

**SISTEM REKOMENDASI SKINCARE DENGAN METODE KEYWORD  
EXTRACTION DAN COSINE SIMILARITY**

**LAPORAN TUGAS AKHIR**

Laporan ini Disusun untuk Memenuhi Salah Satu Syarat Memperoleh  
Gelar Sarjana Strata 1 (S1) pada Program Studi Teknik Informatika  
Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang



**DISUSUN OLEH:**

**ANNISA AMALIA PUTRI**

**NIM 32601900006**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM SULTAN AGUNG  
SEMARANG**

**2023**

**FINAL PROJECT**

**SKINCARE RECOMMENDATION SYSTEM USING KEYWORD  
EXTRACTION AND COSINE SIMILARITY**

Proposed to complete the requirement to obtain a bachelor's degree (S1)  
at Informatics Engineering Departement of Industrial Technology Faculty  
Sultan Agung Islamic University



**DEPARTMENT OF INFORMATICS ENGINEERING  
INDUSTRIAL TECHNOLOGY FACULTY  
SULTAN AGUNG ISLAMIC UNIVERSITY  
SEMARANG  
2023**

## LEMBAR PENGESAHAN PEMBIMBING

Laporan Tugas Akhir dengan judul “**SISTEM REKOMENDASI SKINCARE DENGAN METODE KEYWORD EXTRACTION DAN COSINE SIMILARITY**” ini disusun oleh :

Nama : Annisa Amalia Putri

NIM : 32601900006

Program Studi : Teknik Informatika

Telah disahkan oleh dosen pembimbing pada :

Hari : Kamis

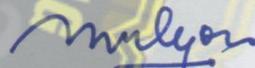
Tanggal : 10 Agustus 2023

Mengesahkan,

Pembimbing I

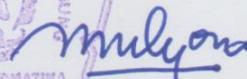
Pembimbing II

  
Sam Farisa Chaerul H, S.T., M.Kom  
NIDN. 0628028602

  
Ir. Sri Mulyono, M.Eng  
NIDN. 0626066601

Mengetahui,

Ketua Program Studi Teknik Informatika  
Fakultas Teknologi Industri  
Universitas Islam Sultan Agung

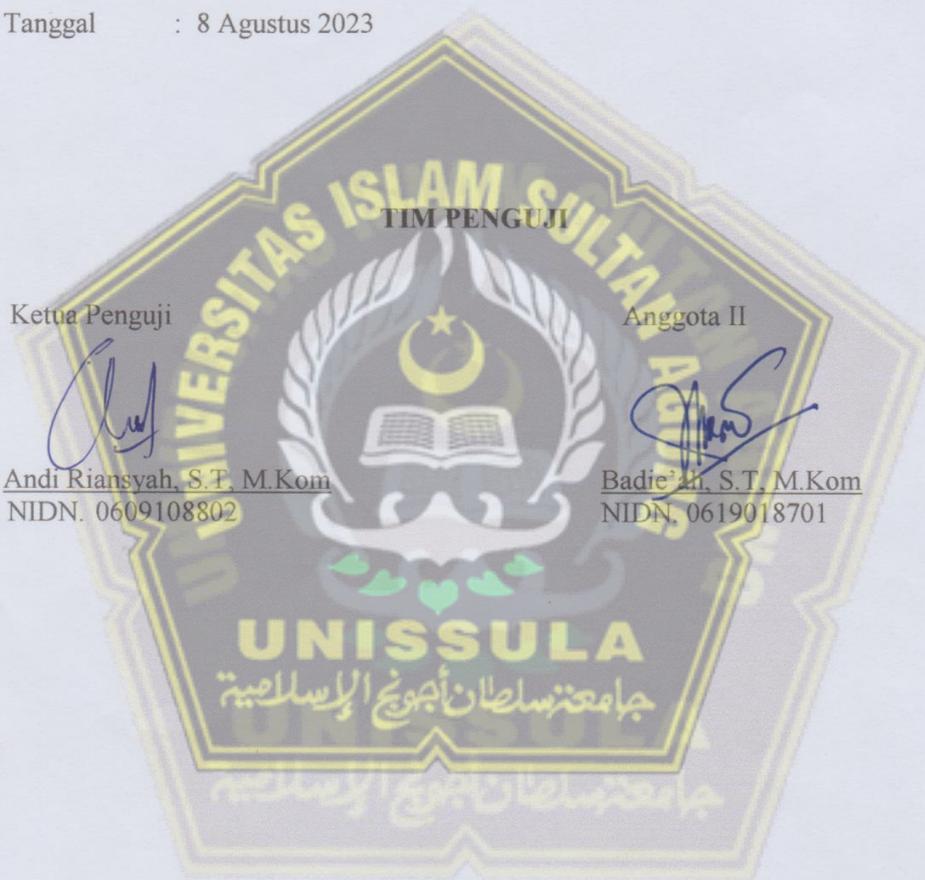
  
Ir. Sri Mulyono, M.Eng  
NIDN. 0626066601

## LEMBAR PENGESAHAN PENGUJI

Laporan tugas akhir dengan judul “SISTEM REKOMENDASI SKINCARE DENGAN METODE KEYWORD EXTRACTION DAN COSINE SIMILARITY” ini telah dipertahankan di depan dosen penguji Tugas Akhir pada:

Hari : Selasa

Tanggal : 8 Agustus 2023



Ketua Penguji

Andi Riansyah, S.T., M.Kom  
NIDN. 0609108802

Anggota II

Badie'ati, S.T., M.Kom  
NIDN. 0619018701

## SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Annisa Amalia Putri

NIM : 32601900006

Judul Tugas Akhir : SISTEM REKOMENDASI SKINCARE DENGAN  
METODE KEYWORD EXTRACTION DAN COSINE  
SIMILARITY

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 2 Agustus 2023

  
Annisa Amalia Putri

## PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Annisa Amalia Putri

NIM : 32601900006

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Alamat Asal : Perum. Taman Batarsari Indah F/81, Mranggem, Demak

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : **SISTEM REKOMENDASI SKINCARE DENGAN METODE KEYWORD EXTRACTION DAN COSINE SIMILARITY**

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan agung.

Semarang, 2 Agustus 2023



Annisa Amalia Putri

## KATA PENGANTAR

Dengan mengucapkan syukur alhamdulillah atas kehadiran Allah SWT yang telah memberikan rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Tugas Akhir dengan judul “Sistem Rekomendasi Skincare Dengan Metode *Keyword Extraction* dan *Cosine Similarity*” ini untuk memenuhi salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar sarjana (S-1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang. Tugas Akhir ini disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terima kasih kepada:

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, S.H., M.H yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Novi Marlyana, S.T., M.T.
3. Dosen pembimbing I penulis Bapak Sam Farisa Chaerul Haviana, S.T., M.Kom yang telah meluangkan waktu dan memberi ilmu.
4. Dosen pembimbing II penulis Bapak Ir. Sri Mulyono, M.Eng yang memberikan banyak nasehat dan saran.
5. Orang tua penulis yang *senantiasa* memberikan semangat serta doa agar tugas akhir ini berjalan dengan lancar,
6. Dan kepada semua pihak yang tidak dapat saya sebutkan satu persatu.

Dengan segala kerendahan hati, penulis menyadari masih terdapat banyak kekurangan dari segi kualitas atau kuantitas maupun dari ilmu pengetahuan dalam penyusunan laporan, sehingga penulis mengharapkan adanya saran dan kritikan yang bersifat membangun demi kesempurnaan laporan ini dan masa mendatang.

Semarang, 25 Juli 2023

Annisa Amalia Putri

## DAFTAR ISI

<b>COVER</b> .....	<b>i</b>
<b>LEMBAR PENGESAHAN PEMBIMBING</b> .....	<b>iii</b>
<b>LEMBAR PENGESAHAN PENGUJI</b> .....	<b>iv</b>
<b>SURAT PERNYATAAN KEASLIAN TUGAS AKHIR</b> .....	<b>v</b>
<b>PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH</b> .....	<b>vi</b>
<b>KATA PENGANTAR</b> .....	<b>vii</b>
<b>DAFTAR ISI</b> .....	<b>viii</b>
<b>DAFTAR TABEL</b> .....	<b>x</b>
<b>DAFTAR GAMBAR</b> .....	<b>xi</b>
<b>ABSTRAK</b> .....	<b>xii</b>
<b>BAB I</b> .....	<b>1</b>
<b>PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah.....	3
1.3 Pembatasan Masalah .....	3
1.4 Tujuan.....	3
1.5 Manfaat.....	4
1.6 Sistematika Penulisan.....	4
<b>BAB II</b> .....	<b>5</b>
<b>TINJAUAN PUSTAKA DAN DASAR TEORI</b> .....	<b>5</b>
2.1 Tinjauan Pustaka.....	5
2.2 Dasar Teori .....	5
2.2.1 <i>Skin Care</i> .....	8
2.2.2 <i>Keyword Extraction</i> .....	9
2.2.3 <i>Cosine Similarity</i> .....	11
2.2.4 Metrik Evaluasi .....	11
<b>BAB III</b> .....	<b>12</b>
<b>METODE PENELITIAN</b> .....	<b>12</b>
3.1 Metode Penelitian.....	12
3.1.1 <i>Study Literature</i> .....	12

3.1.2	<i>Data Collecting</i> .....	12
3.1.3	Perancangan Arsitektur Sistem .....	13
3.1.4	Deskripsi Sistem .....	17
3.1.5	Identifikasi Perangkat Lunak .....	18
3.1.6	Perancangan <i>User Interface</i> .....	20
<b>BAB IV</b> .....		<b>22</b>
<b>HASIL DAN ANALISIS PENELITIAN</b> .....		<b>22</b>
4.1	Implementasi <i>Keyword Extraction</i> .....	22
4.2	Hasil Implementasi Sistem .....	27
4.3	Hasil Pengujian Sistem .....	31
4.4	Validasi Implementasi Algoritma .....	34
<b>BAB V</b> .....		<b>39</b>
<b>KESIMPULAN DAN SARAN</b> .....		<b>39</b>
5.1	Kesimpulan .....	39
5.2	Saran .....	39
<b>DAFTAR PUSTAKA</b> .....		<b>40</b>
<b>LAMPIRAN</b> .....		<b>43</b>



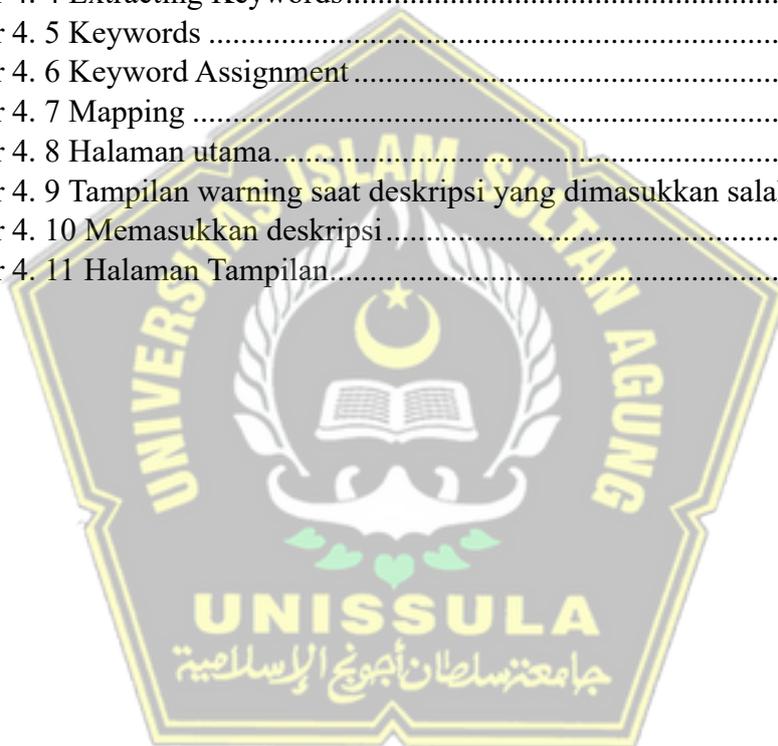
## DAFTAR TABEL

Tabel 3. 1 Contoh perhitungan TF-IDF .....	19
Tabel 4. 1 Pengujian Performa Sistem .....	31
Tabel 4. 2 Pengujian Black Box .....	33
Tabel 4. 3 Hasil Validasi.....	34



## DAFTAR GAMBAR

Gambar 2. 1 Grafik usia pengguna skincare di Indonesia .....	9
Gambar 3. 1 Flowchart Rancangan Arsitektur Sistem .....	14
Gambar 3. 2 Flowchart keyword extraction.....	15
Gambar 3. 3 Rancangan antar muka halaman utama .....	20
Gambar 3. 4 Rancangan antar muka dengan menampilkan hasil rekomendasi....	21
Gambar 4. 1 Case Folding.....	22
Gambar 4. 2 Count Vectorizer.....	23
Gambar 4. 3 TF-IDF Transformer.....	24
Gambar 4. 4 Extracting Keywords.....	24
Gambar 4. 5 Keywords .....	25
Gambar 4. 6 Keyword Assignment.....	26
Gambar 4. 7 Mapping .....	26
Gambar 4. 8 Halaman utama.....	28
Gambar 4. 9 Tampilan warning saat deskripsi yang dimasukkan salah.....	29
Gambar 4. 10 Memasukkan deskripsi.....	30
Gambar 4. 11 Halaman Tampilan.....	30



## ABSTRAK

Perawatan kulit (*skincare*) adalah serangkaian tindakan atau langkah-langkah untuk menjaga dan memelihara kulit yang dilakukan dengan menggunakan produk-produk perawatan kulit dan kosmetik kecantikan dengan bahan yang aman dan cocok untuk jenis kulit wajah masing-masing individu. Oleh karena itu, dibutuhkan sebuah sistem yang mampu memberikan rekomendasi *skincare* yang aman serta baik digunakan sesuai dengan tipe kulit dan permasalahan kulit dengan hasil rekomendasi yang telah disesuaikan untuk penderita alergi bahan kimia tertentu. Dengan begitu pengguna dapat mengatasi permasalahan kulitnya dengan tepat tanpa khawatir timbulnya efek samping. Metode *Keyword Extraction* digunakan untuk mengekstrak *keyword* yang ada pada deskripsi produk, TF-IDF mengekstrak deskripsi produk berdasarkan frekuensi kemunculan kata (*term*) tersebut, semakin tinggi frekuensinya maka semakin besar kemungkinan *keyword* tersebut diambil. Kemudian hasil ekstraksi *keyword* akan dicocokkan dengan inputan *user* menggunakan *cosine similarity*, semakin tinggi *similarity*-nya produk akan berada di daftar paling atas rekomendasi. Sistem ini memberikan rekomendasi dengan rata-rata waktu proses 0,018 detik, *similarity* tertinggi 0,937, dan akurasi 1,0.

