

**Implementasi Metode Content-Based Filtering pada
Sistem Rekomendasi Produk Bodycare**

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

TUGAS AKHIR

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



Avissa Salsabila

19611190

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2023**

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul : Implementasi Metode Content-Based Filtering pada Sistem Rekomendasi Produk *Bodycare*
(Studi Kasus: Review Produk *Bodycare* di *Female Daily*)

Nama Mahasiswa : Avissa Salsabila

NIM : 19611190

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN

Yogyakarta, 06 Oktober 2023

Ketua Prodi Statistika

Pembimbing

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

(Ayundyah Kesumawati, S.Si., M.Si.)

HALAMAN PENGESAHAN

TUGAS AKHIR

Implementasi Metode Content-Based Filtering pada Sistem Rekomendasi Produk Bodycare

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

Nama Mahasiswa : Avissa Salsabila

NIM : 19611190

TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL: 16 Oktober 2023

Nama Penguji:

Tanda Tangan

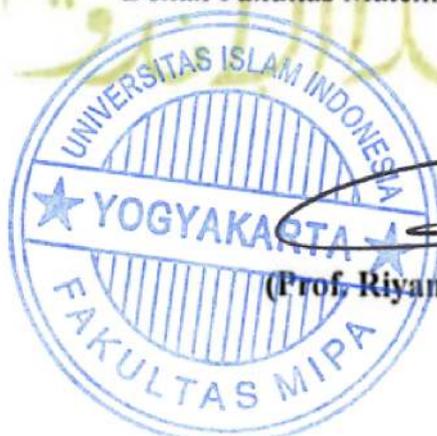
1. Dr.techn.Rohmatul Fajriyah, S.Si., M.Si.

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

3. Ayundyah Kesumawati, S.Si., M.Si.

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



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



KATA PENGANTAR



Assalamu'alaikum Wr.Wb

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

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

1. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta jajarannya.
2. Ibu Dr. Atina Ahdika, S.Si., M.Si., selaku Ketua Program Studi Statistika Universitas Islam Indonesia.
3. Ibu Ayundyah Kesumawati, S.Si., M.Si., selaku dosen pembimbing yang telah memberi bimbingan dan arahan selama mengerjakan tugas akhir ini. Terima kasih atas waktu dan kesabarannya dalam membimbing penulis.
4. Seluruh Dosen Program Studi Statistika Universitas Islam Indonesia yang memberikan banyak ilmu serta pengalaman yang berharga selama perkuliahan.

5. Kedua Orang tua tercinta, Papa dan Mama yang telah memberikan segala dukungan, do'a, dan kasih sayang dari awal proses perkuliahan hingga selesai.
6. Semua pihak yang telah memberikan dukungan dan bantuan yang tidak dapat penulis sebutkan satu per satu, terima kasih.
7. Yang paling akhir, saya mengucapkan terimakasih kepada diri saya sendiri sudah bertahan sampai saat ini melewati segala cobaan, rintangan, segala permasalahan yang ada dengan kesabaran dan hati yang ikhlas.

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

Wassalamualaikum Wr.Wb

Yogyakarta, 06 Oktober 2023



Avissa Salsabila

DAFTAR ISI

HALAMAN SAMPUL	i
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
BAB 2 TINJAUAN PUSTAKA	5
BAB 3 LANDASAN TEORI	16
3.1 Beauty Platform	16
3.2 Female Daily	16
3.3 <i>Bodycare</i>	17
3.4 Text Mining	18
3.4.1 Text Pre-processing	18
3.4.2 Vektorisasi Kata	19
3.5 TF IDF	19
3.5.1 Perhitungan Term-Frequency (TF)	20
3.5.2 Perhitungan Invers Document Frequency (IDF)	20
3.5.3 Perhitungan bobot TF-IDF	21
3.6 Cosine Similarity	21
3.7 Sistem rekomendasi	24
3.7.1 Collaborative Filtering	24
3.7.2 Content Based Filtering	25
BAB 4 METODOLOGI PENELITIAN	27
4.1 Populasi Penelitian	27
4.2 Jenis dan Sumber Data	27
4.3 Variabel Penelitian	27
4.4 Metode Pengambilan Data	29
4.5 Metode Analisis Data	29

4.6	Diagram Penelitian	31
BAB 5	HASIL DAN PEMBAHASAN	32
5.1	Analisis Deskriptif	32
5.2	<i>Pre-processing Data</i>	35
5.2.1	Cleaning Data	35
5.2.2	Case Folding	35
5.2.3	Stopword Removal	36
5.2.4	Stemming	37
5.2.5	Tokenizing	37
5.3	Pembobotan <i>Term Frequency-Invers Document Frequency</i> (TF-IDF)	38
5.4	<i>Cosine Similarity</i>	41
5.5	Hasil Rekomendasi	42
5.6	Tampilan Website	44
5.7	Evaluasi Model	45
BAB 6	PENUTUP	49
6.1	Kesimpulan	49
6.2	Saran	49

DAFTAR TABEL

Tabel 2.1 Perbandingan dengan Penelitian Terdahulu	5
Tabel 4.1 Definisi Operasional Variabel	27
Tabel 5.1 Contoh dokumen pada tahapan <i>Cleaning Data</i>	35
Tabel 5.2 Contoh dokumen pada tahapan <i>Case Folding</i>	36
Tabel 5.3 Contoh dokumen pada tahapan <i>Stopword Removal</i>	36
Tabel 5.4 Contoh dokumen pada tahapan <i>Stemming</i>	37
Tabel 5.5 Contoh dokumen pada tahapan <i>Tokenizing</i>	37
Tabel 5.6 Tabel Perhitungan TF	38
Tabel 5.7 Perhitungan IDF	39
Tabel 5.8 Perhitungan TF-IDF	40
Tabel 5.9 Nilai <i>Cosine Similarity</i>	42
Tabel 5.10 Rekomendasi Produk 5 Teratas	42
Tabel 5.11 <i>Review Produk “Extra White Care & Protect Body Serum” dan “Healthy Bright Aloe Soothe”</i>	43
Tabel 5.12 Evaluasi Model 1	45
Tabel 5.13 Evaluasi Model 2	46
Tabel 5.14 Evaluasi Model 3	47
Tabel 5.15 Total Akurasi Kesesuaian.....	48

DAFTAR GAMBAR

Gambar 3.1 Logo <i>Female Daily</i>	17
Gambar 3.2 Ilustrasi <i>cosine similarity</i> (Sumber: (Seelam, 2021))	22
Gambar 3.3 Ilustrasi Proses <i>Content-Based Filtering</i>	25
Gambar 4.1 Tahapan Analisis Penelitian	31
Gambar 5.1 10 Brand Teratas	32
Gambar 5.2 Pembelian produk	33
Gambar 5.3 Rentang usia yang menggunakan produk <i>bodycare</i>	34
Gambar 5.4 Tampilan Home	44
Gambar 5.5 Tampilan Hasil Rekomendasi	45

DAFTAR LAMPIRAN

Lampiran 1 Dataset dan <i>Script Python</i>	55
--	----

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat 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, 06 Oktober 2023



(Avisa Salsabila)

ABSTRAK

Implementasi Metode Content-Based Filtering pada Sistem Rekomendasi Produk Bodycare

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

Avissa Salsabila

Program Studi Statistika, Fakultas MIPA

Universitas Islam Indonesia

Bodycare merupakan suatu rangkaian perawatan yang dilakukan untuk menjaga kesehatan dan kecantikan kulit tubuh secara keseluruhan. Produk *bodycare* mencakup seperti sabun, *scrub*, *lotion*, dan produk lainnya. Kebutuhan informasi bagi konsumen khususnya dalam memilih produk *bodycare* saat ini tentunya sangat dibutuhkan. *Female Daily* merupakan sebuah *beauty platform* yang dikenal sebagai sumber informasi terpercaya dalam bidang kecantikan. Permasalahan yang ada saat ini adalah terlalu banyak produk *bodycare* yang beredar dipasaran membuat konsumen kesulitan dalam memilih produk yang diinginkan. Solusi yang dapat dilakukan adalah membuat suatu model sistem rekomendasi dalam memilih produk *bodycare* yang di peroleh dari *review* pengguna pada website *femaledaily.com*. Pada penelitian ini menggunakan metode *Content-Based Filtering* untuk mengetahui rekomendasi produk *bodycare* berdasarkan *review* yang diberikan oleh sistem rekomendasi. Metode *Content-Based Filtering* menggunakan konsep pembobotan TF-IDF dan *Cosine Similarity*. Berdasarkan hasil rekomendasi produk “*Extra White Care & Protect Body Serum*” sebagai contoh dalam pengujian. Didapatkan hasil rekomendasi yang memberikan lima rekomendasi produk *bodycare* terhadap pengguna berdasarkan *review* pada produk “*Extra White Care & Protect Body Serum*” berturut-turut yaitu *Healthy Bright Aloe Soothe* dari *Vaseline*, *Body Lotion* dari *Secret Garden*, *Murumuru Butter & Rose Body Lotion* dari *Love Beauty and Planet*, *Fragrance Brightening Body Lotion Jolly* dari *Scarlett*, dan *Sensational Body Lotion Vanilla & Almond Oil* dari *Nivea*. Dengan nilai *cosine similarity* berturut-turut sebesar 0.897903, 0.885389, 0.883158, 0.857375, dan 0.804388.

Kata Kunci: *Bodycare*, *Content-Based Filtering*, *FemaleDaily*, *Review*, *Sistem Rekomendasi*.

ABSTRACT

Implementation of the Content-Based Filtering Method in Bodycare Product Recommendation System

(Case Study: Bodycare Product Review at Female Daily)

Avissa Salsabila

Department of Statistics, Faculty of Mathematics and Natural Sciences

Universitas Islam Indonesia

Bodycare is a series of treatments carried out to maintain the health and beauty of the body's skin as a whole. Product *Bodycare* covers like soap, *scrub*, *lotion*, and other products. Information needs for consumers, especially in choosing products body care This is certainly very much needed now. *Female Daily* is a beauty *platform* which is known as a trusted source of information in the field of beauty. The current problem is too many products *bodycare* circulating on the market makes it difficult for consumers to choose the desired product. The solution that can be done is to create a recommendation system model for selecting products *bodycare* obtained from *review* users on the website *femaledaily.com*. This research uses the method *Content-Based Filtering* to find out product recommendations *bodycare* based on review provided by the recommendation system. Method *Content-Based Filtering* using the TF-IDF weighting concept and *Cosine Similarity*. Based on the results of product recommendations “*Extra White Care & Protect Body Serum*” as an example in testing. Recommendation results were obtained which provided five product recommendations *bodycare* against users based on review on the product “*Extra White Care & Protect Body Serum*” consecutively ie *Healthy Bright Aloe Soothe* dari *Vaseline*, *Body Lotion* from *Secret Garden*, *Murumuru Butter & Rose Body Lotion* from *Love Beauty and Planet*, *Fragrance Brightening Body Lotion Jolly* from *Scarlett*, and *Sensational Body Lotion Vanilla & Almond Oil* from *Nivea*. With *cosine similarity* successively equal to 0.897903, 0.885389, 0.883158, 0.857375, 0.804388.

