twitter Menggunakan TF-IDF & Cosine Similarity. *Jurnal Teknik Informatika dan Sistem Informasi*, Volume 3, Nomor 3, 376-386.
- Maria Sri Wulandari, R. N. (2022). Prediksi Jenis Kulit Wajah Menggunakan Metode Certainty Factor Sebagai Sistem Rekomendasi Produk Berbasis Web. *SEMNAS Teknologi Informasi dan Komunikasi*, Vol. 6, No. 1, Pages. 141-148.
- Mariana Widia Putri, A. M. (2020). Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF. *Journal of Information Technology and Computer Science*, Volume 5, No. 3.
- Michael Wadoe, et al. (2019). Penggunaan dan Pengetahuan Sunscreen Pada Mahasiswa UNAIR. *Jurnal Farmasi Komunitas*, Volume 6, No. 1.
- Min A. Gwak, B. M. (2021). Hyaluronic acid/tannic acid hydrogel sunscreen with excellent anti-UV, antioxidant, and cooling effects. *International Journal of Biological Macromolecules*, Pages 918-924.
- Moh. Afif Rofiqi, et al. (Desember 2019). *ILKOMNIKA: Journal of Computer Science Applied Informatics*, Vol. 2, No. 2.
- Mukaromah A.H., M. E. (2008). Perbedaan Degradasi ION Cr(VI) Dengan Penyinaran UV dan Tanpa Penyinaran UV. *Jurnal Ilmu Kesehatan*, Volume 1, No. 1.
- Nastiti, P. (2019). Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan. *TEKNIKA*, Volume 8, Nomor 1, 1-10.
- Noor Ezati Shuib, D. M. (2018). Sun Protection Awareness Among Schoolchildren In Malaysia. *Proceedings of the Research and Development Seminar Nuklear Malaysia*, Universiti Kebangsaan Malaysia (UKM).
- Peter Goos, D. (2015). *Statistics with JMP : Graphs, Descriptive Statistics and Probability*. United Kingdom : WILEY.



- Pratiwi, N. P. (2022). Formulasi Dan Uji Aktivitas Antioksidan Krim Niacinamide Dengan Metode Dpph (2, 2-Diphenyl-1-Picrylhydrazyl) Menggunakan Spektrofotometri UV-Vis. *Doctoral Dissertation*, Sekolah Tinggi Ilmu Kesehatan Muhammadiyah Ciamis.
- Putri Nastiti. (2019). Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan. *Jurnal Teknologi Informasi dan Komunikasi*, Vol. 8, No. 1.
- Rahman Indra Kesuma, A. I. (2020). Penerapan Content-Boosted Collaborative Filtering untuk Meningkatkan Kemampuan Sistem Rekomendasi Penyedia Jasa Acara Pernikahan. *Jurnal Imiah FIFO*, Vol. 12, No 1.
- Ratih Widolaras, M. N. (2022). Sistem Pendukung Keputusan Pemilihan Tabir Surya Wajah untuk Kulit Berminyak Menggunakan Metode Analytical Hierarchy Process. *Jurnal Teknologi Informatika dan Komputer MH. Thamrin*, Volume 8, No. 2.
- Rosalina Noor Rizky, D. I. (2020). *Pengaruh Harga, Pendapatan, Dan Halal Awareness Terhadap Keputusan Pembelian Produk Kosmetik Impor Pada Konsumen Pengguna E-Commerce Sociolla (Studi Pada Mahasiswa Di Universitas Brawijaya)*. Malang: Fakultas Ekonomi dan Bisnis Universitas Brawijaya.
- Saniyah Nabila Fikriyah, Y. S. (2022). Identify User Behavior based on Tweet Type on Twitter Platform using Mean Shift Clustering. *Jurnal Media Informatika Budidarma*, Volume 6, Nomor 3, Page 1396-1403.
- Siti Hapsah Isfardiyana, S. R. (2014). Pentingnya Melindungi Kulit dari Sinar Ultraviolet dan Cara Melindungi Kulit dengan Sunblock Buatan Sendiri. *Jurnal Inovasi dan Kewirausahaan*, Volume 3, Halaman 126-133.
- Sitti Ramlah, P. L. (2021). Sistem Pakar Penentuan Komposisi Skincare Berdasarkan Karakteristik Jenis Kulit Menggunakan Metode Certainty Factor. *Buletin Sistem Informasi dan Teknologi Islam*, Vol. 2, No. 1, Pages 36-42.
- Toms Bargmanis, S. G. (2018). Context Sensitive Neural Lemmatization with Lematus. *Proceedings of the 2018 Conference of the North American*

*Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, Pages 1391-1400.*

## LAMPIRAN

Lampiran 1	<i>Dataset</i>
	<a href="https://bit.ly/3xWvbGi">https://bit.ly/3xWvbGi</a>
Lampiran 2	Perhitungan <i>Cosine Similarity</i>
	<a href="https://bit.ly/4bjfpDD">https://bit.ly/4bjfpDD</a>
Lampiran 3	<i>Script Analisis Sistem Rekomendasi</i>
	<a href="https://bit.ly/4a3BPYm">https://bit.ly/4a3BPYm</a>
Lampiran 4	<i>Deploy Sistem dan Web</i>
	<a href="https://bit.ly/3UpNH1i">https://bit.ly/3UpNH1i</a>