Kata Kunci: *Skincare, Keyword Extraction, Cosine Similarity*

## ABSTRACT

*Skin care is a procedure or steps to care for the skin that is carried out using skincare products and beauty cosmetics with ingredients that are safe and good for use according to each individual's facial skin type. Therefore, a system is needed that is able to provide skincare recommendations that are safe and good for use according to skin type and skin problems with recommendations that have been adjusted for sufferers of certain chemical allergies. That way users can deal with skin problems properly without worrying about side effects. The Keyword Extraction method is used to extract keywords in product descriptions, TF-IDF extracts product descriptions based on the frequency of occurrence of the word (term), the higher the frequency, the more likely the keyword is taken. Then the keyword extraction results will be matched with user input using cosine similarity, the higher the similarity the product will be at the top of the list of recommendations. This system provides recommendations with an average processing time of 0.018 seconds, the highest similarity is 0.937, and Accuracy is 1,0.*

*Keywords: Skincare, Keyword Extraction, Cosine Similarity*

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Perawatan kulit, yang dikenal juga sebagai *skincare*, merujuk pada serangkaian tindakan dan langkah-langkah yang diambil untuk menjaga dan memelihara kesehatan serta penampilan kulit wajah. Dalam upaya ini, digunakan berbagai produk *skincare* dan kosmetik kecantikan yang dirancang khusus dengan komposisi bahan yang aman dan sesuai dengan jenis kulit individu. Namun, meskipun pentingnya perawatan kulit diakui oleh banyak orang, sebagian konsumen menghadapi tantangan dalam memilih produk *skincare* yang tepat dan efektif.

Salah satu masalah umum yang sering dihadapi adalah kebingungan dalam memahami kandungan dan manfaat dari berbagai produk *skincare* yang tersedia di pasaran. Kehadiran beragam bahan dan formula yang berbeda membuat konsumen kesulitan menilai produk mana yang cocok untuk jenis kulit mereka. Terkadang, pemilihan produk yang tidak sesuai dengan jenis kulit atau masalah kulit tertentu dapat mengakibatkan reaksi alergi atau perburukan kondisi kulit.

Selain itu, setiap individu memiliki karakteristik kulit yang unik, seperti tingkat sensitivitas, kelembapan, atau kecenderungan terhadap masalah kulit tertentu seperti jerawat atau penuaan dini. Oleh karena itu, penting bagi konsumen untuk memahami kondisi kulit mereka dengan baik agar dapat memilih produk *skincare* yang sesuai dengan kebutuhan kulit mereka. Pemakaian produk perawatan kulit yang tidak benar akan memiliki efek negatif pada kulit wajah. Pada penelitian yang dilakukan untuk pemilihan *skincare* yang sesuai dengan jenis kulit wajah dimana penelitian tersebut menjelaskan bahwa terdapat lebih dari satu jenis kulit wajah yang dimiliki manusia, dan setiap jenis kulit wajah tersebut memiliki *treatment* yang berbeda pula contohnya dalam pemilihan *facial foam* untuk kulit sensitif dianjurkan menggunakan *facial foam* dengan pH rendah untuk mencegah timbulnya iritasi (Maarif dkk., 2019). Kemudian penelitian yang dilakukan untuk penentuan jenis kulit wajah dimana penelitian tersebut menjelaskan bahwa jenis

kulit wajah yang biasanya dimiliki oleh manusia adalah kulit wajah normal, berminyak, kering, sensitif, dan kombinasi (Pebrianto dkk., 2020). Dengan beragamnya jenis kulit wajah pada manusia, maka permasalahan yang muncul pada kulit wajah setiap individu pun beragam mulai dari jerawat, flek hitam, kusam, berkomedo, dan lain sebagainya. Tidak hanya itu, beberapa manusia juga memiliki alergi terhadap bahan kimia tertentu. Dikutip dari tulisan (Raison-Peyron, 2019) pada *website* FDA (*U.S Food and Drug Administration*) produk kosmetik dapat memicu reaksi alergi pada sebagian orang. Reaksi alergi adalah reaksi berlebihan sistem kekebalan terhadap zat yang mungkin tidak berbahaya. Alergen dapat memicu sistem kekebalan tubuh untuk melepaskan zat kimia seperti antibodi yang mengakibatkan gejala alergi. Reaksi alergi terhadap kosmetik paling sering muncul yaitu gatal, ruam merah pada kulit atau dermatitis kontak.

Dari ketiga penelitian diatas dapat diambil kesimpulan bahwa tipe kulit manusia sangat beragam, jika salah memilih produk *skincare* dikhawatirkan akan menimbulkan dampak buruk bagi kesehatan kulit wajah. Dengan beragamnya tipe kulit manusia, permasalahan yang muncul juga beragam seperti jerawat, flek hitam, wajah kusam, berkomedo, sehingga untuk penanganan setiap permasalahan kulit juga harus diperhatikan karena setiap *skincare* memiliki fungsi yang berbeda-beda. Namun, produk kosmetik juga dapat menyebabkan reaksi alergi pada sebagian orang yang menyebabkan munculnya gatal dan ruam.

Oleh karena itu, penelitian ini bertujuan untuk membuat sebuah rancang bangun sistem rekomendasi *skincare* dengan mengambil data *skincare* dari sebuah *website* kecantikan, lalu dengan algoritma TF-IDF yang membantu mengekstraksi kata kunci penting dari deskripsi *skincare* kemudian deskripsi pengguna yang telah diinputkan akan dicocokkan dengan kata kunci yang telah diekstrak menggunakan algoritma *Cosine Similarity*, sehingga dapat memudahkan pengguna mencari *skincare* yang sesuai.

Sistem ini diharapkan dapat memberikan rekomendasi *skincare* dengan tepat sesuai dengan tipe kulit dan permasalahan kulit dengan hasil rekomendasi yang telah disesuaikan. Dengan begitu pengguna dapat mengatasi permasalahan kulitnya dengan tepat tanpa khawatir timbulnya efek samping.

## 1.2 Perumusan Masalah

Belum adanya sistem *filtering* atau rekomendasi *skincare* yang memberikan rekomendasi dengan mempertimbangkan jenis kulit, kebutuhan kulit atau keluhan, serta apakah pengguna memiliki alergi. Sistem ini mengambil data dari *website* Female Daily karena *website* tersebut fungsinya seperti forum diskusi antar pengguna *skincare* dan *rating* yang diberikan berdasarkan pengalaman langsung penggunaannya sehingga diharapkan *skincare* yang direkomendasikan adalah *skincare* terbaik dan cocok digunakan pengguna.

## 1.3 Pembatasan Masalah

Adapun batasan masalah dari penulisan proposal adalah sebagai berikut:

1. Dataset yang digunakan merupakan hasil dari scraping *website* Female Daily dengan link berikut, <https://femaledaily.com/category/skincare> dimana terdapat kumpulan *skincare* dengan berbagai jenis dan merk beserta deskripsi dari setiap *skincare*.
2. Sistem ini menggunakan algoritma TF-IDF hanya untuk mengekstrak kata kunci dari deskripsi produk sehingga tidak sampai ke ranah efek samping obat dan sejenisnya.
3. Sistem ini hanya memberikan hasil rekomendasi *skincare* berdasarkan inputan *user* berupa deskripsi tipe kulit dan kondisi atau tujuan menggunakan *skincare* serta apakah pengguna memiliki alergi atau tidak. Inputan tersebut akan dicocokkan oleh sistem dengan deskripsi produk *skincare* sebagai proses rekomendasinya.

## 1.4 Tujuan

Tujuan tugas akhir ini adalah membangun sistem rekomendasi *skincare* berdasarkan tipe dan kebutuhan kulit yang telah disesuaikan untuk penderita alergi bahan kimia tertentu dengan mengekstrak kata kunci (*keyword*) dari setiap deskripsi *skincare* agar mudah dicari.

## 1.5 Manfaat

Berdasarkan permasalahan dan tujuan penelitian, penelitian ini diharapkan dapat memberikan manfaat yang besar bagi penggunanya yaitu memudahkan dalam mencari *skincare* yang cocok dengan tipe kulit dan tujuan/kondisi kulit dengan mempertimbangkan aspek alergi pengguna agar pengguna dapat mengatasi permasalahan kulitnya dengan tepat tanpa khawatir timbulnya efek samping.

## 1.6 Sistematika Penulisan

Adapun sistematika penulisan yang dipakai dalam pembuatan laporan tugas akhir ini adalah sebagai berikut:

### BAB I : PENDAHULUAN

Pada bab ini berisikan latar belakang dalam pemilihan judul penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

### BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Pada bab ini memuat penelitian-penelitian yang telah dilakukan sebelumnya dan dasar teori yang berhubungan dengan metode *Keyword Extraction* dan *Cosine Similarity* untuk penelitian ini.

### BAB III : METODE PENELITIAN

Pada bab ini menjelaskan proses dan tahapan-tahapan penelitian dimulai dari mendapatkan data hingga proses pengolahan data yang sudah didapat.

### BAB IV : HASIL DAN ANALISIS PENELITIAN

Pada bab ini memaparkan hasil penelitian yang sudah dilakukan yaitu hasil dari proses rekomendasi dengan *Keyword Extraction* menggunakan TF-IDF dan *Cosine Similarity*.

### BAB V : KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan yang menggambarkan bahwa permasalahan telah dipecahkan, tujuan telah tercapai dan manfaat telah didapat. Lalu saran yang berisi kekurangan dari sistem untuk bisa diperbaiki lagi pada penelitian berikutnya.

## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Kulit, sebagai organ terbesar yang melapisi tubuh manusia, memiliki peran yang tak tergantikan dalam menjaga keseimbangan dan kesehatan tubuh. Selain menjadi lapisan pelindung utama dari ancaman eksternal seperti sinar UV, polusi, mikroorganisme, dan cedera fisik, kulit juga berperan dalam regulasi suhu tubuh, sintesis vitamin D, serta pengaturan aliran darah dan cairan tubuh. Kulit manusia memiliki sifat yang luar biasa, di mana kemampuannya untuk meregenerasi diri dan beradaptasi dengan kondisi lingkungan membuatnya mampu bertahan dari segala ancaman. Namun, di tengah kondisi iklim tropis seperti yang ditemui di Indonesia, kulit sering mengalami tantangan tersendiri. Iklim yang panas dan lembap cenderung membuat kulit lebih berminyak dan cenderung berkeringat, yang dapat mengakibatkan masalah kulit seperti jerawat, komedo, atau iritasi. Khususnya pada kulit wajah, kondisi tersebut menjadi lebih mencolok karena kulit wajah memiliki karakteristik yang lebih sensitif dan rentan terhadap perubahan lingkungan. Kombinasi antara kelembapan dan panas membuat pori-pori kulit wajah lebih mudah tersumbat, yang kemudian dapat memicu timbulnya masalah kulit. (Riandari, 2017). Oleh karena itu, dengan membuat sistem rekomendasi *skincare* dapat memudahkan dalam mencari *skincare* untuk mengatasi permasalahan kulit wajah dengan tepat dan aman.