Keywords: *Bodycare, Content-Based Filtering, FemaleDaily, Recommendation System, Review.*

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Perawatan tubuh atau *bodycare* adalah suatu rangkaian perawatan yang dilakukan untuk menjaga kesehatan, kebersihan, dan kecantikan kulit tubuh secara keseluruhan. Kulit adalah organ terluar yang melindungi tubuh dari lingkungan luar, termasuk infeksi, radiasi matahari, dan kerusakan lainnya. Banyak sekali hal-hal yang menyebabkan permasalahan pada kulit selain dari faktor lingkungan dan iklim, penggunaan perawatan kulit yang kurang tepat, pola makan yang tidak baik, dan perubahan pada hormon. Jika hal-hal tersebut dibiarkan dan tidak dijaga maka akan membuat kulit tidak sehat. Oleh karena itu penting untuk memberikan perhatian khusus pada kulit melalui perawatan yang tepat, karena kulit yang sehat dapat membantu menjaga kesehatan tubuh secara keseluruhan dan meningkatkan kualitas hidup.

Di Indonesia sendiri, perkembangan industri *bodycare* beberapa tahun terakhir meningkat cukup pesat dikarenakan semakin tingginya kesadaran masyarakat dalam merawat kesehatan kulit. Jika sebelumnya perawatan kulit banyak dilakukan oleh kaum perempuan, namun sekarang kaum laki-laki juga melakukannya. Hal ini dipengaruhi oleh gaya hidup masyarakat modern yang peduli terhadap penampilan. Serta sebagian besar masyarakat Indonesia menganggap penampilan dan gaya hidup menjadi sesuatu yang sangat penting untuk diperhatikan (Maulida, Arsi, & Alimi, 2018).

Terdapat beberapa hal yang harus diperhatikan untuk memastikan bahwa produk yang digunakan sesuai dengan kebutuhana kulit dan aman untuk digunakan. Hal yang pertama yaitu mengetahui jenis kulit, karena terdapat beberapa jenis kulit seperti normal, sensitif, berminyak, kering dan kombinasi. Pemilihan produk harus juga didasarkan sesuai permasalahan kulit yang dimiliki. Jenis produk *bodycare* yang digunakan secara umum yaitu sabun mandi, deodorant, *body lotion*, tabir surya. Namun selain penggunaan *bodycare* secara umum, tentunya dibutuhkan juga penggunaan *bodycare* pendukung lainnya agar memperoleh hasil yang lebih

maksimal. Jenis produk pendukung diantaranya yaitu *body scrub*, *hand cream*, *body treatment*, *oil* dan masih banyak lainnya. Dengan keberagaman produk *bodycare* yang ada serta didukung dengan promosi yang menarik menyebabkan banyak orang yang merasa kesulitan untuk memilih produk *bodycare* di pasaran.

Informasi mengenai produk *bodycare* tentunya sangat dibutuhkan dalam memilih produk yang akan digunakan. Dengan adanya internet, dan semakin majunya perkembangan teknologi membuat pengguna dapat mengakses berbagai sumber informasi seperti *website*, *platform media sosial*, blog kecantikan maupun forum komunitas untuk membaca ulasan, pengalaman, dan rekomendasi produk *bodycare*.

Female Daily merupakan sebuah *Beauty Platform* yang dikenal sebagai sumber informasi terpercaya dalam bidang kecantikan. Dilansir dari website www.femaledaily.com, situs ini menyediakan berbagai ulasan produk, tips perawatan, dan diskusi terkait kecantikan yang mencakup segala aspek dari *bodycare* hingga *makeup*. Selain itu, *Female Daily* juga menyediakan *marketplace* penjualan produk yang bernama *Beauty Studio*. *Female Daily* juga menyajikan artikel-artikel informatif dan edukatif tentang perawatan tubuh, membantu pengguna untuk memahami lebih dalam mengenai jenis kulit, masalah kulit, dan cara merawatnya.

Kendati demikian, dalam menghadapi berbagai preferensi yang ada, ragam produk yang dijual dipasaran, serta promosi yang kerap kali menonjolkan sisi positifnya saja. Maka dapat diatasi dengan penerapan sistem rekomendasi. Dimana informasi yang diberikan dapat lebih diarahkan pada preferensi dan kebutuhan pengguna, serta mengurangi tingkat subjektivitas dan membantu pengguna untuk lebih tepat dalam memilih produk yang sesuai. Sistem rekomendasi juga dapat membantu dengan mengidentifikasi preferensi dan kebutuhan pengguna, kemudian menyajikan rekomendasi yang lebih terfokus.

Sistem rekomendasi merupakan program atau sistem penyaringan informasi yang menjadi solusi dalam masalah kelebihan informasi dengan cara menyaring sebagian informasi penting dari banyaknya informasi yang ada dan bersifat dinamis sesuai dengan preferensi, minat, atau perilaku pengguna terhadap suatu barang.

Sistem rekomendasi dapat membantu dalam mengambil keputusan di dalam informasi yang kompleks dan banyak secara obyektif (Fajriansyah, Adikara, & Widodo, 2021). Terdapat beberapa metode yang dapat digunakan dalam membangun sebuah sistem rekomendasi antara lain *Content-Based Filtering*, *Collaborative Filtering*, *Hybrid Filtering*, dan lain sebagainya.

Oleh karena itu, pada penelitian ini menggunakan pendekatan dengan metode *Content-Based Filtering* karena sistem yang rekomendasi yang dibangun atas dasar kemiripan dari atribut produk yang telah dinilai oleh pengguna. Tahapan dalam analisis menggunakan metode *Content-Based Filtering* dimulai dengan perhitungan vektor dalam teks, pembobotan kata menggunakan *Term Frequency-Invers Document Frequency (TF-IDF)*, serta mengukur kemiripan dokumen menggunakan *Cosine Similarity Methods*. Pada hasil perhitungan *Cosine Similarity*, hasil perhitungan tertinggi yang menjadi vektor terdekat dan vektor yang ingin dibandingkan. Pada kasus ini, nilai tertinggi adalah rekomendasi produk *bodycare* yang akan diberikan kepada pengguna.

Berdasarkan permasalahan tersebut, peneliti akan melakukan penelitian mengenai sistem rekomendasi produk *bodycare* berdasarkan review pengguna pada website *Female Daily*. Sehingga dapat membantu pengguna dalam memilih produk *bodycare* berdasarkan kebutuhan yang ada.

1.2 Rumusan Masalah

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

1. Bagaimana Gambaran umum dari data produk *bodycare* yang tersedia di *beauty platform Femaledaily*?
2. Bagaimana hasil rekomendasi produk *bodycare* berdasarkan review yang diberikan oleh sistem rekomendasi menggunakan *Content-Based Filtering*?

1.3 Batasan Masalah

Agar menghindari perluasan pokok bahasan, maka diperlukan batasan masalah pada penelitian ini. Sebagai berikut:

1. Data yang digunakan adalah data review dari produk *bodycare* pada platform *Femaledaily* yang diperoleh pada tanggal 12 Maret 2023. Data *review* produk mencakup 22 kategori dengan jumlah data sebanyak 11.496 data.
2. Alat yang digunakan untuk melakukan pengambilan data atau *scrapping* adalah Data Miner. *Microsoft Excel* digunakan untuk menyimpan hasil dari *scrapping* data. Data diolah menggunakan bahasa Pemrograman Python pada *Google Colab*.

1.4 Tujuan Penelitian

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

1. Bertujuan untuk memberikan gambaran umum mengenai produk *bodycare* yang tersedia di platform *Femaledaily*
2. Bertujuan untuk mengetahui hasil rekomendasi pada produk *bodycare* berdasarkan *review* yang diberikan oleh sistem rekomendasi menggunakan metode *Content-Based Filtering*.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memberikan pengetahuan mengenai cara kerja pada sistem rekomendasi menggunakan metode *Content-Based Filtering*.
2. Dengan adanya model sistem rekomendasi, dapat memberikan rekomendasi produk *bodycare* yang lebih relevan dan sesuai dengan preferensi pengguna.
3. Penelitian ini diharapkan dapat digunakan sebagai referensi untuk penelitian lain yang berkaitan dengan sistem rekomendasi.

BAB 2 TINJAUAN PUSTAKA

Bab ini memuat penelitian terdahulu yang digunakan sebagai bahan acuan peneliti untuk melakukan penelitian ini. Beberapa ringkasan penelitian yang pernah dilakukan sebelumnya terkait studi kasus yang peneliti gunakan adalah sebagai berikut:

Tabel 2.1 Perbandingan dengan Penelitian Terdahulu

No	Nama Peneliti	Judul Penelitian	Tahun	Isi	Persamaan Penelitian	Perbedaan Penelitian
1.	Aprilia Saptu Ningrum, Heru Cahya Rustamaji, Yuli Fauziah	<i>Content Based</i> dan <i>Collaborative Filtering</i> pada Rekomendasi tujuan pariwisata di daerah Yogyakarta	2019	Penelitian ini memberikan rekomendasi tempat wisata berdasarkan hasil prediksi nilai rating baru menggunakan metode <i>hybrid</i> .	Menggunakan metode yang sama yaitu <i>Content-Based Filtering</i> .	Data yang digunakan peneliti yaitu data pariwisata di kota yogyakarta dan pada penelitian ini menggunakan dua metode, yaitu metode

No	Nama Peneliti	Judul Penelitian	Tahun	Isi	Persamaan Penelitian	Perbedaan Penelitian
						<i>Collaborative Filtering</i> dan <i>Content-Based Filtering</i> . Metode pengembangan sistem yang digunakan dalam penelitian ini menggunakan metode <i>grapple</i> .
2.	Rhesa Havilah Mondi, Ardhi Wijayanto, Winarno	<i>Recommendation system with Content-Based Filtering method for culinary tourism in Mangan application</i>	2019	Sistem rekomendasi untuk studi kasus aplikasi MANGAN. Hasil rekomendasi dipengaruhi oleh kemiripan item <i>profile</i>	Menggunakan metode yang sama yaitu <i>Content-Based Filtering</i>	Data yang digunakan pada penelitian yaitu data kuliner yang memuat gambar,

No	Nama Peneliti	Judul Penelitian	Tahun	Isi	Persamaan Penelitian	Perbedaan Penelitian
				restoran yang dipilih oleh pengguna dengan item profile restoran lainnya. Pengujian dilakukan dengan cara membandingkan hasil rekomendasi dengan data pembanding restoran pada kategori Khas Indonesia yang merupakan salah satu kriteria dari fitur konten pada item profile Ayam Geprek Keprabon.		nama, dan jarak restoran.
3.	Firmahsyah, Tiur Gantini	Penerapan <i>Metode Content-Based Filtering</i> Pada Sistem Rekomendasi Kegiatan	2016	Hasil penelitian berfokus kepada sistem rekomendasi dengan penerapan Algoritma Naïve Bayes.	Penerapan metode yang sama yaitu	Data pada penelitian adalah data siswa dan data

No	Nama Peneliti	Judul Penelitian	Tahun	Isi	Persamaan Penelitian	Perbedaan Penelitian
		Ekstrakurikuler (Studi Kasus di Sekolah ABC)		Atribut yang dipilih dan digunakan sebagai data untuk sistem rekomendasi ekstrakurikuler di Sekolah ABC adalah atribut bulan lahir, jenis kelamin, usia, dan pekerjaan orang tua	<i>Content-Based Filtering</i>	ekstrakurikuler di Sekolah ABC khusus tingkat PG, TKA dan TKB. Penerapan metode Naïve Bayes dengan Information Gain sebagai seleksi fitur
4.	Hilmi Hidayat Arfisko, Agung Toto Wibowo	Sistem Rekomendasi Film Menggunakan Metode <i>Hybrid Collaborative Filtering</i> dan <i>Content-Based Filtering</i>	2022	Pembuatan sistem rekomendasi film menggunakan metode <i>Hybrid Collaborative Filtering</i> dan <i>Content-Based Filtering</i> . Hasil pengujian yang	Penerapan metode yang sama yaitu <i>Content-Based Filtering</i>	Data yang digunakan pada penelitian ini merupakan dataset movielens yang

No	Nama Peneliti	Judul Penelitian	Tahun	Isi	Persamaan Penelitian	Perbedaan Penelitian
				dilakukan menghasilkan list rekomendasi item film yang lebih baik dibandingkan 3 metode lainnya yang diuji cobakan terhadap keseluruhan pengguna dalam dataset pengujian.		telah diproses sebelumnya. Menggunakan metode <i>Hybrid Collaborative Filtering</i> Dan <i>Content-Based Filtering</i>
5.	Ririn Intan Kumalsari, Ajib Susanto	SISTEM REKOMENDASI PADA E-MARKET PRODUK UMKM DINPORA PROPINSI JAWA TENGAH DENGAN MENGGUNAKAN	2015	Pengembangan sistem aplikasi menggunakan <i>Content Based Filtering</i> dengan metode i-CRM, sistem ini merupakan sistem rekomendasi produk berdasarkan <i>user profile</i> yang berfokus terhadap pelayanan	Penerapan metode yang sama yaitu <i>Content-Based Filtering</i>	Data yang digunakan pada penelitian ini merupakan data UMKM dan menggunakan 3 metode pengumpulan

No	Nama Peneliti	Judul Penelitian	Tahun	Isi	Persamaan Penelitian	Perbedaan Penelitian
		CONTENT BASED FILTERING		pelanggan guna menunjang dalam memperoleh loyalitas pelanggan dan mempertahankan pelanggan.		data seperti : Wawancara, Studi Literatur, dan <i>Research and site visit</i> .
6.	RIDHO ADE PUTRA	SISTEM REKOMENDASI PRODUK PADA E-COMMERCE MENGGUNAKAN METODE CONTENT BASED DAN COLLABORATIVE FILTERING	2019	Membangun sistem yang dapat memberikan rekomendasi produk yang relevan dengan pertimbangan kemiripan produk dan pola interaksi <i>user</i> . dibutuhkan sistem rekomendasi yang dapat memberikan rekomendasi produk bukan hanya sekedar pada kemiripan konten, melainkan	Penerapan metode yang sama yaitu <i>Content-Based Filtering</i>	Data yang digunakan terdiri dari produk kategori fashion, kategori elektronik, kategori obat & kosmetik, kategori kamera & gadget, kategori rumah

No	Nama Peneliti	Judul Penelitian	Tahun	Isi	Persamaan Penelitian	Perbedaan Penelitian
				juga berdasarkan minat masing-masing pengguna yang tidak sama satu dengan lainnya.		tangga yang dikumpulkan dari beberapa <i>e-commerce</i> di Indonesia.
7.	Dodyk Kurniawan, Patmi Kasih	Sistem Rekomendasi Menu Paket Miklur Menggunakan Metode <i>Content-Based Filtering</i> dan <i>Algoritma Apriori</i>	2021	Pada penelitian ini dibangun sebuah sistem yang dapat memberikan sebuah rekomendasi untuk memperkirakan informasi yang menarik bagi pengguna dan membantu calon <i>customer</i> dalam memutuskan menu paket apa saja yang dipilih. Penelitian ini memberikan sebuah	Penerapan metode yang sama yaitu <i>Content-Based Filtering</i>	Penelitian ini menggunakan metode Algoritma Apriori. Dataset pada penelitian ini meliputi data produk dan kriteria produk menu Miklur.

No	Nama Peneliti	Judul Penelitian	Tahun	Isi	Persamaan Penelitian	Perbedaan Penelitian
				rekomendasi menu paket minuman yang tersedia di Kedai Miklur menggunakan metode <i>Content-Based Filtering</i> dan <i>Algoritma Apriori</i>		
8.	Evan Salim, Jeanny Pragantha, Manatap Dolok Lauro	IMPLEMENTASI <i>CONTENT BASED FILTERING</i> PADA SISTEM REKOMENDASI FILM BERBASIS WEB	2021	Membangun dan menganalisa sebuah sistem rekomendasi film pada dataset MovieLens dengan metode <i>Content-Based Filtering</i> menggunakan <i>Cosine Similarity</i> dengan Tag pada film untuk membantu pengguna dengan cara memberikan rekomendasi	Penerapan metode yang sama yaitu <i>Content-Based Filtering</i>	Penelitian ini menggunakan dataset <i>MovieLens</i> .

No	Nama Peneliti	Judul Penelitian	Tahun	Isi	Persamaan Penelitian	Perbedaan Penelitian
				terhadap suatu film berdasarkan selera pengguna.		
9.	Didik Irawan, Kusrini, Muhammad Rudyanto Arie	Rekomendasi Penjual Bahan Makanan Dengan Metode Filtering Berbasis Konten dan Lokasi Pada Aplikasi Resep Masakan	2019	Penerapan <i>Algoritma Text Processing</i> dan TF-IDF untuk <i>Filtering</i> konten dan untuk menerapkan <i>algoritma klasifikasi K-Nearest Neighbors</i> (KNN) dan formula <i>haversine</i> untuk mencari irisan antara bahan masakan dan jarak lokasi penjual terdekat yang menjual sebagian besar atau seluruh bahan pada resep masakan.	Penerapan metode yang sama yaitu <i>Content-Based Filtering</i>	Penelitian ini menerapkan algoritma klasifikasi K-Nearest Neighbors (KNN) dan formula <i>haversine</i>

No	Nama Peneliti	Judul Penelitian	Tahun	Isi	Persamaan Penelitian	Perbedaan Penelitian
10.	DITA AISHA	SISTEM REKOMENDASI TOKO ONLINE MENGGUNAKAN ALGORITMA COLLABORATIVE FILTERING DAN CONTENT BASED FILTERING	2022	Penelitian ini menggunakan dua metode untuk membandingkan metode mana yang lebih baik dan untuk memudahkan konsumen dalam menyelesaikan masalah sistem rekomendasi seperti kesulitan dalam memilih produk yang paling banyak diminati, memudahkan konsumen dalam bertransaksi tanpa harus berlama lama memilih produk. sehingga perusahaan bisa memperluas	Menggunakan metode yang sama yaitu <i>Content-Based Filtering</i>	Penelitian ini menggunakan data dari toko <i>online</i> , dan menggunakan dua metode yaitu <i>collaborative filtering</i> dan <i>Content-Based Filtering</i>

No	Nama Peneliti	Judul Penelitian	Tahun	Isi	Persamaan Penelitian	Perbedaan Penelitian
				penjualan wilayah pemasarannya.		

Berdasarkan tinjauan pustaka diatas, perbedaan penelitian terdahulu dengan penelitian saat ini adalah menitikberatkan pada bentuk data yang diolah. Pada penelitian ini data yang digunakan berasal dari review produk yang diberikan oleh user. Sementara pada data peneliatian terdahulu menggunakan data pariwisata, data kuliner, data film, dan lainnya. Pada penelitian ini juga menggunakan tahap pre-processing yang lebih kompleks.

BAB 3 LANDASAN TEORI

3.1 Beauty Platform

Beauty platform adalah istilah yang merujuk kepada sebuah platform atau layanan digital yang menyediakan informasi, produk, serta layanan yang berhubungan dengan industri kecantikan. *Platform* ini dapat berupa aplikasi, situs web ataupun perangkat lunak yang dirancang guna menghubungkan pengguna dengan berbagai aspek kecantikan termasuk perawatan kulit, perawatan rambut, produk kecantikan, interaksi dan koneksi dengan komunitas kecantikan, dan banyak lainnya. *Beauty platform* merupakan sebuah ekosistem digital yang mencakup berbagai layanan informasi yang berkaitan dengan kecantikan (armeliafarah, 2023).

Tujuan dari adanya *beauty platform* ini adalah memberika solusi, informasi ataupun pengalaman yang relevan sesuai dengan kebutuhan serta preferensi dari pengguna. Dengan tujuan yang ada diharapkan dapat membantu pengguna mencapai tujuan yang diinginkan, dan dapat membantu untuk menjelajahi tren baru serta dapat berinteraksi dengan komunitas yang memiliki minat serupa (armeliafarah, 2023)..

3.2 Female Daily

Female Daily merupakan platform kecantikan yang berbasis di laman www.femaledaily.com dan aplikasi *mobile*. *Female Daily* memiliki fitur yang memungkinkan terjadinya interaktif dikalangan penggunanya. Seperti mendapatkan informasi seputar kecantikan lewat membaca blog atau editorial dan menyaksikan video yang diunggah melalui akun *Youtube Female Daily Network*, berinteraksi dengan sesama pengguna atau berdiskusi tentang topik-topik kecantikan tertentu melalui fitur Forum, serta menuliskan *review* terkait produk kecantikan tertentu dan menambahkan produk kecantikan ke fitur *database* (Ekaputri, Karolina, & Abdullah, 2020).

FEMALE DAILY

Gambar 3.1 Logo *Female Daily*

Platform ini juga memberikan kemudahan bagi penggunanya dalam mencari dan memperoleh informasi kecantikan yang berguna dalam pengambilan keputusan dalam konsumsi produk kecantikan. *Female Daily* menjadi objek penelitian karena *platform* ini termasuk salah satu *platform* kecantikan yang paling populer di Indonesia. *Female Daily* setiap harinya memiliki 450 pengguna baru yang mendaftar sebagai anggota: 7,5 juta kunjungan laman web per-bulan, lebih dari 355.000 pengikut di berbagai channel media social; 2 juta *unique users* per-bulannya. Penelitian ini bertujuan untuk mengetahui tingkat literasi informasi kecantikan dikalangan pengguna *Female Daily* (Ekaputri, Karolina, & Abdullah, 2020).

3.3 *Bodycare*

Bodycare berasal dari Bahasa Inggris, yaitu *Body* yang berarti tubuh dan *care* yang berarti perawatan, sehingga jika digabungkan berarti perawatan tubuh. Dalam skala yang lebih luas, *bodycare* dapat diartikan pula sebagai suatu rangkaian perawatan-perawatan tubuh yang digunakan untuk menjaga kondisi kulit agar tetap sehat dan terawat (Kristanto, 2020).