Dampak buruk yang bisa terjadi saat menggunakan *skincare* yang tidak tepat adalah alergi, muncul ruam yang berwarna kemerahan, perih, gatal-gatal bahkan fatalnya bisa mengalami peradangan kulit kronis. Penelitian yang dilakukan untuk memberikan rekomendasi *skincare* menggunakan metode *content-based filtering* dan algoritma apriori dengan sampel data 12 produk *skincare* berbagai macam kategori. Penelitian tersebut memberikan rekomendasi *skincare* berdasarkan inputan *user* berupa jenis kulit, range harga, kategori usia, dan kondisi kulit lalu dicocokkan dengan kriteria yang dimiliki *skincare*, kesimpulan yang didapat adalah sistem rekomendasi *skincare* mampu membantu menyelesaikan permasalahan

pengguna *skincare* dalam pemilihan *skincare* yang cocok, dengan hasil confidence sebesar 88,89%. (Safitri dkk., 2021)

Penelitian yang dilakukan untuk memberikan rekomendasi *skincare* menggunakan metode *collaborative-filtering* dengan data beberapa produk *skincare* Korea yang dipisah menjadi data sampel dan data rekomendasi. Penelitian tersebut memberikan rekomendasi *skincare* berdasarkan nilai rating yang diinput *user* pada sampel produk *skincare*, selanjutnya dilakukan *similarity* antar produk sehingga mendapatkan kesimpulan bahwa prediksi rating produk mendapatkan skor MAE 0,25 yang berarti prediksinya tidak menyimpang jauh dengan data yang sebenarnya. (Ayuningrum dkk., 2022)

Penelitian yang dilakukan untuk pengambilan keputusan pemilihan *skincare* menggunakan 10 data *skincare* jenis *facial wash* dengan *merk* yang berbeda-beda. Penelitian tersebut memberikan rekomendasi *skincare* (*facial wash*) berdasarkan inputan *user* berupa jenis kulit, usia, dan rentang harga mendapatkan kesimpulan bahwa dengan penerapan algoritma *Weighted Product* menunjukkan bahwa hasil perankingan terbesar atau tertinggi akan ditampilkan sebagai rekomendasi terbaik (Haslindah dkk., 2022).

Penelitian yang dilakukan untuk memberikan saran produk Emina Cosmetics menggunakan teknik TF-IDF dan *Cosine Similarity* dengan data 10 produk Emina Cosmetics. Penelitian tersebut proses sarannya dilakukan dengan membandingkan kesamaan antara produk yang dicari pengguna dengan deskripsi produk yang ada. Hasilnya, penelitian ini tidak memerlukan penilaian atau peringkat, dan algoritma similaritas kosinus cocok digunakan pada data yang tidak memiliki nilai subjektif seperti kesamaan antara teks berdasarkan kata-kata yang mirip dalam teks dengan hasil perhitungan similaritas kosinus tertinggi 0,7195. (Larasati & Februriyanti, 2021). Dari keempat penelitian diatas dapat disimpulkan bahwa metode TF-IDF yang diterapkan pada penelitian yang dilakukan oleh Larasati dan Februriyanti dalam metode *Content-Based Filtering* sangat cocok dengan penelitian ini yang akan membuat sistem rekomendasi *skincare* berdasarkan tipe kulit dan kondisi kulit yang sudah tertulis di deskripsi produk *skincare*.

Metode TF-IDF adalah pendekatan untuk menentukan bobot kata dalam dokumen dengan memperhitungkan frekuensi kemunculan kata dalam dokumen (TF) dan kebalikan frekuensi dokumen yang mengandung kata tersebut (IDF). Dengan menggunakan model ruang vektor dan pembobotan TF-IDF, dokumen dapat direpresentasikan sebagai vektor numerik, sehingga kita dapat menghitung kedekatan antar dokumen. Metode ini berguna dalam mengevaluasi pentingnya kata dalam dokumen dan mengidentifikasi kemiripan antar dokumen berdasarkan perhitungan vektornya.

Penelitian yang dilakukan untuk klasifikasi berita online menggunakan TF-IDF dan Cosine Similarity dengan data berita online yang diambil dari Kompas.com. Penelitian tersebut mendapatkan kesimpulan bahwa proses yang ada di dalam TF-IDF berupa *tokenizing*, *stopword* dan *stemming*. *Preprocessing* tersebut dapat memperkecil term sehingga dapat mempercepat proses perhitungan pembobotan term menggunakan TF-IDF dan mempercepat proses *Cosine Similarity*. (Herwijayanti dkk., 2018)

Fungsi ukuran kemiripan (*similarity measure*) digunakan untuk menghitung kemiripan antara dua objek, seperti dokumen. Semakin besar hasil dari fungsi kemiripan tersebut, maka kedua objek yang dievaluasi dianggap semakin mirip. Dalam konteks pencarian informasi, fungsi kemiripan ini memungkinkan perankingan dokumen berdasarkan relevansinya terhadap suatu *query*. Dengan menghitung kemiripan antara *query* dan dokumen-dokumen yang ada, dapat mengidentifikasi dan menampilkan dokumen-dokumen yang paling relevan dengan *query* tersebut di urutan teratas. Penelitian yang dilakukan untuk memberikan rekomendasi seminar dengan 20 data seminar. Penelitian tersebut menggunakan TF-IDF untuk mengambil *keyword* lalu dengan *cosine similarity* sistem dapat memberikan rekomendasi seminar dengan hasil yang sudah bagus dan relevan dengan seminar pembandingan sebagai data uji mendapatkan hasil similarity tertinggi sebesar 49,35% (Yusuf & Cherid, 2020).

Dari keenam metode diatas dapat disimpulkan bahwa untuk memberikan perawatan kulit yang maksimal harus dengan *skincare* yang sesuai dengan tipe kulit dan permasalahan kulit atau tujuan menggunakan *skincare*. Dalam sistem yang

akan dibuat metode *Keyword Extraction* akan berperan dalam pengolahan data mentah hasil dari *scraping website* kecantikan untuk mengambil kata kunci (*keyword*) berupa tipe kulit dan permasalahan kulit atau tujuan menggunakan *skincare* yang sudah tertulis dalam deskripsi produk setiap *skincare*. Untuk metode *cosine similarity* sendiri akan berperan dalam membandingkan dua buah objek data untuk mencari tingkat kemiripan antara inputan *user* dengan *keyword* yang sudah tersedia, yang memiliki kemiripan lebih dari 0.25 akan ditampilkan untuk rekomendasi *skincare*.

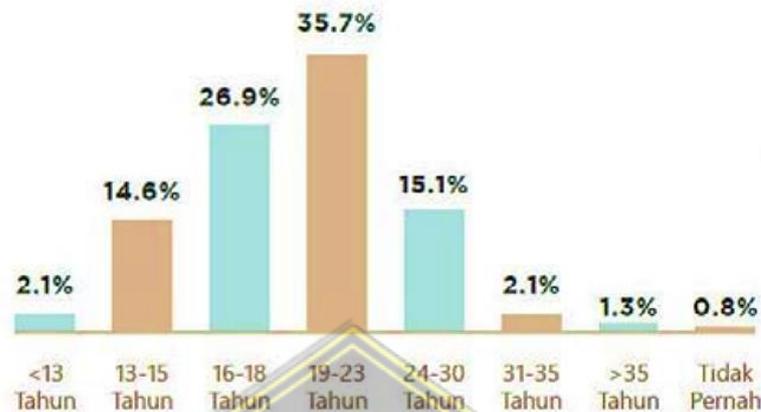
## 2.2 Dasar Teori

### 2.2.1 Skin Care

Upaya untuk mendapatkan kesehatan kulit sudah dilakukan sejak dulu kala, mulai dari perawatan tubuh hingga wajah, hal tersebut dilakukan oleh para wanita dengan meracik bahan-bahan tradisional sendiri sesuai resep turun-temurun. Seiring berkembangnya zaman, prosedur perawatan kulit pun juga berkembang dan semakin canggih. Seperti yang diungkapkan Nindhira dari penjelasan yang disampaikan oleh Prihatini dan Mastawan, faktor yang membuat seorang perempuan merasa lebih percaya diri saat bergaul dengan lingkungannya adalah penampilan secara fisik. Agar selalu terlihat menarik dan indah, diperlukan perawatan kecantikan untuk menunjang penampilan terutama dalam hal kecantikan, meskipun menurut sebagian besar orang lain kecantikan fisik itu bersifat relatif. Dari perspektif ini, perempuan memberikan penilaian bahwa kecantikan menjadi hal yang utama dalam menilai orang lain yang baru dikenal. (Ivada dkk., 2022).

Perawatan kulit, yang dikenal juga sebagai *skincare*, merujuk pada serangkaian tindakan dan langkah-langkah yang diambil untuk menjaga dan memelihara kesehatan serta penampilan kulit wajah. Dalam upaya ini, digunakan berbagai produk *skincare* dan kosmetik kecantikan yang dirancang khusus dengan komposisi bahan yang aman dan sesuai dengan jenis kulit individu. Grafik penggunaan *skincare* dapat dilihat dari survey yang dilakukan MarkPlus, Inc dan ZAP Beauty Index pada Gambar 2.1 yang menunjukkan bahwa pengguna *skincare*

di Indonesia dimulai dari usia 13 tahun, kemudian konsumen tertinggi *skincare* berada di rentang usia 19-23 tahun (Syauki & Amalia Avina, 2020).



Gambar 2. 1 Grafik usia pengguna *skincare* di Indonesia

### 2.2.2 Keyword Extraction

*Keyword extraction* merupakan metode dalam bidang Pemrosesan Bahasa Alami (NLP) yang bertujuan menganalisis teks dengan fokus pada kata kunci yang terdapat dalam teks tersebut. Pendekatan ini membantu dalam menggambarkan isi teks secara ringkas dan mengenali entitas utama dalam teks tersebut. Proses *keyword extraction* dimulai dengan langkah-langkah seperti mengubah huruf menjadi huruf kecil (*case folding*), memecah teks menjadi kata-kata (*tokenizing*), menghapus kata-kata umum (*stopword removal*), dan menghitung jumlah kemunculan kata (*count vectorizer*). Melalui tahapan ini, metode *keyword extraction* mempermudah pemahaman dan analisis teks dengan mengidentifikasi kata-kata kunci yang paling relevan. (Kurniawan, 2017). *Case folding* adalah proses mengubah seluruh huruf menjadi huruf kecil untuk menyamakan huruf agar tidak mempengaruhi analisis pada proses selanjutnya. *Tokenisasi* adalah metode pemisahan kalimat menjadi kata-kata. Penghilangan *stopword* adalah teknik untuk menghapus kata-kata umum yang ada dalam kalimat, seperti "dan", "atau", "juga", dan lain-lain. *Count Vectorizer* menghitung frekuensi kemunculan setiap kata dalam teks dan menghasilkan vektor numerik yang mewakili teks tersebut. Setiap komponen dalam vektor tersebut mewakili jumlah kemunculan kata dalam teks. (Prasetyo dkk., 2021).

*TF-IDF Transformer*, yang merupakan singkatan dari Term Frequency-Inverse Document Frequency, adalah pendekatan yang digunakan untuk memberikan nilai berbobot pada frekuensi token dalam representasi vektor. Metode ini bertujuan untuk mengukur dan menyoroti pentingnya kata-kata dalam teks berdasarkan seberapa sering muncul dalam dokumen serta seberapa jarang muncul dalam seluruh koleksi dokumen. Dengan cara ini, TF-IDF Transformer membantu mengidentifikasi kata-kata yang paling signifikan dalam konteks analisis teks. (Husin, 2023). Frekuensi Istilah (TF) mengukur seberapa sering istilah tertentu muncul dalam teks. Frekuensi istilah sering dibagi dengan panjang teks (jumlah total kata) untuk normalisasi. Rumus TF adalah rasio kejadian istilah dalam dokumen dan jumlah total kata dalam dokumen yang sama seperti pada persamaan (1) di bawah ini.

$$TF(t, d) = \frac{\text{Berapa kali } (t) \text{ muncul dalam } (d)}{\text{Jumlah kata dalam } (d)} \quad (1)$$

IDF mengukur betapa pentingnya kata dalam teks yang diberikan dengan menghitung berat kata-kata langka di semua kalimat dalam teks. Kata-kata yang jarang terjadi dalam teks memiliki skor IDF yang tinggi seperti pada persamaan (2) di bawah ini.

$$IDF(t) = \log_{10} \frac{N}{1 + df} \quad (2)$$

Persamaan (3) merupakan rumus skor TF-IDF untuk sebuah kata didefinisikan sebagai produk dari Term Frequency (TF) dan Inverse Document Frequency (IDF) (Adi dkk., 2018).

$$TF - IDF(d, t) = TF(d, t) * IDF(t) \quad (3)$$

### 2.2.3 Cosine Similarity

*Cosine Similarity* adalah metode dalam bidang text mining yang digunakan untuk mengklasifikasikan dokumen atau teks berdasarkan konsep normalisasi panjang vektor. Algoritma ini membandingkan dua dokumen, yaitu dokumen A dan B, dengan cara menghitung kesamaan arah vektor antara mereka. Dengan menerapkan teknik ini, *Cosine Similarity* membantu mengukur tingkat kemiripan antara dua dokumen berdasarkan orientasi vektor yang menggambarkan representasi teks. (Yusuf & Cherid, 2020). *Cosine Similarity* menggunakan perhitungan angka dari 0 sampai 1. Apabila hasil pengujian kemiripan mendekati angka 0 kemungkinan besar objek yang dibandingkan tidak mirip namun jika hasil pengujian mendekati angka 1 kemungkinan besar objek yang dibandingkan memiliki kemiripan (Iriananda dkk., 2019). Persamaan (4) di bawah ini merupakan rumus yang umum digunakan untuk *Cosine Similarity*:

$$\text{Cos } \alpha = \frac{AB}{|A||B|} + \frac{\sum_{i=1}^n Ai \times Bi}{\sqrt{\sum_{i=1}^n (Ai)^2} \times \sqrt{\sum_{i=1}^n (Bi)^2}} \quad (4)$$

Keterangan:

- A = Vektor A, yang akan dibandingkan kemiripan
- B = Vektor B, yang akan dibandingkan kemiripan
- A.B = Dot product antara vektor A dan vektor B
- |A| = Panjang vektor A
- |B| = Panjang vektor B
- |A||B| = Cross product antara |A| dan |B|

### 2.2.4 Metrik Evaluasi

Metrik evaluasi adalah ukuran yang digunakan untuk mengevaluasi kinerja model atau sistem dalam pemrosesan data atau pembelajaran mesin. Metrik ini dapat membantu untuk mengukur sejauh mana model atau sistem berhasil mencapai tujuannya dan memberikan gambaran kualitas yang dihasilkan. Akurasi, presisi, *Recall*, dan *F1-Score* adalah metrik evaluasi yang umum digunakan untuk

mengukur kinerja model dalam berbagai tugas, terutama dalam klasifikasi. Berikut adalah peran masing-masing metrik:

1. *Accuracy*: *Accuracy* mengukur sejauh mana model berhasil mengklasifikasikan secara benar keseluruhan *instance*, baik yang positif maupun negatif. Ini didapatkan dari hasil pembagian jumlah prediksi yang benar (*True Positives* dan *True Negatives*) dengan total jumlah data yang dievaluasi seperti pada persamaan (5) di bawah ini.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

2. *Precision*: *Precision* mengukur sejauh mana prediksi positif dari model sesuai dengan realitas. Ini didapatkan dari hasil pembagian jumlah prediksi positif yang benar (*True Positives*) dengan total jumlah prediksi positif (*True Positives* dan *False Positives*) seperti pada persamaan (6) di bawah ini.