Kesehatan dari kulit tubuh merupakan suatu hal yang paling penting dan paling diperhatikan dikalangan wanita maupun pria, karena bagian kulit sangat terlihat dalam segi penampilan yang dimana harus dijaga dan dirawat. Penampilan juga sangat mendukung dalam tingkat kepercayaan diri pada seseorang (Revitasari, 2022).

Kulit memiliki fungsi untuk melindungi, mengeluarkan zat-zat yang tidak berguna dari sisa metabolisme tubuh menyimpan kelebihan lemak serta mengatur suhu yang ada di tubuh (Achroni & Famadiani, 2017). Oleh karena itu kulit harus senantiasa dijaga dan dirawat menyesuaikan dengan kondisi dari kulit.

3.4 Text Mining

Text Mining merupakan salah satu bagian dari *Data Mining*. *Text Mining* merupakan suatu proses menambang data berupa teks di mana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah guna mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen tersebut. (A.S & Shalahuddin, 2018).

Text Mining dapat menganalisis dokumen, menentukan kesamaan antara dokumen untuk mengetahui bagaimana mereka saling berhubungan dengan variabel lainnya, dan mengelompokkan dokumen berdasarkan kata-kata yang terkandung di dalamnya. (Deni Rusdian, 2019).

Pada proses ini diperoleh informasi berguna yang berasal dari sumber data melalui proses identifikasi dan eksplorasi yang tidak dalam bentuk *database record*, melainkan dalam teks yang tidak terstruktur.

3.4.1 Text Pre-processing

Text Pre-processing merupakan suatu tahapan awal dari *text mining*. Tujuan dari *text preprocessing* adalah guna mempersiapkan dokumen teks yang tidak terstruktur menjadi data yang terstruktur dan disimpan kedalam basis data (Langgeni, Baizal, & Firdaus, 2010). *Text Pre-processing* juga bertujuan untuk mengembalikan teks menjadi bahasa yang alami.

Pada penelitian ini menggunakan beberapa tahapan, diantaranya:

1. *Cleaning data* merupakan proses membersihkan data dari *noise* seperti tanda baca dan karakter yang tidak berpengaruh terhadap penelitian ini seperti tanda baca titik (.), koma(,), penghapusan angka dan lainnya.
2. *Case Folding* merupakan proses mengubah huruf kapital pada dokumen teks menjadi huruf kecil. Hal ini bertujuan agar menghindari huruf kapital tidak terdeteksi berbeda dengan kata yang tidak mengandung huruf kapital.
3. *Stopword Removal* merupakan suatu tahap penghapusan kata yang sering muncul namun tidak bermakna atau yang tidak memberikan nilai informasi penting dalam analisis teks.

4. *Stemming* merupakan suatu proses pengubahan dari bentuk kata menjadi kata dasar atau tahap mencari *root* kata dari setiap kata hasil *filtering*. *Stemming* digunakan untuk meningkatkan kualitas dari informasi yang akan
5. *Tokenizing* merupakan suatu proses untuk memecah sebuah kalimat menjadi sebuah kumpulan kata yang berdiri sendiri. Tokenisasi memecah teks yang semula berupa kalimat menjadi beberapa kata. Tokenisasi dapat menghilangkan delimiter seperti koma(,), titik(.), spasi dan karakter angka pada kalimat tersebut (Ratino, Hafidz, Anggraeni, & Gata, 2020)

3.4.2 Vektorisasi Kata

Vektorisasi kata pada lingkup *text mining* merupakan proses membuat vektor dengan nilainya berupa angka-angka kemunculan term (yaitu kata unik) dalam dokumen. Untuk melakukan vektorisasi maka teks harus dipisah menjadi kata per kata. Pemisahan teks menjadi kata-kata disebut dengan tokenisasi. Tokenisasi merupakan proses pemecahan dokumen menjadi komponen-komponen individual (Aulia & Patriya, 2019).

3.5 TF IDF

TF-IDF (*Term Frequency-Invers Document Frequency*) merupakan metode yang digunakan untuk menentukan seberapa jauh keterhubungan kata (*term*) terhadap dokumen dengan memberikan bobot pada setiap dokumen. Data yang sudah melalui tahap *pre-processing* harus berbentuk numerik dan untuk mengubah data tersebut menjadi numerik dibutuhkan metode pembobotan TF-IDF. Metode ini juga digunakan untuk menandakan pentingnya kata tersebut pada sebuah dokumen. Metode ini menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut (Simatupang & Utomo, 2019).

Dalam perhitungan bobot menggunakan TF-IDF terdapat beberapa tahapan yaitu sebagai berikut:

3.5.1 Perhitungan Term-Frequency (TF)

Term Frequency (TF) merupakan banyaknya kemunculan suatu kata (*term*) dalam suatu dokumen. Semakin sering suatu kata muncul dalam sebuah dokumen, berarti semakin penting kata tersebut atau akan memberikan nilai kesesuaian yang semakin besar. Ada empat cara yang bisa digunakan untuk mendapatkan nilai TF (Siregar, Sinaga, & Arianto, 2017)

a) *Raw Tf*

Metode yang digunakan untuk mendapatkan nilai TF sebuah kata yang didapatkan berdasarkan kemunculan kata dalam sebuah teks atau dokumen. Misalkan suatu kata muncul dalam dokumen sebanyak 1 kali maka nilai dari TF adalah 1.

b) *Logarithmic Tf*

Cara yang digunakan untuk mendapatkan nilai TF menggunakan fungsi matematika yaitu logaritmik, Berikut rumus yang digunakan:

$$TF = 1 + \text{Log}(TF) \quad (3.1)$$

c) *Binnary Tf*

Dalam metode ini, munculnya kata atau term pada dokumen tidak memiliki pengaruh karena TF akan bernilai 0 jika sebuah kata tidak ada pada dokumen tersebut, sedangkan jika sebuah kata atau term ada pada dokumen tersebut TF akan bernilai 1.

d) *Augmented Tf*

Nilai TF adalah jumlah kemunculan term pada sebuah dokumen. Nilai $\max(Tf)$ adalah jumlah kemunculan terbanyak term pada dokumen yang sama.

$$TF = 0,5 + 0,5 \times \frac{TF}{\max(TF)} \quad (3.2)$$

3.5.2 Perhitungan *Invers Document Frequency* (IDF)

Pada proses perhitungan IDF berkebalikan dengan TF, yang mana semakin tinggi frekuensi kemunculan term maka nilai bobot term itu sendiri akan semakin kecil (Apriani, Zakiyudin, & Marzuki, 2021) .Berikut persamaan dari IDF.

$$IDF_t = \ln\left(\frac{D}{df_t}\right) \quad (3.3)$$

Keterangan :

- D = Total dokumen
- df_t = Jumlah dokumen yg mengandung term t
- t = Kata ke-t dari kata kunci
- IDF = *Invers document frequency* dari term t

3.5.3 Perhitungan bobot TF-IDF

Berikut persamaan dari TF-IDF (Nurjannah, Hamdani, & Astuti, 2013).

$$W_{d,t} = TF_{d,t} \times IDF_{d,t} \quad (3.4)$$

$$W_{d,t} = TF_{d,t} \times \ln\left(\frac{D}{df_t}\right) \quad (3.5)$$

Keterangan :

- tf = Banyaknya kata yang dicari pada sebuah dokumen
- $W_{d,t}$ = Bobot dokumen ke-d terhadap kata ke-t
- d = Dokumen ke-d
- t = Kata ke-t dari kata kunci
- IDF = *Invers document frequency* dari term t

Berdasarkan persamaan 3.5, terlihat bahwa berapapun nilai dari $TF_{d,t}$ akan bernilai 0 jika nilai $D = df_t$, maka ditambahkan nilai 1 pada sisi IDF agar menghindari nilai $W_{d,t}$ tidak menjadi 0 (Apriani, Zakiyudin, & Marzuki, 2021). Persamaan dari TF-IDF akan menjadi seperti berikut:

$$W_{d,t} = TF_{d,t} \times \left(\ln\left(\frac{D}{df_t}\right) + 1 \right) \quad (3.6)$$

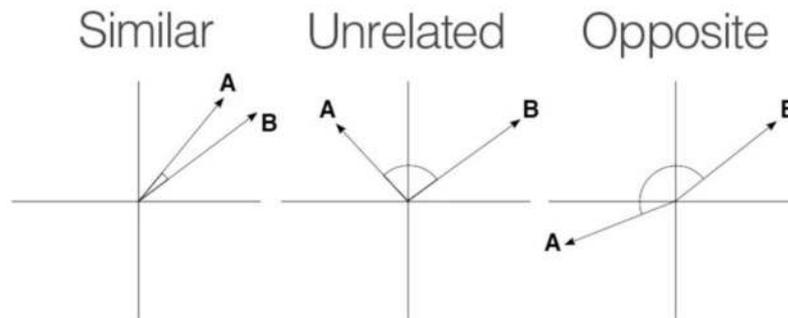
3.6 Cosine Similarity

Cosine Similarity adalah ukuran dari kesamaan antara dua buah vektor dalam sebuah ruang dimensi yang didapat dari nilai *cosinus* sudut dari perkalian dua buah vektor yang dibandingkan karena *cosinus* dari 0° adalah 1 dan kurang dari 1 untuk nilai sudut lainnya, maka nilai kesamaan dari dua buah vektor dikatakan mirip ketika nilai dari *cosine similarity* bernilai 1 (Ariantini, Lumenta, & Jacobus, 2016).

Cosine Similarity merupakan metrik yang mengukur kosinus dari sudut antara dua vektor yang diproyeksikan dalam ruang multi-dimensi. Semakin kecil sudut diantara dua vektor, semakin mirip keduanya.

Sebagai contoh:

Sudut antara dua vektor adalah 90 derajat, maka *cosine similarity* akan memiliki nilai 0, hal ini berarti dua vektor tersebut saling tegak lurus satu sama lain dimana berarti mereka tidak memiliki korelasi satu sama lain. Tetapi apabila saat pengukuran *cosine similarity* semakin mendekati 1, maka sudut antara dua vektor A dan B menjadi lebih kecil, dan dalam hal ini A dan B menjadi lebih mirip atau memiliki kesamaan satu sama lain (Seelam, 2021).



Gambar 3.2 Ilustrasi *cosine similarity* (Sumber: (Seelam, 2021))

Cosine Similarity dijelaskan secara sistematis sebagai pembagian antara perkalian titik (*dot product*) dari vektor-vektor dan hasil perkalian norma *Euclidean* atau magnitudo dari masing-masing vektor.

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$$

$$\|\vec{a}\| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \dots + a_n^2}$$

$$\|\vec{b}\| = \sqrt{b_1^2 + b_2^2 + b_3^2 + \dots + b_n^2}$$
(3.7)

Pada rumus diatas, a dan b adalah vektor-vektor dalam ruang multi-dimensi.

Karena nilai $\cos(\theta)$ berada dalam rentang $[-1,1]$:

- Nilai -1 akan menunjukkan vektor yang sangat berlawanan, yaitu tidak adanya kesamaan

- Nilai 0 mengindikasikan vektor yang *independen* (atau ortogonal) yang berarti tidak ada hubungan antara keduanya.
- Nilai 1 mengindikasikan tingkat kesamaan yang tinggi antara vektor-vektor tersebut.