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

3. *Recall*: *Recall* (juga dikenal sebagai sensitivitas atau *true positive rate*) mengukur sejauh mana model dapat mengenali semua item positif yang sebenarnya. Ini didapatkan dari hasil pembagian jumlah prediksi positif yang benar (*True Positives*) dengan total jumlah item positif sebenarnya (*True Positives* dan *False Negatives*) seperti pada persamaan (7) di bawah ini.

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

4. *F1-Score*: *F1-Score* adalah rata-rata harmonik antara presisi dan *Recall*. Ini menggambarkan keseluruhan tentang kualitas model yang seimbang antara kemampuan model untuk mengidentifikasi *instance* positif (*Recall*) dan menghindari mengklasifikasikan *instance* negatif sebagai positif (*Precision*) seperti pada persamaan (8) di bawah ini.

$$F1\ Score = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

## BAB III METODE PENELITIAN

### 3.1 Metode Penelitian

Dalam penelitian ini, metode atau algoritma yang digunakan adalah *Keyword Extraction* TF-IDF dan *Cosine Similarity*, dimana dengan mengkombinasikan kedua metode ini yaitu *Keyword Extraction* untuk mengambil kata kunci (*keyword*) dari setiap deskripsi *skincare*, kemudian metode *Cosine Similarity* sebagai perhitungan *similarity* antara input yang diberikan *user* dan *keyword* hasil ekstraksi dapat menghasilkan sistem rekomendasi *skincare*. Adapun tahapan yang harus dilakukan dalam penelitian ini, antara lain:

#### 3.1.1 Study Literature

Peneliti meninjau beberapa e-book, artikel, jurnal, hasil penelitian terdahulu seperti tesis dan skripsi, mempelajari teori dan praktik mengenai *Natural Language Processing* (NLP), *Keyword Extraction*, dan *Cosine Similarity* melalui berbagai sumber seperti artikel, jurnal, berbagai situs *website* dan mempelajari karakteristik *skincare* beserta penggunaannya dari berbagai sumber di internet.

#### 3.1.2 Data Collecting

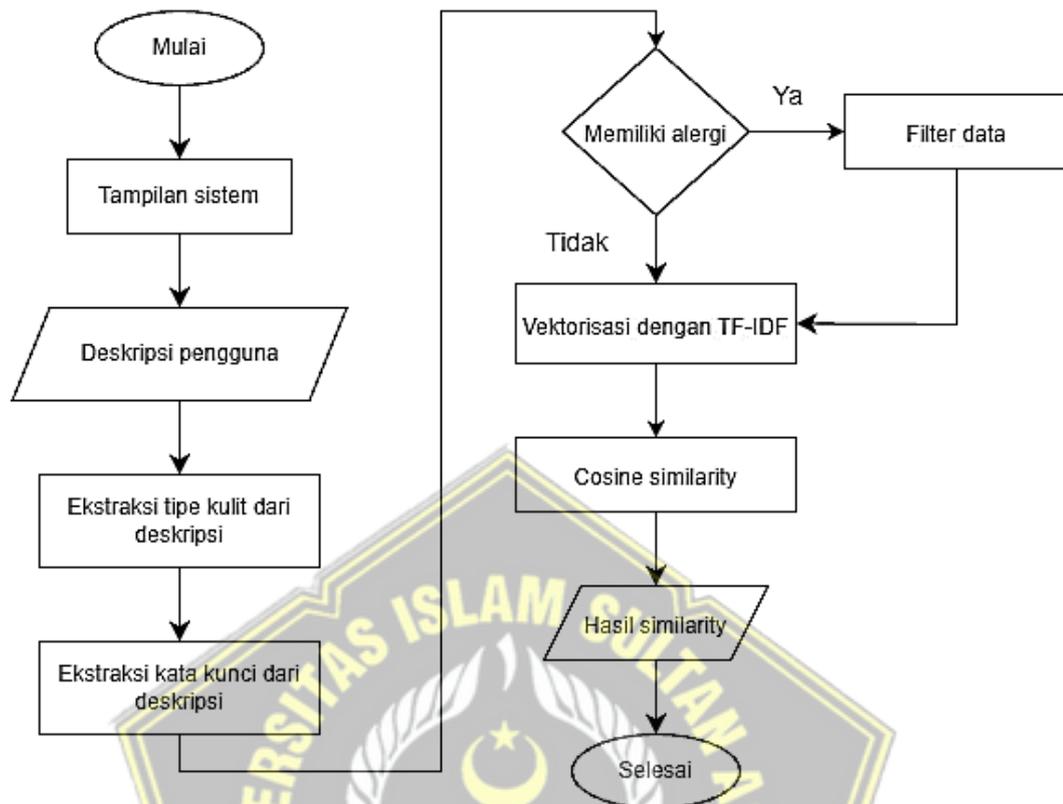
Teknik yang digunakan untuk pengumpulan data dan informasi pada penelitian ini adalah dengan cara *web scraping* situs kecantikan bernama Female Daily dengan *link* sebagai berikut, <https://femaledaily.com/category/skincare>. *Web scraping* adalah proses pengambilan data dari halaman *web* secara otomatis atau terprogram (Levi dkk., 2020). Tujuan dari *web scraping* untuk mendapatkan informasi dari halaman *web* dalam jumlah besar dan dapat diproses secara otomatis. Dalam penelitian ini, *web scraping* dilakukan dengan menggunakan *tools chrome extension* bernama *Web Scraper* kemudian data diekspor menjadi file Excel. Data *skincare* yang diambil berupa rating, kategori, gambar, nama produk, *brand*, deskripsi, dan harga, sedangkan kategori yang diambil adalah *serum*, *eye treatment*, *skin soothing*, *night cream*, *day cream*, *facial wash*, *makeup remover*, *toner*, *peeling*,

*sun protection*, dan *scrub* dengan setiap kategori diambil kurang lebih 20 data. Sehingga jumlah data yang digunakan kurang lebih sekitar 209 data *skincare*. Sebelum data *skincare* digunakan akan dilakukan penggabungan isi field yang nantinya menjadi teks yang akan dicari kata kunci (*keyword*). Untuk data *field* yang digabung adalah data nama produk dan deskripsi produk *skincare*. Data tersebut dilakukan proses *Keyword Extraction* menggunakan TF-IDF untuk dapat mengambil kata kunci (*keyword*) dari deskripsi produk, tipe kulit, dan memberikan label untuk *skincare* yang mengandung bahan allergen.

Untuk mendapatkan data kedua yaitu dari inputan *user* pada *website* berupa deskripsi yang berisi tipe kulit serta tujuan menggunakan *skincare* atau kondisi kulit dan status alergi *user*. Deskripsi tersebut diekstrak setiap kata-katanya jika kata tersebut ada di dalam list yang berisi nilai unik dari kolom *keyword* dan tipe kulit maka akan dimasukkan ke list kosong, lalu masuk ke proses *cosine similarity* untuk mencocokkan deskripsi pengguna dengan deskripsi produk *skincare*, data yang memiliki similaritas lebih dari 0.25 akan ditampilkan pada *website*.

### 3.1.3 Perancangan Arsitektur Sistem

Dalam merancang arsitektur suatu sistem, diperlukan *flowchart* yang menunjukkan langkah-langkah bagaimana sistem berjalan. Pada gambar 3.1 merupakan *flowchart* rancangan sistem yang akan dibangun :



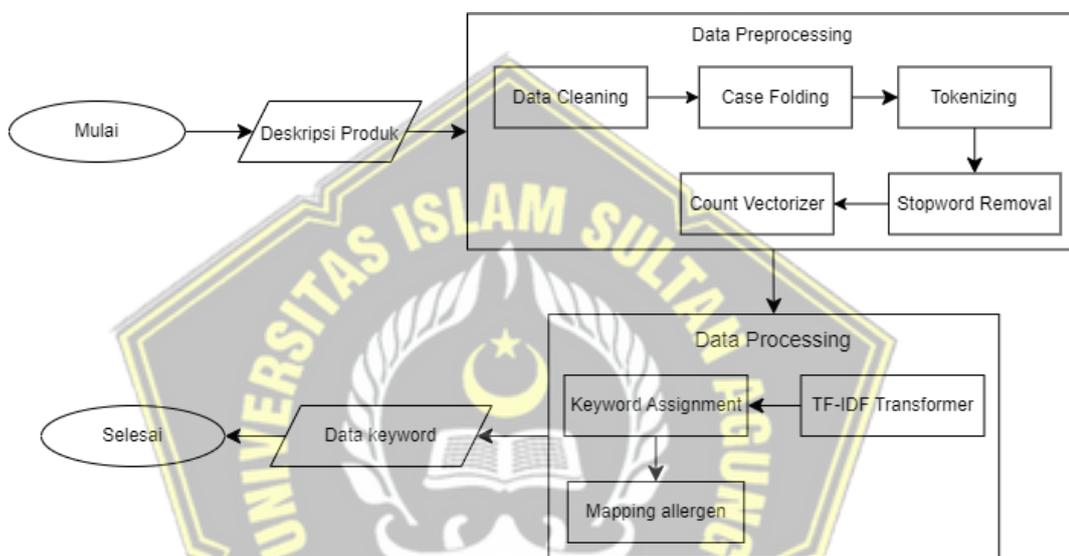
Gambar 3. 1 *Flowchart* Rancangan Arsitektur Sistem

Gambar 3.1 merupakan *flowchart* rancangan arsitektur sistem yang akan dibangun dimana tahapannya adalah:

1. Ekstraksi tipe kulit: Pada tahap ini jenis kulit pengguna diekstraksi dari deskripsi dan diperiksa apakah ada kesesuaian dengan jenis kulit yang ada dalam kolom tipe kulit.
2. Ekstraksi kata kunci: Kata kunci yang relevan diekstraksi dari deskripsi pengguna dan diperiksa apakah ada kesesuaian dengan kata kunci yang ada dalam kolom *keywords*. Jika terdapat kata kunci yang cocok, data akan difilter untuk hanya mempertahankan produk dengan kata kunci yang sesuai.
3. Jika opsi alergen adalah "Ya", data akan difilter untuk hanya mempertahankan produk yang tidak mengandung allergen, begitu pula sebaliknya.
4. TF-IDF Vectorizer: TF-IDF Vectorizer diinisialisasi untuk melakukan vektorisasi TF-IDF pada kolom "*keywords*" dalam data. Data diberikan vektor TF-IDF untuk masing-masing kata kunci produk, dan deskripsi pengguna juga diberikan vektor TF-IDF.

5. *Cosine Similarity* dihitung antara vektor deskripsi pengguna dan vektor kata kunci produk. Hasilnya disimpan dan diurutkan berdasarkan *similarity* secara menurun. Data difilter untuk hanya mempertahankan produk dengan *similarity* di atas 0.1 dan sesuai dengan jenis kulit yang telah ditemukan. Data yang telah difilter dan diurutkan dikembalikan sebagai hasil rekomendasi.

Pada sistem ini *Keyword Extraction* TF-IDF diterapkan untuk data *processing* sebagai langkah pengambilan kata kunci dari deskripsi produk *skincare*.



Gambar 3. 2 Flowchart keyword extraction

Pada gambar 3.2 merupakan *flowchart* penerapan *Keyword Extraction* dengan algoritma TF-IDF dimana tahapannya adalah :

1. Pembersihan data (*Data Cleaning*), pembersihan data berupa penghapusan data yang kosong, tidak lengkap, dan kurang akurat. Hal ini dilakukan agar data yang diolah dapat memberikan hasil yang baik.
2. *Case Folding*, tahap ini mencakup proses lowercasing, penghapusan angka dan karakter yang tidak diperlukan seperti seperti `[!'"#$%&'()*+,- ./:;<=>?@[ \ ] ^ _ ` { } ~ ]`. *Case Folding* berguna untuk menyetarakan kata agar tidak terjadi duplikasi seperti “Kulit” dan “kulit”, kedua kata tersebut memiliki makna yang sama yaitu organ terbesar dalam tubuh dan bagian terluar dari tubuh manusia, namun dapat menjadi dua *term* yang berbeda karena perbedaan kapitalisasi.