Cosine Distance:

Cosine Similarity biasanya digunakan sebagai metrik kesamaan antara vektor-vektor. *Cosine distance* dapat didefinisikan sebagai berikut:

Cosine distance = 1- *cosine similarity*

Intuisi ini adalah jika dua vektor sama persis, maka kesamaannya adalah 1 (sudut = 0 sehingga $\cos(\theta) = 1$), dan oleh sebab itu, jaraknya adalah 0 ($1 - 1 = 0$) (Seelam, 2021).

Dalam proses *cosine similarity* yang menjadi masukan yaitu bobot dari *term* setiap data, bobot *term* tersebut digunakan untuk proses perhitungan pada jarak kemiripan dengan kata kluster, kemudian dari setiap nilai akan menentukan *centroid* setiap kluster (Herwijayanti, Ratnawati, & Muflikhah, 2018).

Berikut adalah rumus dari *cosine similarity*:

$$\cos \alpha = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (3.8)$$

Keterangan :

A = Vektor A, yang akan dibandingkan kemiripannya

B = Vektor B, yang akan dibandingkan kemiripannya

A_i = bobot *term* i dalam blok A_i

B_i = bobot *term* I dalam blok B_i

i = jumlah *term* dalam kalimat

n = jumlah vektor

$|A|$ = Panjang vektor A

$|B|$ = Panjang vektor B

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

Pada penelitian ini *cosine similarity* bekerja untuk membandingkan similaritas antar data. Kemudian mencari nilai similaritas tertinggi antar data. Apabila total similaritas yang didapatkan adalah 0 maka dokumen yang diolah tidak memiliki kesamaan dan jika nilai yang didapatkan maksimal 1 maka dokumen tersebut memiliki kemiripan (Riyan, Naf'an, & Burhanuddin, 2019). Nilai *cosine similarity* berada di rentang 0 – 1, apabila nilai *cosine similarity* semakin mendekati 0 maka data tersebut dikatakan rendah kemiripannya sebaliknya jika nilai *cosine similarity* mendekati 1 maka semakin tinggi kemiripan antar dokumen.

3.7 Sistem rekomendasi

Sistem rekomendasi merupakan program atau sistem penyaringan informasi yang menjadi solusi dalam masalah pada kelebihan informasi dengan cara menyaring sebagian informasi penting dari banyaknya informasi yang ada dan bersifat dinamis sesuai dengan preferensi, minat, atau perilaku pengguna terhadap suatu barang. Sistem rekomendasi dirancang untuk memahami dan memprediksi preferensi pengguna berdasarkan perilaku pengguna (Fajriansyah, Adikara, & Widodo, 2021).

Sistem rekomendasi bertujuan untuk menyediakan dan merekomendasi suatu item dalam membuat suatu keputusan yang diinginkan oleh pengguna. Sistem rekomendasi melakukan penyarian informasi yang disesuaikan dengan kebutuhan serta preferensi pengguna (Sebastia, Garcia, Onaindia, & Guzman, 2009).

3.7.1 Collaborative Filtering

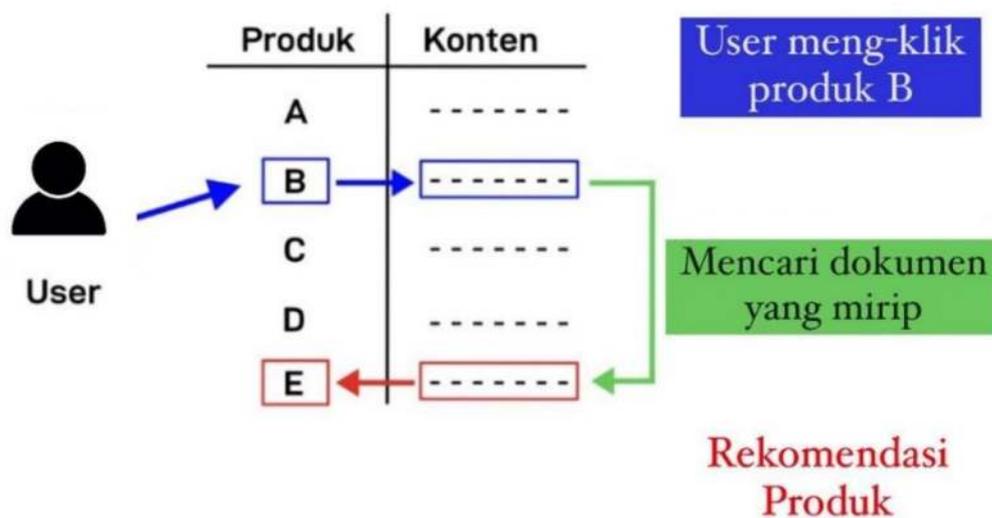
Metode *Collaborative Filtering* merupakan suatu proses dalam melakukan penyaringan atau pengevaluasian item dengan menggunakan opini dari orang lain. Tujuan utamanya ialah guna mengeksplorasi informasi mengenai perilaku di masa lampau maupun opini dari suatu komunitas pengguna yang kemudian digunakan untuk memprediksi item mana yang akan menarik bagi pengguna tersebut (Dzumiroh & Saptono, 2012).

Metode ini juga bekerja dengan cara membangun sebuah basis data yang menyimpan item-item yang disukai oleh pengguna. Lalu transaksi baru yang dibuat oleh seorang pengguna selanjutnya akan dicocokkan dengan basis data tersebut

untuk mengetahui data historis mana yang paling sesuai dengan data baru tersebut. data historis yang paling sesuai akan ditampilkan sebagai sebuah rekomendasi bagi pengguna yang akan melakukan transaksi tersebut (Marthasari, Azhar, & Puspitaningrum, 2015).

3.7.2 Content Based Filtering

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



Gambar 3.3 Ilustrasi Proses *Content-Based Filtering*

Berdasarkan **Gambar 3.2**. apabila user meng-klik produk B, maka sistem akan mengecek dokumen B, dalam hal ini yaitu review pada produk B. Selanjutnya

sistem mencari dokumen yang memiliki kemiripan dengan dokumen B, kemudian didapatkan produk E yang paling mirip dengan dokumen B. Maka sistem akan merekomendasikan produk E kepada user.

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

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

BAB 4 METODOLOGI PENELITIAN

4.1 Populasi Penelitian

Populasi pada penelitian ini adalah data pengguna dan review produk *bodycare* di website <https://femaledaily.com/>. Sedangkan sampel yang digunakan adalah data *review* berdasarkan produk dengan kategori *Bath & shower*, *Moisturizer*, *Personal Care*, dan *Hand & Foot*. Terdapat 22 sub-kategori. Data diambil pada tanggal 12 Maret 2023 - 12 April 2023 dengan jumlah keseluruhan data 11.496.

4.2 Jenis dan Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder. Data ini diperoleh dengan melakukan *scrapping* data pada website *Femaledaily.com* yang diambil pada tanggal 12 Maret 2023.

4.3 Variabel Penelitian

Terdapat 12 variabel yang digunakan pada penelitian ini.

Tabel 4.1 Definisi Operasional Variabel

No.	Variabel	Definisi Operasional	Keterangan
1	<i>Category</i>	Pembagian <i>product bodycare</i> berdasarkan jenis <i>bodycare</i> yang digunakan	1. Bath & Shower 2. Moisturizer 3. Personal Care 4. Hand & Foot
2	<i>Sub Category</i>	Pembagian produk <i>Bodycare</i> secara luas dari variabel <i>category</i>	1. Bar Soap 2. Lotion 3. Deodorant, dan lainnya
3	<i>Brand</i>	Identitas dari suatu produk yang diciptakan oleh suatu perusahaan	1. The Body Shop 2. Nivea 3. Vaseline, dan lainnya

No.	Variabel	Definisi Operasional	Keterangan
4	<i>Product Name</i>	Identitas yang digunakan untuk melabeli suatu produk	<ol style="list-style-type: none"> 1. Lemon Fresh Soap Bar 2. Firm Glow Gluta Vitamin Youthful Serum 3. Bio Oil 4. Dan lain sebagainya
5	<i>Rating</i>	Suatu penilaian yang diberikan oleh pengguna terhadap suatu produk yang digunakan	Rentang rating yang digunakan dari 4.0-5.0
6	<i>Reviews</i>	Informasi mengenai testimoni suatu produk yang diberikan oleh pengguna	Contoh review: awalnya aku ditawari ini sama nenekku, dan ternyata cocok banget buat kulitku yang kering dan sensitif kalo terkena sinar matahari berlebih! dia emang nggak mencerahkan
7	Nama Pengguna	Alamat profil dari pengguna yang memberikan <i>review</i>	@kohkim
8	Jenis Kulit	Keterangan mengenai kondisi kulit pengguna	<ol style="list-style-type: none"> 1. Combination 2. Oily 3. Normal, dan lainnya

No.	Variabel	Definisi Operasional	Keterangan
9	Umur	Keterangan usia pengguna	1. Dibawah 18 Tahun 2. 18 Tahun 3. Diatas 18 Tahun, dan lainnya
10	Penggunaan	Durasi lama produk yang digunakan	1. Kurang dari 1 minggu 2. Lebih dari 1 minggu, 3. 1 - 3 Bulan, dan lainnya
11	Pembelian	Keterangan informasi mengenai dimana produk dibeli	1. Shopee 2. Tiktokshop 3. Tokopedia, dan lainnya
12	Rating Pembeli	Penilaian secara individu dari pengguna	Rentang rating dari 1 hingga 5

4.4 Metode Pengambilan Data

Pengambilan data dilakukan melalui website dari *FemaleDaily.com* dengan cara *scrapping* menggunakan *Data Miner*. Data diambil pada tanggal 12 Maret 2023. Kemudian data yang telah di *scrapping* disimpan ke dalam *Microsoft Excel* 2019. Data memuat 4 kategori *bodycare*, 22 sub-kategori *bodycare*, dan jumlah produk pada setiap sub-kategori yang memiliki jumlah keseluruhan 11.496 produk.

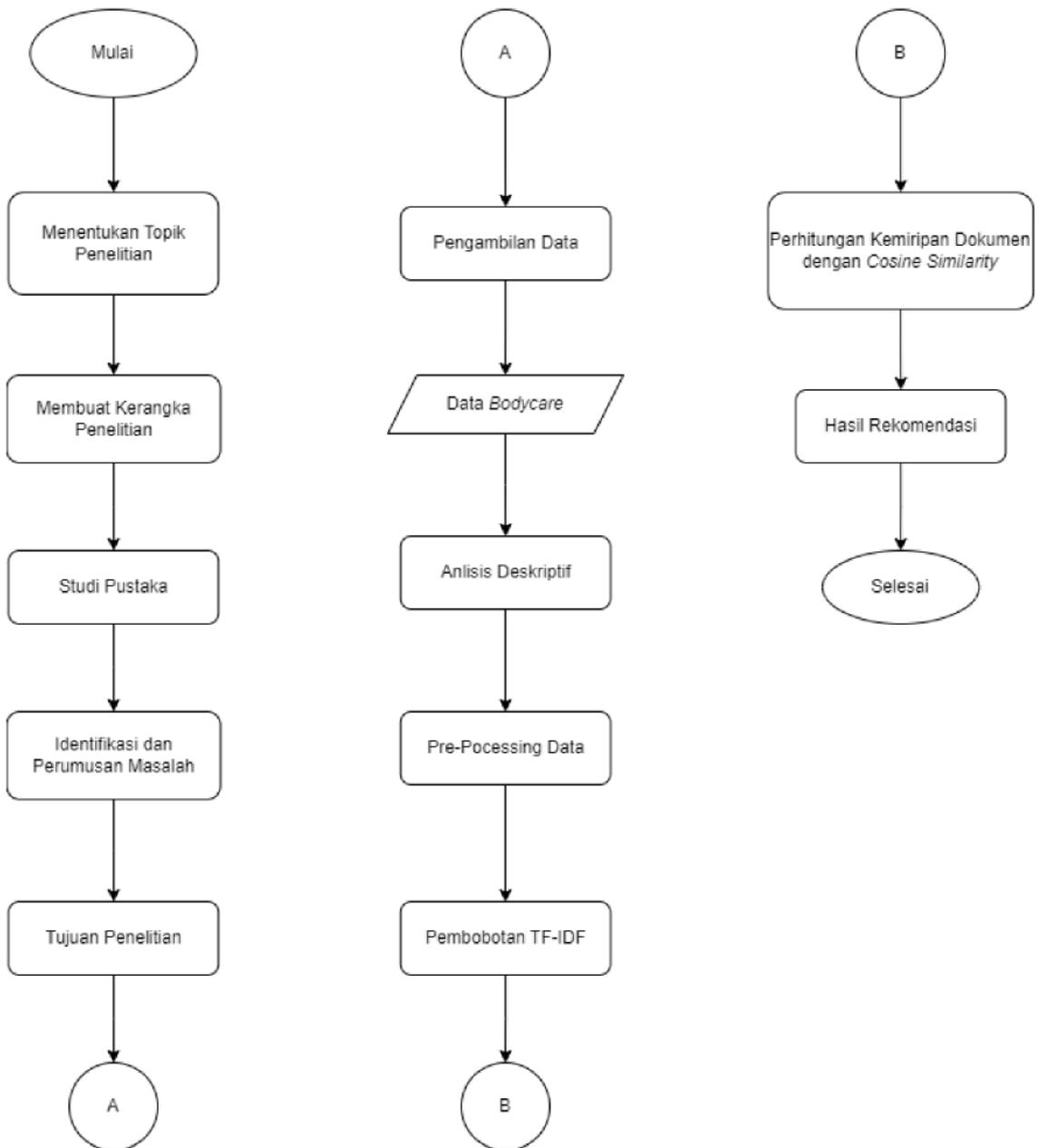
4.5 Metode Analisis Data

Pada penelitian ini metode yang digunakan adalah *Content Based Filtering* dengan menggunakan konsep perhitungan vektor pada teks yang kemudian akan dilakukan suatu pembobotan TF-IDF, dilanjutkan dengan mencari nilai kemiripan

pada dokumen untuk menjadi pembandingan dengan dokumen yang akan diuji menggunakan *Cosine Similarity* untuk menghasilkan suatu rekomendasi produk yang serupa dengan produk yang diinginkan pengguna berdasarkan konten produk pada brand tersebut dan *review* yang terdapat pada produk tersebut.

4.6 Diagram Penelitian

Tahapan analisis dari penelitian ini dijelaskan dari penentuan topik penelitian hingga tahap analisis disajikan dalam **Gambar 4.1**

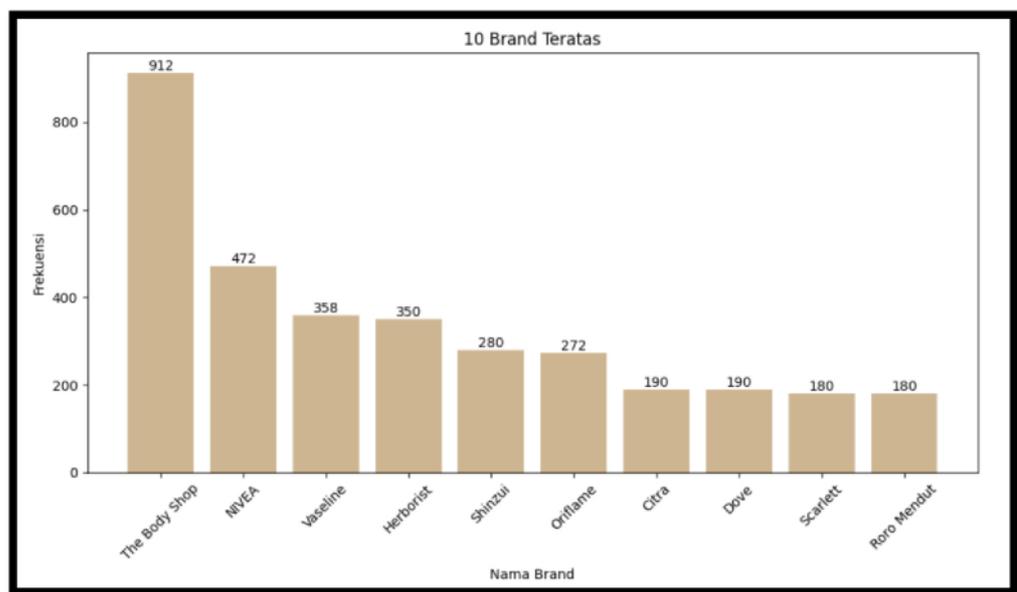


Gambar 4.1 Tahapan Analisis Penelitian

BAB 5 HASIL DAN PEMBAHASAN

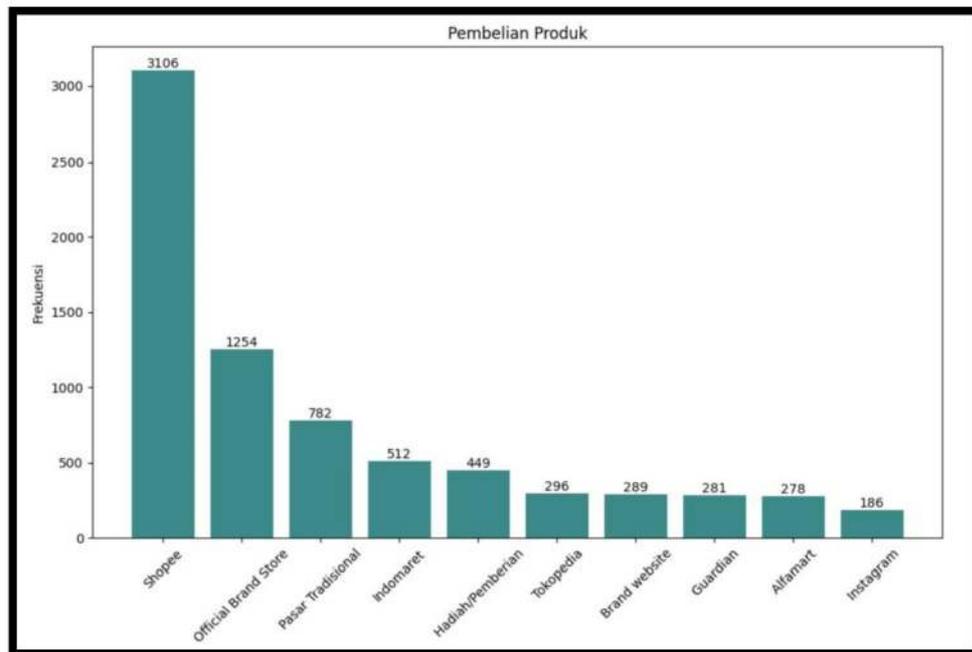
5.1 Analisis Deskriptif

Analisis deskriptif pada penelitian ini berguna untuk memberikan gambaran umum yang mencakup informasi 10 brand teratas, pembelian produk, dan rentang usia yang menggunakan produk *bodycare*. Hal ini bertujuan untuk memberikan pandangan mengenai tren pasar dan perilaku pada konsumen.



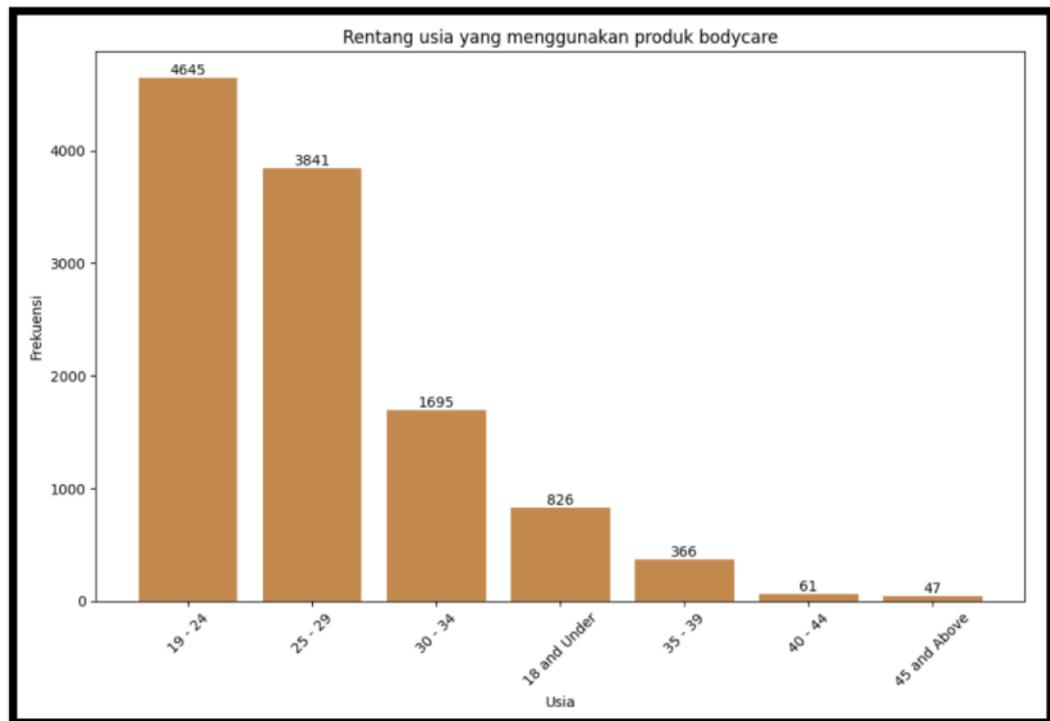
Gambar 5.1 10 Brand Teratas

Pada **Gambar 5.1** menampilkan 10 brand teratas atau brand dengan frekuensi paling banyak di review oleh pengguna *Femaledaily*. Brand dari The Body Shop berada di peringkat pertama yang paling sering di *review* oleh pengguna, dimana hal ini juga berarti produk *bodycare* dari brand tersebut memiliki popularitas yang tinggi diantara brand lainnya. Selain itu hal ini juga disebabkan oleh banyaknya jenis produk yang dikeluarkan oleh brand tersebut.