3. *Tokenizing* atau tokenisasi adalah pemisahan urutan teks (kalimat) menjadi potongan-potongan, yang disebut token . Token dapat berupa kata, karakter, atau bagian teks lainnya tergantung pada kebutuhan. Misalnya, dalam tokenisasi kata, teks dibagi menjadi kata-kata individu.
4. *Stopword Removal* merupakan tahapan penghapusan kata yang tidak diperlukan namun memiliki frekuensi yang tinggi seperti kata “dan”, “atau”, “untuk”, dsb. Tujuannya adalah untuk mengurangi ukuran kosakata dan meningkatkan kualitas representasi vektor dengan fokus pada kata-kata yang lebih informatif.
5. *Count Vectorizer* merupakan alat yang digunakan untuk mengubah teks menjadi representasi vektor frekuensi token. *Count Vectorizer* akan membangun kosakata dari token-token yang ada dalam teks, menghitung frekuensi kemunculan setiap token dalam setiap dokumen, dan menghasilkan matriks dengan jumlah token sebagai nilai dalam setiap sel (Riadi Silitonga, 2019). Matriks ini mewakili kumpulan teks dalam bentuk yang dapat digunakan oleh model *machine learning* atau NLP.
6. *TF-IDF Transformer (Term Frequency-Inverse Document Frequency)* adalah metode yang digunakan untuk memberikan bobot pada frekuensi token dalam representasi *vector* (Husin, 2023). *Term Frequency (TF)* mengukur seberapa sering suatu istilah terjadi dalam suatu teks. Istilah frekuensi sering dibagi dengan panjang teks (jumlah total kata) sebagai cara normalisasi.
7. *Keyword Assignment* adalah proses pengambilan kata kunci (*keyword*) yang di-assign secara custom atau ditentukan sendiri kata kunci apa yang akan diambil. Dalam penelitian ini menggunakan *keyword assignment* karena banyak tipe kulit yang tidak terambil pada proses *keyword extraction* dikarenakan frekuensinya yang sedikit.

8. *Mapping* adalah proses menghubungkan atau memetakan satu set nilai ke set nilai lainnya. Ini melibatkan pemetaan nilai dari satu kumpulan sumber ke nilai dalam kumpulan target. Dalam penelitian ini terdapat satu set bahan kimia allergen, jika suatu produk mengandung salah satu bahan kimia tersebut maka akan diberi label 'Ya'.

Proses diatas akan menghasilkan data excel yang sudah lengkap dengan kolom *keyword*, tipe kulit, dan allergen. Selanjutnya data tersebut akan menjadi data utama pad *website* yang sudah ditanamkan algoritma *cosine similarity* untuk memberikan rekomendasi yang sesuai dengan deskripsi pengguna.

#### 3.1.4 Deskripsi Sistem

Penelitian ini akan mengembangkan sistem berbasis *web application* untuk memberikan rekomendasi *skincare* yang cocok untuk pengguna sesuai dengan deskripsi yang diinputkan pengguna. Untuk prosesnya, pengguna bisa mendeskripsikan tipe kulit dan kondisi kulit atau tujuan pengguna menggunakan *skincare*, kemudian deskripsi tersebut akan diekstraksi atau diambil kata yang mengandung tipe kulit dan *keyword* untuk dicocokkan dengan data *skincare*, lalu opsi pilihan alergi jika memilih 'Ya' data akan difilter agar menampilkan produk yang tidak mengandung bahan kimia allergen. Data *Skincare* diambil dari *website* Female Daily (<https://femaledaily.com/category/skincare>) dengan cara *web scraping* menggunakan *chrome extension Web Scraper*. Data *skincare* tersebut melalui proses *keyword extraction* untuk mengekstrak kata kunci (*keyword*) dari deskripsi produk. Hasil ekstraksi dari deskripsi pengguna dan data hasil *keyword extraction* dikonversi menjadi vector dengan TF-IDF Vectorizer lalu kedua data tersebut diuji kemiripannya menggunakan metode *Cosine Similarity*. Data yang memiliki *similarity* lebih dari 0.25 akan ditampilkan kepada pengguna sebagai *skincare* yang direkomendasikan oleh sistem.

### 3.1.5 Identifikasi Perangkat Lunak

Pada tahap identifikasi perangkat lunak, peneliti menganalisis apa saja perangkat lunak yang dibutuhkan dalam pengembangan aplikasi ini sehingga proses input hingga mengeluarkan hasil dari rekomendasi yang dapat berjalan sesuai dengan yang diinginkan. Adapun perangkat lunak yang digunakan untuk mengembangkan sistem ini adalah:

1. Python 3.11.3

Python dipilih dalam penelitian ini karena python merupakan bahasa pemrograman tingkat tinggi yang umum digunakan untuk pengembangan perangkat lunak, analisis data, kecerdasan buatan, dan beragam aplikasi lainnya.

2. *Library* Pandas 2.0.1

*Library* Pandas adalah sebuah *library* Python yang digunakan untuk melakukan manipulasi dan analisis data. Pandas menyediakan struktur data dan fungsi yang efisien untuk bekerja dengan data tabular, seperti DataFrames, serta menyediakan berbagai alat untuk membersihkan, memanipulasi, dan menganalisis data.

3. *Library* Re 2023.6.3

*Library* Re adalah *library* Python yang digunakan untuk bekerja dengan ekspresi regular (*regular expression*). Ekspresi regular memungkinkan kita untuk mencocokkan, mencari, dan memanipulasi teks berdasarkan pola tertentu. *Library* Re menyediakan fungsi dan metode yang kuat untuk memanipulasi teks dengan menggunakan ekspresi regular.

4. *Library* Scikit-Learn 1.2.2

*Library* Scikit-Learn adalah *library* Python yang digunakan untuk melakukan pembelajaran mesin (*machine learning*). Pada penelitian ini *library* Scikit-Learn digunakan untuk proses *Cosine Similarity* dan TF-IDF Vectorizer. Selain itu, *library* ini juga menyediakan alat untuk melakukan pemrosesan data, evaluasi model, dan pemilihan fitur. Berikut adalah contoh dari TF-IDF Vectorizer:

A = "Dokumen ini adalah contoh dokumen pertama."

B = "Contoh dokumen kedua ini sama panjang."

Tabel 3. 1 Contoh perhitungan TF-IDF

Kata	TF		IDF	TF*IDF	
	A	B		A	B
adalah	1/6	0	$\text{Log}(2/1) = 0.3$	0.05	0
contoh	1/6	1/6	$\log(2/2) = 0$	0	0
dokumen	2/6	1/6	$\log(3/2) = 0.17$	0.056	0
ini	1/6	1/6	$\log(2/2) = 0$	0	0
kedua	0	1/6	$\log(2/1) = 0.3$	0	0.05
panjang	0	1/6	$\log(2/1) = 0.3$	0	0.05
pertama	1/6	0	$\log(2/1) = 0.3$	0.05	0
sama	0	1/6	$\log(2/1) = 0.3$	0	0.05

Tabel 3.1 merupakan proses perhitungan TF-IDF dari 2 dokumen. Kalimat pada Dokumen A dan Dokumen B dijadikan huruf kecil (*lowercase*) lalu dipecah menjadi kata-kata kemudian dihitung nilai TF-IDF dari masing-masing kata.

#### 5. Streamlit 1.23.1

Streamlit merupakan kerangka kerja Python yang bersifat terbuka. Streamlit memiliki kapabilitas untuk mengubah skrip menjadi aplikasi web yang dapat ditampilkan dalam waktu singkat. Kelebihan ini memudahkan dalam pembuatan aplikasi berbasis web, terutama bagi mereka yang memiliki pengetahuan terbatas tentang pengembangan tampilan depan web. (Fikri dkk., 2023).

#### 6. Jupyter Notebook

Jupyter Notebook adalah sebuah aplikasi yang digunakan untuk membuat dan berbagi dokumen yang menggabungkan teks, kode, visualisasi, dan hasil eksekusi. Jupyter Notebook mendukung bahasa pemrograman Python dan banyak bahasa pemrograman lainnya.

#### 7. Visual Studio Code

Visual Studio Code (VS Code) adalah sebuah editor kode populer dan gratis yang dikembangkan oleh Microsoft. VS Code menyediakan fitur-fitur yang kaya untuk pengembangan perangkat lunak, seperti *highlighting syntax*, otomatisasi penyelesaian kode, *debugging*, dan integrasi.

### 3.1.6 Perancangan *User Interface*

Berikut ini merupakan rancangan/ design system yang digunakan pada penelitian ini :

1. Halaman utama

Pada gambar 3.2 merupakan perancangan antar muka untuk halaman utama dimana halaman utama adalah halaman awal yang akan dilihat oleh pengguna.

Gambar 3. 3 Rancangan antar muka halaman utama

Gambar 3.2 merupakan rancangan antar muka halaman utama dimana halaman inilah yang akan dilihat pertama kali oleh pengguna. Di halaman utama ini nantinya pengguna akan melihat header bertuliskan “SISTEM REKOMENDASI *SKINCARE*”, kemudian form *input text* untuk menuliskan deskripsi pengguna dilengkapi dengan *help button* untuk memberikan contoh inputan deskripsi kepada pengguna, dan *radio button* untuk pengguna yang memiliki alergi. Setelah mengisi form, pengguna dapat meng-klik *button* rekomendasi untuk melihat *skincare* apa saja yang direkomendasikan oleh sistem.

## 2. Halaman rekomendasi

Pada gambar 3.4 merupakan rancangan halaman antar muka untuk halaman hasil rekomendasi.

**SISTEM REKOMENDASI SKINCARE**

serum x eye treatment x toner x

Deskripsikan jenis kulit dan keluhan atau tujuan penggunaan skincare

kulit kering berjerawat noda hitam

Apakah memiliki alergi?

Ya

Tidak

Rekomendasi

Rating	Kategori	Gambar	Produk	Brand	Deskripsi	Harga
4	serum		Niacinamide 10% + Zinc 1%	The Ordinary	The Ordinary Niacinamide 10% + Zinc 1% menyamarkan noda bekas jerawat pada wajah serta menyeimbangkan kadar sebum pada wajah.	Rp. 157.000
4.9	eye treatment		Marigold Eye Power Serum Concentrate	NPURE	Perawatan mata anti kerut alami dengan 8x vitamin esensial 5x peptida dan korektor lingkaran hitam alami untuk membantu mengurangi munculnya kerutan garis halus lingkaran hitam dan kantung mata	Rp. 145.000
4	toner		AHA/BHA Clarifying Treatment Toner	COSRX	Cosrx AHA/BHA Clarifying Treatment Toner merupakan toner dengan bahan-bahan natural dan air mineral. Produk ini lebih dari toner biasa karena memiliki kandungan AHA/BHA organik yang dapat berfungsi sebagai exfoliating toner.	Rp. 190.000

Gambar 3. 4 Rancangan antar muka dengan menampilkan hasil rekomendasi

Gambar 3.4 merupakan rancangan antar muka yang menampilkan hasil rekomendasi untuk pengguna setelah mengisi form inputan. Terdapat tabel yang berisi rekomendasi *skincare* dengan kolom rating, gambar, produk, brand, deskripsi, dan harga.

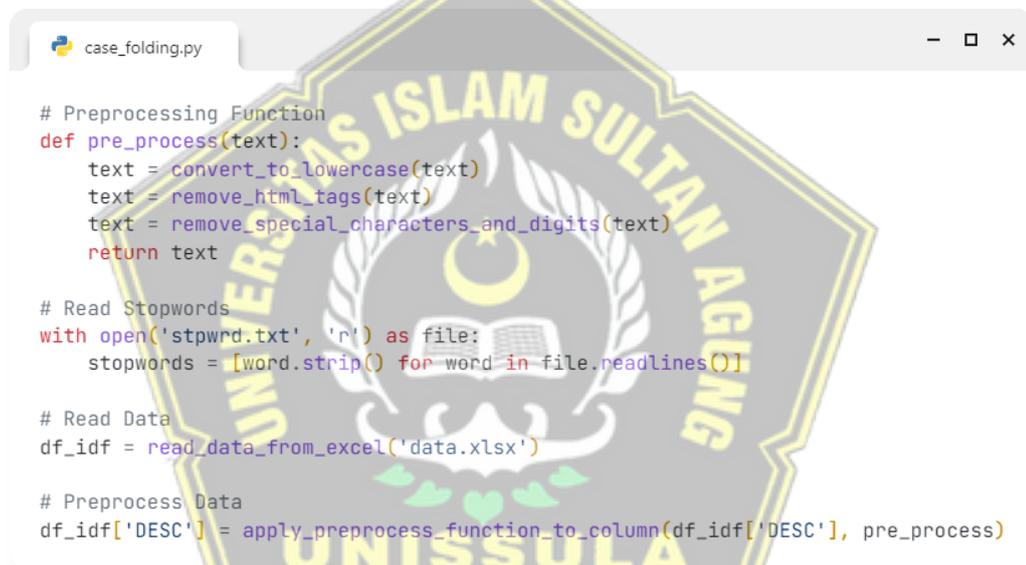
## BAB IV

### HASIL DAN ANALISIS PENELITIAN

#### 4.1 Implementasi *Keyword Extraction*

##### 1. *Case Folding*

*Case Folding* berguna untuk menyetarakan kata agar tidak terjadi duplikasi seperti “Kulit” dan “kulit”, kedua kata tersebut memiliki makna yang sama yaitu organ terbesar dalam tubuh dan bagian terluar dari tubuh manusia, namun dapat menjadi dua term yang berbeda karena perbedaan kapitalisasi.



```
case_folding.py
# Preprocessing Function
def pre_process(text):
    text = convert_to_lowercase(text)
    text = remove_html_tags(text)
    text = remove_special_characters_and_digits(text)
    return text

# Read Stopwords
with open('stopwr.txt', 'r') as file:
    stopwords = [word.strip() for word in file.readlines()]

# Read Data
df_idf = read_data_from_excel('data.xlsx')

# Preprocess Data
df_idf['DESC'] = apply_preprocess_function_to_column(df_idf['DESC'], pre_process)
```

Gambar 4. 1 *Case Folding*

Gambar 4.1 merupakan tampilan *source code* untuk *case folding*. Dimana yang pertama dilakukan adalah membuat sebuah fungsi untuk membersihkan data, kemudian memasukkan data yang akan dibersihkan seperti di bawah ini:

- *Preprocessing Function* (*pre\_process*):  
Ini adalah fungsi pra-pemrosesan yang diterapkan pada setiap teks. Fungsi ini melibatkan beberapa langkah seperti mengonversi teks menjadi huruf kecil, menghapus tag HTML, dan menghilangkan karakter khusus dan digit dari teks.