Gambar 5.2 Pembelian produk

Pada **Gambar 5.2** dapat dilihat sebaran dan frekuensi pembelian produk melalui berbagai tempat. Shopee berada di peringkat tertinggi yang mendominasi sebagai tempat yang paling sering digunakan oleh konsumen dalam membeli produk. Selain itu *official brand store* juga menjadi pilihan kedua yang cukup populer, hal ini dikarenakan konsumen memilih untuk membeli langsung ketempat *offline store* dari produk yang ingin dibeli seperti contoh *The Body Shop* yang memiliki *offline store*. **Gambar 5.2** memberikan wawasan mengenai preferensi konsumen dalam memilih tempat untuk pembelian produk *bodycare* dengan Shopee sebagai tempat yang paling sering digunakan.



Gambar 5.3 Rentang usia yang menggunakan produk *bodycare*

Pada **Gambar 5.3** menggambarkan rentang usia dari konsumen yang memberikan *review* produk *bodycare* di *Femaledaily* yang dimana memberikan ulasan dalam penggunaan produk yang dibeli. Dari data tersebut, terlihat bahwa kelompok usia 19-24 tahun dan 25-29 tahun mendominasi dalam memberikan ulasan produk *bodycare* di *Femaledaily*. Hal ini mengindikasikan bahwa *Femaledaily* menjadi *beauty platform* favorit di kalangan generasi muda dan dewasa muda sebagai sumber informasi mengenai produk *bodycare*. Rentang usia 30-34 tahun juga menunjukkan kontribusi yang signifikan dalam memberikan ulasan, meskipun jumlahnya lebih rendah dibandingkan dengan dua kelompok usia sebelumnya. Dengan demikian, ini juga dapat menjadi informasi berharga bagi produsen dan pemasar produk *bodycare* untuk mengarahkan upaya pemasaran dan promosi mereka kepada kelompok usia yang dominan dalam memberikan ulasan di *Femaledaily*.

5.2 *Pre-processing Data*

Pre-processing data diperlukan guna untuk merapikan data agar lebih terstruktur. Data hasil *scrapping* tentunya juga memiliki format yang tidak terstruktur, oleh karena itu sebelum data diolah, diperlukan tahap *preprocessing*. Pada tahapan ini menghilangkan bagian-bagian yang tidak penting atau yang tidak diperlukan sehingga data yang didapat lebih berkualitas. Tahapan *preprocessing* pada tahapan ini terdiri dari *Cleaning Data*, *Case Folding*, *Stopwords Removal*, *Stemming*, dan *Tokenizing*.

5.2.1 *Cleaning Data*

Pada tahapan ini dilakukan penghapusan komponen-komponen yang tidak diperlukan dalam data seperti tanda baca dan karakter yang tidak berpengaruh terhadap penelitian ini seperti tanda baca titik (.), koma(,), penghapusan angka dan lainnya, seperti contoh yang ada pada **Tabel 5.1**.

Tabel 5.1 Contoh dokumen pada tahapan *Cleaning Data*

Sebelum	Sesudah
bagus banget untuk mencerahkan kulit. kalo yang ini biasanya aku pakai siang hari. lembab, cepet meresap, wanginya enak, sama kudu rajin scrub badan sih. itu juga yang bikin tambah ngaruh, tambah cepet mencerahkan kulit. overall vaseline tidak pernah salah	bagus banget untuk mencerahkan kulit kalo yang ini biasanya aku pakai siang hari lembab cepet meresap wanginya enak sama kudu rajin scrub badan sih itu juga yang bikin tambah ngaruh tambah cepet mencerahkan kulit overall vaseline tidak pernah salah

5.2.2 *Case Folding*

Case Folding merupakan proses perubahan huruf kapital menjadi huruf kecil. Hal ini bertujuan agar semua kata menjadi sama dan dapat terdeteksi, karena jika terdapat kata yang berhuruf kapital akan membuat perbedaan dan tidak dianggap oleh sistem. Maka tahapan *case folding* berguna untuk menyeragamkan kata pada data. seperti contoh yang ada pada **Tabel 5.2**.

Tabel 5.2 Contoh dokumen pada tahapan *Case Folding*

Sebelum	Sesudah
bagus banget untuk mencerahkan kulit kalo yang ini biasanya aku pakai siang hari lembab cepet meresap wanginya enak sama kudu rajin scrub badan sih itu juga yang bikin tambah ngaruh tambah cepet mencerahkan kulit overall vaseline tidak pernah salah	bagus banget untuk mencerahkan kulit kalo yang ini biasanya aku pakai siang hari lembab cepet meresap wanginya enak sama kudu rajin scrub badan sih itu juga yang bikin tambah ngaruh tambah cepet mencerahkan kulit overall vaseline tidak pernah salah

5.2.3 *Stopword Removal*

Stopword Removal berfungsi untuk menghilangkan kata-kata umum yang tidak memberikan nilai informasi penting dalam analisis teks. Kata-kata ini disebut "*stopwords*" seperti kata "yang" dan "ini". Dapat dilihat pada contoh yang ada di **Tabel 5.3**.

Tabel 5.3 Contoh dokumen pada tahapan *Stopword Removal*

Sebelum	Sesudah
bagus banget untuk mencerahkan kulit kalo yang ini biasanya aku pakai siang hari lembab cepet meresap wanginya enak sama kudu rajin scrub badan sih itu juga yang bikin tambah ngaruh tambah cepet mencerahkan kulit overall vaseline tidak pernah salah	bagus mencerahkan kulit biasanya siang hari lembab cepat meresap wanginya enak rajin scrub badan tambah ngaruh tambah cepat mencerahkan kulit vaseline tidak pernah salah

5.2.4 Stemming

Stemming merupakan proses yang bertujuan untuk menghilangkan akhiran atau awalan kata-kata agar hanya menyisakan bentuk dasar atau kata dasar (r dari kata-kata tersebut. Seperti contoh yang ada pada **Tabel 5.4**

Tabel 5.4 Contoh dokumen pada tahapan *Stemming*

Sebelum	Sesudah
bagus mencerahkan kulit biasanya siang hari lembab cepat meresap wanginya enak rajin scrub badan tambah ngaruh tambah cepat mencerahkan kulit vaseline tidak pernah salah	bagus cerah kulit biasa siang hari lembab cepat resap wangi enak rajin scrub badan tambah ngaruh cepat cerah kulit vaseline tidak pernah salah

5.2.5 Tokenizing.

Tokenizing merupakan proses yang dilakukan untuk memisahkan atau memecahkan yang dimana awalnya berupa kalimat menjadi potongan kata yang akan menjadi suatu entitas dan memiliki nilai dalam penyusunan matriks dokumen pada proses berikutnya. Seperti contoh yang ada pada **Tabel 5.5**.

Tabel 5.5 Contoh dokumen pada tahapan *Tokenizing*

Sebelum	Sesudah
bagus cerah kulit biasa siang hari lembab cepat resap wangi enak rajin scrub badan tambah ngaruh cepat cerah kulit vaseline tidak pernah salah	"bagus", "cerah", "kulit", "biasa", "siang", "hari", "lembab", "cepat", "resap", "wangi", "enak", "rajin", "scrub", "badan", "tambah", "ngaruh", "cepat", "cerah", "kulit", "vaseline", "tidak", "pernah", "salah"

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

Setelah melewati tahap pre-processing data, dilanjutkan dengan tahap pembobotan TF-IDF, pada tahap ini menggabungkan dua konsep untuk melakukan perhitungan bobot, yaitu frekuensi dari kemunculan sebuah kata (term) didalam dokumen yang disebut Term Frequency (TF) dan frekuensi dokumen yang mengandung kata yang disebut dengan Invers Document Frequency (IDF). Langkah pertama peneliti melakukan perhitungan pada Term Frequency (TF), dan peneliti mengambil beberapa kata dari dokumen yang akan dijadikan sampel seperti kata “bagus”, “wangi”, “lembab”, “nyaman”, “kulit”. Maka berikut hasil nilai dai TF yang dapat dilihat pada **Tabel 5.6**

Tabel 5.6 Tabel Perhitungan TF

Dokumen	wangi	kulit	bagus	lembab	...	nyaman
1	0	2	1	0	...	0
2	0	0	0	0	...	0
3	1	3	0	0	...	0
4	0	1	2	0	...	0
5	0	1	1	0	...	0
6	0	1	0	0	...	0
7	0	1	0	0	...	0
8	2	2	0	0	...	0
9	1	0	0	0	...	0
10	1	0	0	0	...	0
:	:	:	:	:	...	:
11496	1	0	0	0	...	1
Jumlah	6906	6292	1308	2012	...	542

Setelah melakukan perhitungan TF, langkah selanjutnya yaitu menghitung nilai *Inverse Document Frequency* (IDF). IDF merupakan nilai \ln dari jumlah seluruh dokumen dibagi dengan jumlah dokumen yang mengandung kata terpilih (df) yang tertera pada **Persamaan 3.3**. Sebagai contoh misalnya diambil kata “kulit”, maka dalam melakukan perhitungan nilai df akan menampilkan jumlah

dokumen yang memuat kata “kulit”. Dari proses perhitungan df akan didapatkan banyaknya dokumen yang mengandung kata “kulit” yaitu sebanyak 6.282 sehingga nilai tersebut merupakan nilai DF. Berikut perhitungan IDF dari kata “kulit”:

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

$$IDF_{kulit} = \ln\left(\frac{11496}{6292}\right) = 0.602720$$

D merupakan jumlah dokumen, pada penelitian ini dokumen berjumlah 11.496. sedangkan df_{kulit} merupakan nilai *document frequency* dari kata kulit yaitu sebanyak 6282 dokumen.

Tabel 5.7 Perhitungan IDF

Dokumen	wangi	kulit	bagus	lembab	...	nyaman
1	0	0.602720	2.173499	0	...	0
2	0	0	0	0	...	0
3	0.509609	0.602720	0	0	...	0
4	0	0.602720	2.173499	0	...	0
5	0	0.602720	2.173499	0	...	0
6	0	0.602720	0	0	...	0
7	0	0.602720	0	0	...	0
8	0.509609	0.602720	0	0	...	0
9	0.509609	0	0	0	...	0
10	0.509609	0	0	0	...	0
:	:	:	:	:	...	:
11496	0.509609	0	0	0	...	3.054488

Tahap Selanjutnya yaitu melakukan pembobotan TF-IDF dengan cara mengalikan nilai TF dengan nilai IDF yang tertera pada **Persamaan 3.6**. Diketahui jumlah dokumen pada perhitungan ini yang dilambangkan dengan D berjumlah 11.496 dokumen. Sebagai contoh, perhitungan manual untuk kata “kulit” pada dokumen ke-4 sebagai berikut:

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

$$W_{kulit,4} = 1 \times \left(\ln \left(\frac{11496}{6292} \right) + 1 \right)$$

$$W_{kulit,4} = 1 \times (0.602720 + 1)$$

$$W_{kulit,4} = 1.602720$$

Berdasarkan hasil pembobotan yang telah dicari, maka bobot dari kata “kulit” pada dokumen ke-4 memiliki bobot kata sebesar 1.602720.

Sehingga untuk hasil perhitungan TF-IDF didapatkan hasil seperti pada **Tabel 5.8**.