- *Read Stopwords:*  
Membaca daftar kata-kata *stopwords* dari file "stpwrtd.txt". *Stopwords* adalah kata-kata umum yang sering diabaikan dalam analisis teks karena kurangnya informasi semantik.
- *Read Data:*  
Membaca data dari file *Excel* "data.xlsx" menggunakan fungsi `read_data_from_excel`. Ini menghasilkan *DataFrame* `df_idf` yang berisi kolom 'DESC' yang berisi teks yang akan diproses.

## 2. *Count Vectorizer*

*Count Vectorizer* merupakan alat yang digunakan untuk mengubah teks menjadi representasi vektor frekuensi token.



```

tokenizing_and_cv.py
# Tokenization and Count Vectorization
docs = extract_text_column_as_list(df_idf['DESC'])
word_count_vector = create_count_vector_with_options(docs, min_df=0.07, stopwords=stopwords, ngram_range=(1, 2))

```

Gambar 4. 2 *Count Vectorizer*

Gambar 4.2 merupakan proses *Count Vectorizer* dimana langkah awal yang dilakukan adalah menerapkan fungsi pra-pemrosesan (`pre_process`) pada kolom 'DESC' dari `df_idf` menggunakan fungsi `apply_preprocess_function_to_column`. Kemudian mengekstrak kolom 'DESC' dari `df_idf` sebagai daftar teks. Membuat vektor hitungan (*count vector*) dengan opsi seperti batas minimum (`min_df`), *stopwords*, dan rentang n-gram (`ngram_range`).

- `min_df=0.07`: Mengatur batas bawah frekuensi kata dalam dokumen minimal 7% kemunculannya dalam dokumen.
- `stop_words=stopwords`: Menggunakan list *stopwords* yang telah dibuat sebelumnya untuk menghapus kata-kata yang tidak relevan.
- `ngram_range=(1,2)`: Mengatur rentang n-gram yaitu unigram, dan bigram sehingga kombinasi kata-kata hingga tiga kata akan menjadi fitur.

### 3. TF-IDF Transformer

TF-IDF Transformer (*Term Frequency-Inverse Document Frequency*) adalah metode yang digunakan untuk memberikan bobot pada frekuensi token dalam representasi vector.



```

tf-idf.py
# Calculate TF-IDF
tfidf_transformer = create_tfidf_transformer()
tfidf_transformer.fit(word_count_vector)
tfidf_matrix = transform_count_vector_to_tfidf(word_count_vector, tfidf_transformer)

```

Gambar 4. 3 *TF-IDF Transformer*

Gambar 4.3 merupakan tahapan TF-IDF Transformer dengan langkah awal membuat objek `tfidf_transformer` sebagai transformasi TF-IDF. Lalu melatih transformasi ini pada `word_count_vector` dengan menggunakan metode `fit`. Kemudian, mengubah `word_count_vector` menjadi matriks TF-IDF menggunakan metode `transform`.



```

test_data.py
# Load Test Data
df_test = read_data_from_excel('sc-price.xlsx')
df_test['text'] = apply_preprocess_function_to_column(df_test['DESC'], pre_process)

# Preprocess Test Data
docs_test = extract_text_column_as_list(df_test['DESC'])

# Calculate TF-IDF Vector for Test Data
tf_idf_vector = calculate_tfidf_vector_for_test_data(docs_test, tfidf_transformer)

```

Gambar 4. 4 *Extracting Keywords*

Gambar 4.4 adalah langkah awal dan pendefinisian fungsi untuk *Keyword Extraction* menggunakan model TF-IDF yang telah dibuat sebelumnya. Berikut adalah penjelasan langkah-langkah yang dilakukan oleh kode tersebut:

- *Load Test Data:*  
Membaca data dari file Excel "data.xlsx" dan menambahkan kolom 'text' yang berisi teks yang akan diuji.

- *Preprocess Test Data:*  
Menerapkan fungsi pra-pemrosesan (`pre_process`) pada kolom 'text' dari `df_test` menggunakan `apply_preprocess_function_to_column`.
- *Calculate TF-IDF Vector for Test Data:*  
Ekstrak kolom 'text' dari `df_test` sebagai daftar teks. Kemudian menghitung vektor TF-IDF untuk data uji menggunakan `tfidf_transformer` yang telah dilatih sebelumnya.



```
# Extract Keywords for Test Data
results = []
for i in range(tfidf_vector.shape[0]):
    curr_vector = get_vector_for_single_document(tfidf_vector, i)
    sorted_items = sort_tfidf_vector(curr_vector)
    keywords = extract_top_keywords_from_vector(feature_names, sorted_items, topn=10)
    results.append(keywords)

# Create DataFrame with Keywords
df_keywords = create_dataframe_with_keywords(docs, results)
```

Gambar 4.5 *Keywords*

Gambar 4.5 merupakan proses transformasi TF-IDF pada dokumen uji yang dimulai dengan terasi melalui matriks TF-IDF vektor data uji dan ekstrak kata kunci berdasarkan skor TF-IDF tertinggi menggunakan fungsi-fungsi seperti `sort_tfidf_vector` dan `extract_top_keywords_from_vector`. Terakhir, membuat DataFrame `df_keywords` yang berisi daftar teks dan kata kunci yang diekstraksi untuk setiap dokumen.

#### 4. *Keyword Assignment*

*Keyword Assignment* adalah proses pengambilan kata kunci (*keyword*) yang di-assign secara *custom* atau ditentukan sendiri kata kunci apa yang akan diambil.



```

assign_keyword_skin_type.py
# Function to Assign Keywords to Data
def assign_keywords(data, keyword_column, keywords):
    data[keyword_column] = data["DESC"].apply(lambda x: [word for word in x.split() if word.lower() in keywords])
    return data

# Read Data from Excel
data = read_data_from_excel('data.xlsx')

# Define Keyword Column Name and Keywords List
keyword_column = "skin_type"
keywords = ["kering", "berminyak", "sensitif", "normal", "kombinasi", "semua"]

# Assign Keywords to Data
data = assign_keywords(data, keyword_column, keywords)

```

Gambar 4. 6 *Keyword Assignment*

Gambar 4.6 merupakan proses *Keyword Assignment* dengan langkah awal mendefinisikan sebuah fungsi bernama `assign_keywords` yang menerima input data, kolom *keyword*, dan daftar kata kunci. Dalam fungsi tersebut, dilakukan pembacaan data dari file Excel kemudian fungsi `assign_keywords` dipanggil dengan argumen data, nama kolom *keyword* (`keyword_column`), dan daftar kata kunci (`keywords`). Di dalam fungsi `assign_keywords`, dilakukan pemrosesan pada kolom "DESC" menggunakan metode `apply` dan fungsi lambda. Hanya kata-kata yang ada dalam daftar kata kunci yang akan dimasukkan ke dalam kolom *keyword*. Setelah selesai, data yang telah dilengkapi dengan kata kunci akan dikembalikan dan disimpan kembali ke variabel data.

#### 5. *Mapping*

*Mapping* menghubungkan atau memetakan satu set nilai ke set nilai lainnya dengan pemetaan nilai dari satu kumpulan sumber ke nilai dalam kumpulan target.



```

mapping.py
# Split 'DESC' Column and Process Allergen Information
key = df_idf['DESC'].str.split('Ingredients:\n', expand=True)

key[1] = key[1].str.lower().replace({r"[.:(){}/*]":""}, regex=True)

key['allergen'] = key[1].str.contains('benzyl alcohol|benzyl salicylate|citraal|citronelol').map({True: 'Ya', False: 'Tidak'})

```

Gambar 4. 7 *Mapping*

Gambar 4.7 merupakan proses mapping komposisi yang mengandung bahan kimia allergen. Langkah pertama yang dilakukan adalah memisahkan komposisi dari deskripsi produk ke variabel baru bernama `key`. Kemudian komposisi dilakukan pembersihan karakter dan *lowercasing*. Selanjutnya proses *mapping* dilakukan dengan memasukkan nama bahan kimia apa saja yang mengandung allergen, dengan metode map nilai boolean tersebut diubah menjadi 'Ya' jika bernilai True, dan 'Tidak' jika bernilai False. Hasilnya akan disimpan dalam kolom baru bernama 'allergen' pada variabel `key`. Terakhir, dilakukan penggabungan dataframe `df` dengan data frame `key['allergen']`. Penggabungan dilakukan berdasarkan indeks (baris) pada kedua dataframe.

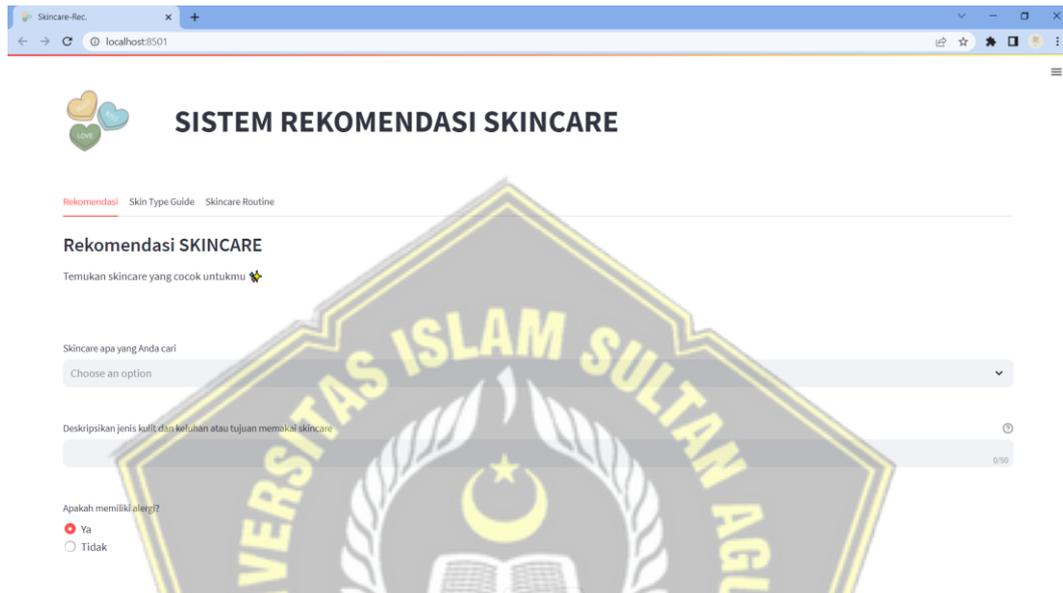
#### 4.2 Hasil Implementasi Sistem

Sistem ini menggunakan algoritma TF-IDF untuk mengekstraksi *keyword* yang ada pada kolom deskripsi produk. TF-IDF akan mengambil kata kunci yang frekuensinya tinggi pada dokumen tersebut, dengan syarat kata kunci tersebut juga harus disebut berulang kali pada dokumen lainnya. Jadi, hanya kata kunci penting yang dapat terekstraksi yaitu kata kunci yang sering disebutkan.

Kemudian algoritma *Cosine Similarity* berperan pada *website* untuk mencocokkan deskripsi pengguna dengan *keyword* yang telah diekstraksi. *Cosine Similarity* bekerja dengan membandingkan 2 vektor yaitu deskripsi pengguna dan *keyword* yang sudah dikonversi ke *vector* menggunakan `TfidfVectorizer`. Kemudian akan menghasilkan nilai *similarity* dan yang akan ditampilkan adalah produk dengan *similarity* lebih dari 0,25. Berikut merupakan hasil implementasi sistem berupa *website* yang dibangun menggunakan Streamlit.

## 1. Halaman Utama

Halaman utama dalam *web* adalah halaman pertama yang muncul ketika pengguna mengunjungi sebuah situs *web*. Halaman utama umumnya berfungsi sebagai titik awal navigasi di situs *web* dan memberikan gambaran singkat tentang konten, tujuan, atau layanan yang disediakan oleh situs tersebut.

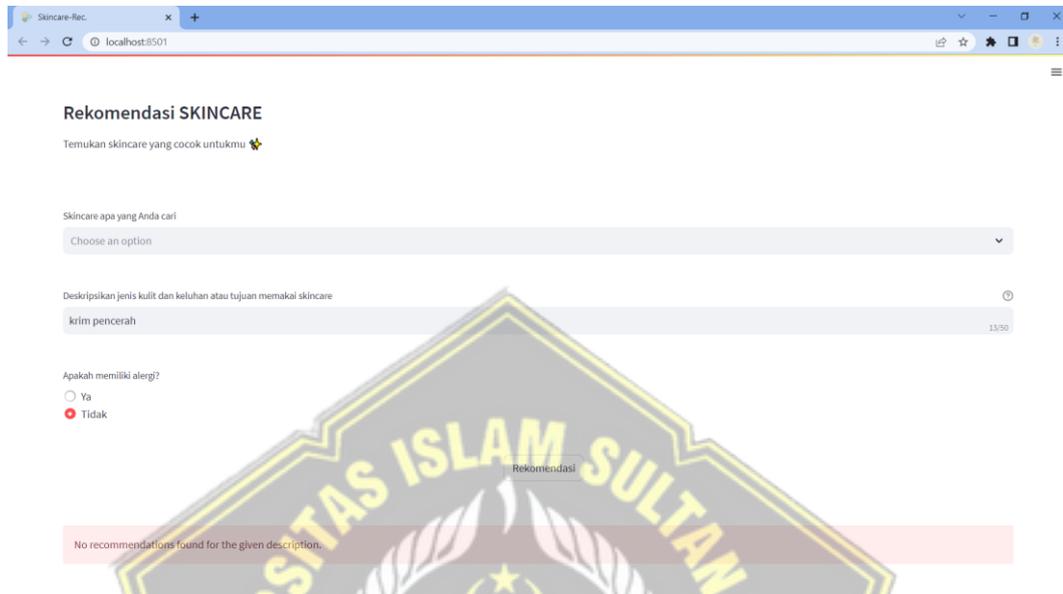


Gambar 4. 8 Halaman utama

Gambar 4.8 merupakan tampilan halaman utama dimana tampilan tersebut yang dilihat pengguna saat pertama kali membuka halaman *web* dengan url <https://skincare-rec.streamlit.app>. Halaman ini berisi judul tugas akhir yaitu “SISTEM REKOMENDASI SKINCARE” yang memiliki 3 tab berbeda dimana tab pertama untuk mencari rekomendasi skincare, tab kedua dan ketiga berisi informasi umum tentang cara menentukan jenis kulit wajah dan panduan penggunaan *skincare*. Kemudian terdapat *filter* kategori produk dimana pengguna bisa memilih lebih dari satu kategori produk untuk ditampilkan, lalu *form* inputan deskripsi pengguna dilengkapi dengan *help button* yang berisi contoh inputan, *radio button* sebagai opsi untuk pengguna yang memiliki alergi, dan *button* “Rekomendasi” untuk memproses inputan.

## 2. Halaman saat deskripsi yang dimasukkan salah

Di halaman yang sama seperti halaman utama akan memunculkan notifikasi bahwa deskripsi yang dimasukkan salah.

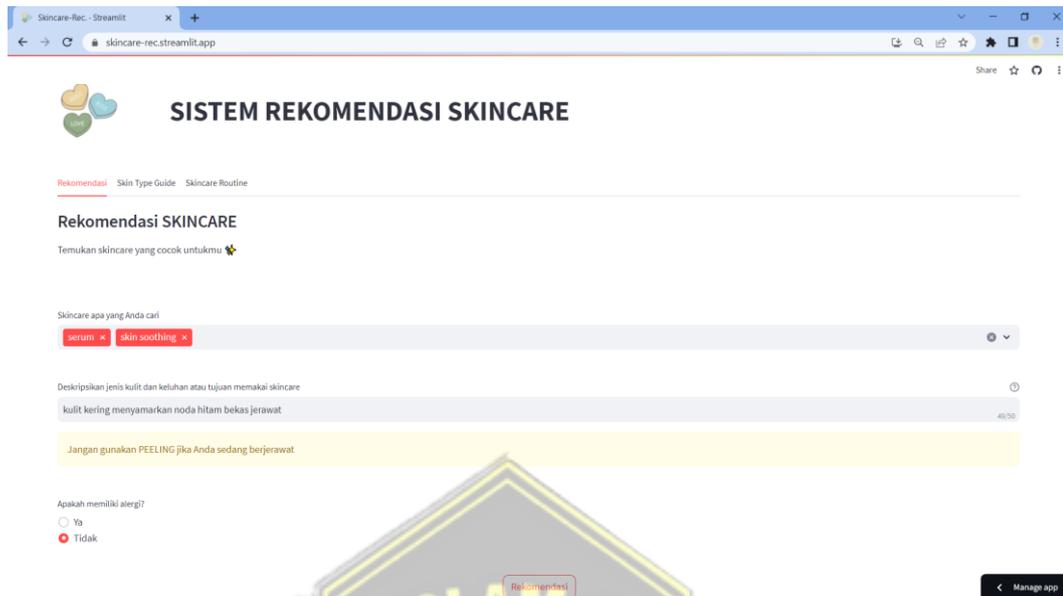


Gambar 4. 9 Tampilan *warning* saat deskripsi yang dimasukkan salah

Gambar 4.9 merupakan tampilan halaman disaat sistem tidak menerima input deskripsi yang sesuai dengan format yaitu harus disertai dengan tipe kulit. Sistem akan memunculkan *warning* bahwa tidak ada rekomendasi yang cocok untuk deskripsi yang diberikan.

## 3. Halaman Rekomendasi

Di halaman yang sama seperti halaman utama akan memunculkan hasil rekomendasi jika inputan yang dimasukkan *user* sudah benar. Pada Gambar 4.10 merupakan contoh inputan dengan *query* “kulit kering menyamarkan noda hitam bekas jerawat” tidak memiliki alergi dan menggunakan *filter* untuk melihat rekomendasi produk kategori serum dan *skin soothing*.



Gambar 4. 10 Memasukkan deskripsi



Gambar 4. 11 Halaman Tampilan

Gambar 4.11 merupakan tampilan ketika pengguna sudah memberikan input yang sesuai. Sistem menampilkan rekomendasi yang cocok dengan deskripsi pengguna dimana produk akan diurutkan secara *descending* berdasarkan similaritas. Terdapat notifikasi berhasil dilengkapi dengan *timer* sistem melakukan pencarian rekomendasi, serta jumlah produk yang direkomendasikan. Tampilan

tabel berisi kolom *rating*, kategori, gambar, produk, *brand*, deskripsi, harga, dan *similarity*.

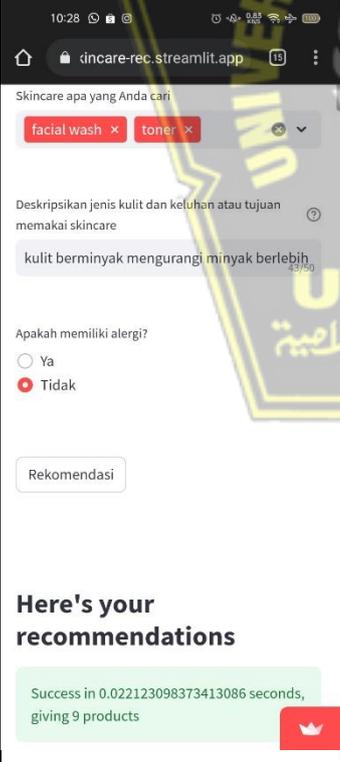
### 4.3 Hasil Pengujian Sistem

Ada dua macam pengujian yang dilakukan untuk menguji sistem rekomendasi lowongan *skincare* antara lain pengujian performa sistem dan pengujian Black Box. Berikut merupakan rincian dari hasil pengujian sistem rekomendasi *skincare*:

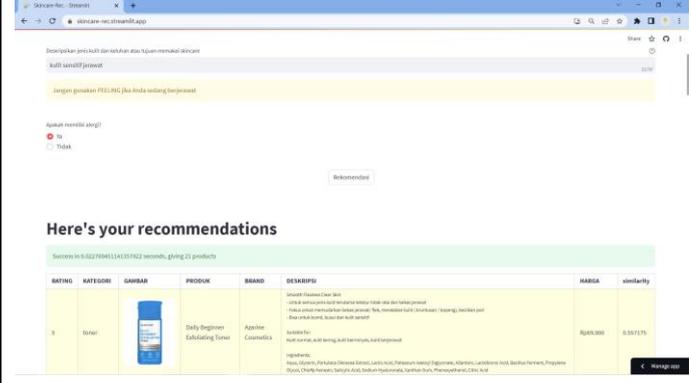
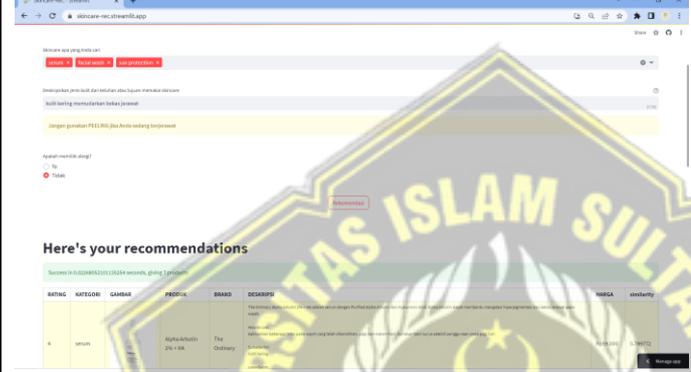
#### 1. Pengujian Performa Sistem

Pada tabel 4.1 merupakan hasil dari pengujian performa dari sistem rekomendasi *skincare* yang meliputi *input*, *output*, waktu yang dibutuhkan sistem serta hasil yang diberikan sistem, dan *similarity* produk yang direkomendasikan.

Tabel 4. 1 Pengujian Performa Sistem

Input	Output	Performa	
		Waktu & Produk	<i>Similarity</i> tertinggi
		0,022 detik / 9 produk	0,714

	<table border="1"> <thead> <tr> <th>DESKRIPSI</th> <th>HARGA</th> <th>similarity</th> </tr> </thead> <tbody> <tr> <td>Azarine Cosmetics C White Lightening Serum adalah serum dengan kandungan tinggi bahan aktif alami untuk whitening. Campuran kakadu dan lainnya mencerahkan kulit termasuk flek hitam pada kulit. Diperkaya asam amino dan hyaluronic acid untuk mengencangkan, mengembalikan dan melembakan hingga lapisan terdalam kulit. Serum diagen ini juga membantu proses pembentukan kolagen pada kulit dan mengurangi</td> <td>Rp220.000</td> <td>0.560694</td> </tr> </tbody> </table>	DESKRIPSI	HARGA	similarity	Azarine Cosmetics C White Lightening Serum adalah serum dengan kandungan tinggi bahan aktif alami untuk whitening. Campuran kakadu dan lainnya mencerahkan kulit termasuk flek hitam pada kulit. Diperkaya asam amino dan hyaluronic acid untuk mengencangkan, mengembalikan dan melembakan hingga lapisan terdalam kulit. Serum diagen ini juga membantu proses pembentukan kolagen pada kulit dan mengurangi	Rp220.000	0.560694	<p>0,013 detik / 13 produk</p>	<p>0,561</p>
DESKRIPSI	HARGA	similarity							
Azarine Cosmetics C White Lightening Serum adalah serum dengan kandungan tinggi bahan aktif alami untuk whitening. Campuran kakadu dan lainnya mencerahkan kulit termasuk flek hitam pada kulit. Diperkaya asam amino dan hyaluronic acid untuk mengencangkan, mengembalikan dan melembakan hingga lapisan terdalam kulit. Serum diagen ini juga membantu proses pembentukan kolagen pada kulit dan mengurangi	Rp220.000	0.560694							
	<table border="1"> <thead> <tr> <th>DESKRIPSI</th> <th>HARGA</th> <th>similarity</th> </tr> </thead> <tbody> <tr> <td>Sabi Beet Brightening Serum merupakan serum yang tu kulit tampak lebih anahmbat ukan pigmen kulit yamarkan noda nyembuhkan luka, bkan, menghaluskan juga anti aging. dua varian yaitu 5% ide + Moisture Sabi m dan 10% ide + Moisture Sabi Brightening Serum. amide + Moisture : Serum cocok untuk ang takut mencoba% % gi, dan orang yang toleransi % rendah ahan aktif, terutama ide.</td> <td>Rp170.000</td> <td>0.719678</td> </tr> </tbody> </table>	DESKRIPSI	HARGA	similarity	Sabi Beet Brightening Serum merupakan serum yang tu kulit tampak lebih anahmbat ukan pigmen kulit yamarkan noda nyembuhkan luka, bkan, menghaluskan juga anti aging. dua varian yaitu 5% ide + Moisture Sabi m dan 10% ide + Moisture Sabi Brightening Serum. amide + Moisture : Serum cocok untuk ang takut mencoba% % gi, dan orang yang toleransi % rendah ahan aktif, terutama ide.	Rp170.000	0.719678	<p>0,013 detik / 21 produk</p>	<p>0,719</p>
DESKRIPSI	HARGA	similarity							
Sabi Beet Brightening Serum merupakan serum yang tu kulit tampak lebih anahmbat ukan pigmen kulit yamarkan noda nyembuhkan luka, bkan, menghaluskan juga anti aging. dua varian yaitu 5% ide + Moisture Sabi m dan 10% ide + Moisture Sabi Brightening Serum. amide + Moisture : Serum cocok untuk ang takut mencoba% % gi, dan orang yang toleransi % rendah ahan aktif, terutama ide.	Rp170.000	0.719678							

	<p>0,022 detik / 21 produk</p>	<p>0,557</p>
	<p>0,022 detik / 7 produk</p>	<p>0,789</p>

2. Pengujian *Black Box*

Pengujian *black-box* adalah untuk meng-input dan menguji apakah fungsi-fungsi yang terdapat pada sistem sudah sesuai dengan tujuan, dan tercermin dalam hasil keluaran. Adapun rencana yang akan dibuat yaitu input, hasil yang ingin di harapkan, *output* dan kesimpulan, yang akan diuji adalah inputan berupa deskripsi, dan hasil rekomendasi Pada tabel 4.2 merupakan hasil pengujian dari *Black Box*.

Tabel 4. 2 Pengujian Black Box

Input	Hasil yang diharapkan	Output	Kesimpulan
Input deskripsi “Kulit berminyak bekas jerawat bopeng”	Menampilkan rekomendasi <i>skincare</i>	Setelah di-klik tombol “Rekomendasi” muncul hasil rekomendasi	Berhasil
Input deskripsi “kombinasi mencerahkan	Menampilkan rekomendasi <i>skincare</i>	Setelah di-klik tombol “Rekomendasi” muncul hasil rekomendasi	Berhasil

Input deskripsi “Kulit berminyak noda hitam”	Menampilkan rekomendasi <i>skincare</i>	Setelah di-klik tombol “Rekomendasi” muncul hasil rekomendasi	Berhasil
Input deskripsi “normal”	Memberikan peringatan deskripsi yang diberikan kurang sesuai	Setelah di-klik tombol “Rekomendasi” muncul peringatan “ <i>No recommendations found for the given description.</i> ”	Berhasil
Input deskripsi “krim pencerah wajah”	Memberikan peringatan deskripsi yang diberikan kurang sesuai	Setelah di-klik tombol “Rekomendasi” muncul peringatan “ <i>No recommendations found for the given description.</i> ”	Berhasil

#### 4.4 Validasi Implementasi Algoritma

Dalam penelitian ini menggunakan validasi *Precision*, *Recall*, *F1-Score*, dan *Accuracy* karena validasi tersebut dapat mengukur kinerja model dalam mengklasifikasikan hasil *cosine similarity* dengan lebih rinci, termasuk kemampuan model dalam menghindari kesalahan dalam mengklasifikasikan hasil, seberapa baik model menemukan semua dokumen yang mirip, dan bagaimana keseimbangan antara *Precision* dan *Recall*. Tabel 4.3 merupakan hasil validasi yang cara kerjanya dengan melakukan perbandingan antara hasil rekomendasi dari sistem dengan data yang sebenarnya.

Tabel 4. 3 Hasil Validasi

No.	Input	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Accuracy</i>
1	kulit kering menyamarkan noda hitam	0.556	1.0	0.714	0.714
2	kulit berminyak menyamarkan noda hitam	0.470	1.0	0.639	0.64

3	kulit normal menyamarkan noda hitam	0.470	1.0	0.639	0.64
4	kulit sensitif menyamarkan noda hitam	0.529	1.0	0.692	0.692
5	kulit kombinasi menyamarkan noda hitam	0.444	1.0	0.615	0.615
6	kulit kering mencerahkan kulit	0.666	0.875	0.756	0.756
7	kulit berminyak mencerahkan kulit	0.777	0.933	0.848	0.843
8	kulit normal mencerahkan kulit	0.875	0.933	0.903	0.903
9	kulit sensitif mencerahkan kulit	0.857	0.857	0.857	0.857
10	kulit kombinasi mencerahkan kulit	0.857	0.857	0.857	0.857
11	kulit berminyak minyak berlebih	1.0	1.0	1.0	1.0
12	kulit kering melembabkan	1.0	0.586	0.739	0.739
13	kulit normal melembabkan	1.0	0.5	0.666	0.666
14	kulit sensitif melembabkan	1.0	0.458	0.628	0.628
15	kulit kombinasi melembabkan	1.0	0.476	0.645	0.645
16	kulit kering jerawat	1.0	0.75	0.857	0.857

17	kulit berminyak jerawat	1.0	0.608	0.756	0.756
18	kulit normal jerawat	1.0	0.722	0.838	0.838
19	kulit sensitif jerawat	1.0	0.75	0.857	0.857
20	kulit kombinasi jerawat	1.0	0.769	0.869	0.869

Berdasarkan validasi yang dilakukan menggunakan metrik evaluasi diatas dengan memberikan 20 *query* berbeda, hasil yang didapatkan yaitu:

1. Rata-rata *Precision* sebesar 0.825
2. Rata-rata *Recall* sebesar 0.803
3. Rata-rata *F1-Score* sebesar 0.769
4. Rata-rata *Accuracy* sebesar 0.768

Dari hasil evaluasi yang didapatkan, terlihat bahwa sistem memiliki performa yang baik dalam menghasilkan rekomendasi yang relevan. Berikut adalah beberapa kesimpulan yang dapat diambil:

- a. Rata-rata *Precision* sebesar 0.825: Nilai presisi yang tinggi menunjukkan bahwa sebagian besar item yang direkomendasikan oleh sistem memang relevan untuk pengguna. Dengan nilai ini, kemungkinan pengguna mendapatkan rekomendasi yang bermanfaat cukup tinggi.
- b. Rata-rata *Recall* sebesar 0.803: Nilai *recall* yang tinggi mengindikasikan bahwa sistem mampu menemukan dan merekomendasikan sebagian besar item yang sebenarnya relevan. Dengan kata lain, sistem dapat mengenali sebagian besar item yang seharusnya direkomendasikan.
- c. Rata-rata *F1-Score* sebesar 0.769: F1-score merupakan harmonik dari presisi dan recall. Nilai *F1-score* yang tinggi menunjukkan bahwa sistem memiliki keseimbangan yang baik antara kemampuan untuk merekomendasikan item yang relevan dan kemampuan untuk mengenali item yang relevan.
- d. Rata-rata *Accuracy* sebesar 0.768: Nilai akurasi yang cukup baik menunjukkan sejauh mana sistem dapat memberikan rekomendasi yang benar secara keseluruhan.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Hasil penelitian menunjukkan bahwa metode Keyword Extraction TF-IDF telah berhasil diterapkan dan dapat memberikan *keyword* yang sesuai untuk menggambarkan deskripsi produk. Selanjutnya, algoritma *Cosine Similarity* berhasil diintegrasikan ke dalam sistem rekomendasi sehingga sistem dapat memberikan rekomendasi *skincare* sesuai dengan deskripsi yang diinputkan oleh user. Sistem memberikan rekomendasi dengan rata-rata waktu 0,014 detik, *similarity* tertinggi mencapai 0,937, dan mendapat nilai akurasi tertinggi sebesar 1,0. Dari hasil tersebut dapat disimpulkan bahwa sistem sudah bekerja dengan cepat karena prosesnya hanya membutuhkan waktu dibawah 1 detik dengan *similarity* dan akurasi tertinggi yang didapat lebih dari 90%.

#### 5.2 Saran

Saran yang dapat diterapkan untuk pengembangan sistem ini lebih lanjut nantinya yaitu : Pada sistem rekomendasi *skincare* dengan metode *keyword extraction* dan *cosine similarity* adalah sistem memiliki fitur yang lebih lengkap seperti *allergen* yang bisa lebih spesifik lagi, dan menambahkan metode *collaborative-filtering*.

## DAFTAR PUSTAKA

- Adi, K., #1, N., & Sebastian, D. (2018). Pembentukan Dataset Topik Kata Bahasa Indonesia pada Twitter Menggunakan TF-IDF & Cosine Similarity. *Jurnal Teknik Informatika dan Sistem Informasi*, 4, 2443–2229. <http://dx.doi.org/10.28932/jutisi.v4i3.862>
- Ayuningrum, E., Azhar, Y., & Marthasari, G. I. (2022). Sistem Rekomendasi Produk Skincare Korea Berbasis *Web* Menggunakan Metode Collaborative Filtering. *Repositor*, 4(4), 497–506. <https://doi.org/10.22219/repositor.v4i4.1418>
- Fikri, M., Syahbani, N., & Ramadhan, N. G. (2023). Klasifikasi Gerakan Yoga dengan Model Convolutional Neural Network Menggunakan Framework Streamlit. *Media Informatika Budidarma*, 7(1), 509–519. <https://doi.org/10.30865/mib.v7i1.5520>
- Haslindah, A., Suharni, Nadiya, N. mujahidah, & Sanpratiwi. (2022). Sistem Pendukung Keputusan Pemilihan Jenis Skincare Berdasarkan Jenis Kulit Wajah Menggunakan Metode Weighted Product (WP). *Jurnal Teknologi dan Komputer (JTEK)*, 2(02), 196–201. <https://doi.org/10.56923/jtek.v2i02.98>
- Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(1), 306–312. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/796>
- Husin, N. (2023). *Komparasi Algoritma Random Forest , Naïve Bayes , dan Bert Untuk Multi-Class Classification Pada Artikel Cable News Network ( CNN)*. 7(1), 75–84. <https://doi.org/https://doi.org/10.55886/infokom.v7i1>
- Iriananda, S. W., Muslim, M. A., & Dachlan, H. S. (2019). Identifikasi Kemiripan Teks Menggunakan Class Indexing Based dan Cosine Similarity Untuk Klasifikasi Dokumen Pengaduan. *Matics*, 10(2), 30. <https://doi.org/10.18860/mat.v10i2.5327>
- Ivada, D., Darmawan, D. R., & Novianti, N. (2022). Beauty Care Kebutuhan Kultural Perempuan Metropolitan. *Journal of Arts and Humanities*, 26, 216–

225. <http://theartsjournal.org/index.php/site/article/view/478/270>
- Kurniawan, T. (2017). Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naïve Bayes Classifier Dan Support Vector Machine Media Mainstream Menggunakan Naïve Machine. *IT Journal*, 23, 1. [https://repository.its.ac.id/48557/1/1313100075-Undergraduate\\_Theses.pdf](https://repository.its.ac.id/48557/1/1313100075-Undergraduate_Theses.pdf)
- Larasati, F. B. A., & Februriyanti, H. (2021). Sistem Rekomendasi Product Emina Cosmetics Dengan Menggunakan Metode Content - Based Filtering. *Jurnal Manajemen Informatika dan Sistem Informasi*, 4(1), 45. <https://doi.org/10.36595/misi.v4i1.250>
- Levi, M., Palit, H. N., Rostianingsih, S., & Siwalankerto, J. (2020). Perbandingan Performa Tools *Web Scraping* pada *Website* dengan Data Statis dan Dinamis. *Jurnal Infra*, 8, 1–7. <https://publication.petra.ac.id/index.php/teknik-informatika/article/view/10521>
- Maarif, V., Nur, H. M., & Septianisa, T. A. (2019). Sistem Pendukung Keputusan Pemilihan Skincare Yang Sesuai Dengan Jenis Kulit Wajah Menggunakan Logika Fuzzy. *EVOLUSI: Jurnal Sains dan Manajemen*, 7(2), 73–80. <https://doi.org/10.31294/evolusi.v7i2.6755>
- Pebrianto, R., Nugraha, S. N., & Gata, W. (2020). Perancangan Sistem Pakar Penentuan Jenis Kulit Wajah Menggunakan Metode Certainty Factor. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 83–93. <https://doi.org/10.31294/ijcit.v5i1.7408>
- Prasetyo, V. R., Benarkah, N., & Chrisintha, V. J. (2021). Implementasi Natural Language Processing Dalam Pembuatan Chatbot Pada Program Information Technology Universitas Surabaya. *Teknika*, 10(2), 114–121. <https://doi.org/10.34148/teknika.v10i2.370>
- Raison-Peyron, N. (2019). Allergens in cosmetics. In *Revue Francaise d'Allergologie* (Vol. 59, Nomor 3, hal. 214–215). <https://doi.org/10.1016/j.reval.2019.02.218>
- Riadi Silitonga, Y. (2019). Sistem Pendeteksi Berita Hoax di Media Sosial dengan Teknik Data Mining Scikit Learn. *Jurnal Ilmu Komputer*, 4, 173.

www.beritasatu.com,

- Riandari, F. (2017). Sistem Pakar Mendiagnosa Penyakit Kulit Wajah. *Jurnal Mantik Penusa*, 1(2), 85–89.
- Safitri, D. A. N., Halilintar, R., & Wahyuniar, L. S. (2021). Sistem Rekomendasi Skincare Menggunakan Metode Content-Based Filtering dan Algoritma Apriori. *Seminar Nasional Inovasi Teknologi (SEMNAS INOTEK)*, 242–248. <https://doi.org/10.29407/inotek.v5i2.1136>
- Syauki, W. R., & Amalia Avina, D. A. (2020). Persepsi dan preferensi penggunaan skincare pada perempuan milenial dalam perspektif komunikasi pemasaran. *Jurnal Manajemen Komunikasi*, 4(2), 42. <https://doi.org/10.24198/jmk.v4i2.25719>
- Yusuf, M., & Cherid, A. (2020). Implementasi Algoritma Cosine Similarity Dan Metode TF-IDF Berbasis PHP Untuk Menghasilkan Rekomendasi Seminar. *Jurnal Ilmiah Fakultas Ilmu Komputer*, 9(1), 8–16. <https://publikasi.mercubuana.ac.id/index.php/fasilkom/article/view/8830>