Tabel 5.8 Perhitungan TF-IDF

Dokumen	wangi	kulit	bagus	lembab	nyaman
1	0.000000	3.205440	3.173499	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000
3	1.509609	4.808160	0.000000	0.000000	0.000000
4	0.000000	1.602720	6.346998	0.000000	0.000000
5	0.000000	1.602720	3.173499	0.000000	0.000000
6	0.000000	1.602720	0.000000	0.000000	0.000000
7	0.000000	1.602720	0.000000	0.000000	0.000000
8	3.019218	3.205440	0.000000	0.000000	0.000000
9	1.509609	0.000000	0.000000	0.000000	0.000000
10	1.509609	0.000000	0.000000	0.000000	0.000000
:	:	:	:	:	:
11497	1.509609	0.000000	0.000000	0.000000	4.054488

Hasil pembobotan kata dari metode TF-IDF menunjukkan seberapa berpengaruh kata dari sebuah dokumen terhadap seluruh dokumen. Sebuah kata dapat dikatakan penting apabila banyak kemunculan kata di dalam dokumen tersebut tinggi. Semakin banyak suatu kata muncul di dalam suatu dokumen maka bobot hubungan antara suatu kata dan suatu dokumen akan semakin tinggi.

Sebaliknya, jika kemunculan suatu kata di dalam dokumen sedikit maka bobot hubungan antara suatu kata dan suatu dokumen akan semakin rendah.

Nilai TF-IDF yang didapatkan ini selanjutnya akan digunakan pada perhitungan nilai *cosine similarity*. Alasan menggunakan nilai TF-IDF untuk perhitungan *cosine similarity* karena nilai TF-IDF memberikan hasil rekomendasi yang lebih relevan berdasarkan kesamaan kata atau konten yang mendasar. TF-IDF memberikan penekanan yang kuat pada kata-kata kunci dalam dokumen. TD-IDF juga membantu dalam mengidentifikasi kata-kata yang mewakili konsep atau topik tertentu dalam dokumen. Serta TF-IDF mencakup komponen normalisasi yang dimana memungkinkan perbandingan yang lebih adil antara dokumen dengan panjang yang berbeda. Dokumen yang lebih panjang cenderung memiliki nilai TF yang lebih tinggi karena mereka memiliki lebih banyak kata, tetapi komponen IDF dari TF-IDF membantu dalam mengatasi perbedaan ini, sehingga pengukuran kesamaan yang didapatkan akan lebih adil.

5.4 *Cosine Similarity*

Setelah melakukan perhitungan TF-IDF, dilanjutkan dengan menghitung nilai *cosine similarity* yang dimana bertujuan untuk membandingkan kemiripan antar dokumen. Nilai *cosine similarity* berada di rentang 0 – 1, apabila nilai *cosine similarity* semakin mendekati 0 maka data tersebut dikatakan rendah kemiripannya sebaliknya jika nilai *cosine similarity* mendekati 1 maka semakin tinggi kemiripan antar dokumen.

Peneliti akan mencari nilai *cosine similarity* antara dokumen ke-4869 dengan dokumen ke-4923 menggunakan **Persamaan 3.4**. sebagai berikut:

$$\begin{aligned} \cos \alpha &= \frac{A \cdot B}{\|A\| \|B\|} \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \\ \cos \alpha &= \frac{A \cdot B}{\|A\| \|B\|} \frac{\sum_{i=1}^n D4869_i \times D4923_i}{\sqrt{\sum_{i=1}^n (D4869_i)^2} \times \sqrt{\sum_{i=1}^n (D4923_i)^2}} \\ \cos \alpha &= \frac{463.6961}{\sqrt{28.78567} \times \sqrt{25.79907}} = 0.897903 \end{aligned}$$

Nilai kemiripan dokumen ke-4869 terhadap dokumen ke-4923 yaitu sebesar 0.897903. Lalu peneliti akan mencari 5 dokumen yang paling mirip dengan dokumen ke-4869 dengan mencari nilai *cosine similarity* yang diurutkan dari nilai terbesar, seperti pada **Tabel 5.9**.

Tabel 5.9 Nilai *Cosine Similarity*

No	Dokumen	Nilai Cosine Similarity
1.	4923	0.897903
2.	5009	0.885389
3.	5034	0.883158
4.	4982	0.857375
5.	5021	0.804388

5.5 Hasil Rekomendasi

Berdasarkan hasil nilai *cosine similarity*, diperoleh rekomendasi produk *bodycare* yang mempunyai nilai *cosine similarity* atau nilai kemiripan yang paling besar. Pada dokumen ke-4869 yaitu produk *lotion* dari brand *Nivea* dengan nama produk yaitu “*Extra White Care & Protect Body Serum*” menampilkan 5 produk yang memiliki kemiripan yang tinggi dengan dokumen tersebut. Maka berdasarkan hasil dari **Tabel 5.9**, berikut merupakan 5 urutan teratas hasil rekomendasi pada produk *bodycare* yang tertera pada **Tabel 5.10**.

Tabel 5.10 Rekomendasi Produk 5 Teratas

No	Dokumen	Nama Produk	Nilai Cosine Similarity
1.	4923	<i>Healthy Bright Aloe Soothe</i>	0.897903
2.	5009	<i>Body Lotion</i>	0.885389
3.	5034	<i>Murumuru Butter & Rose Body Lotion</i>	0.883158
4.	4982	<i>Fragrance Brightening Body Lotion Jolly</i>	0.857375

5.	5021	<i>Sensational Body Lotion Vanilla & Almond Oil</i>	0.804388
----	------	---	----------

Berdasarkan hasil yang didapatkan pada **Tabel 5.9**, maka ketika pengguna melakukan pencarian pada produk “Extra White Care & Protect Body Serum”, akan muncul 5 rekomendasi produk bodycare yang mempunyai kemiripan dengan produk yang dicari berdasarkan hasil perhitungan dari cosine similarity. Kemiripan ini diperoleh dari review pengguna. Sebagai contoh, apabila pengguna produk “*Extra White Care & Protect Body Serum*”, menuliskan sebuah *review* yang memuat kesan dari penggunaan produk yang dipakai ataupun kondisi kulit yang sesuai dengan produk, maka hasil rekomendasi produk akan memiliki review pengguna yang mirip dengan produk “*Extra White Care & Protect Body Serum*”.

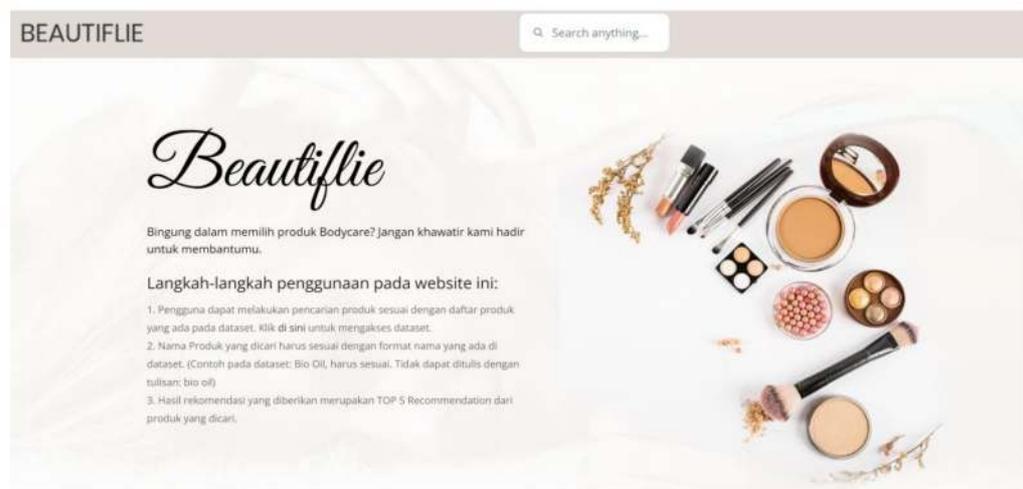
Tabel 5.11 *Review Produk “Extra White Care & Protect Body Serum” dan “Healthy Bright Aloe Soothe”*

Extra White Care & Protect Body Serum	Healthy Bright Aloe Soothe
Pakai body lotion ini semenjak aku kuliah, sekitar 10 tahun lalu. Udah repurchase puluhan kali. Dan suka banget pakai ini karena melembabkan di kulit badanku yang kering . Wanginya juga seger. Kalau beli produk ini jadi ada nilai nostalgia buat aku karena pakai ini terus sat dulu kuliah.	awalnya aku ditawari ini sama nenekku, dan ternyata cocok banget buat kulitku yang kering dan sensitifkalo terkena sinar matahari berlebih! dia emang nggak mencerahkan, tapi lembabnya juara!! dia bisa nyembuhin bekas luka. produknya gampang dicari di dekat rumah, tapi berkali2 repurchase

Berdasarkan **Tabel 5.11** dapat dilihat bahwa produk Review Produk “*Extra White Care & Protect Body Serum*” dan “*Healthy Bright Aloe Soothe*” mempunyai kemiripan pada *review* penggunaannya, hal ini selaras dengan nilai *cosine similarity* yang cukup besar yaitu 0.897903.

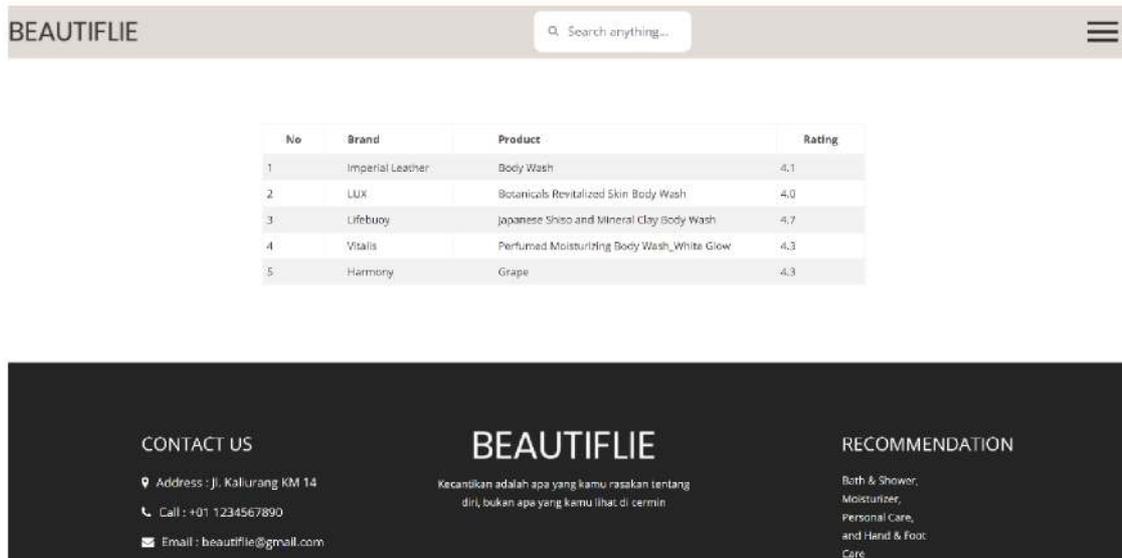
5.6 Tampilan Website

Dilakukan deploy pada sistem rekomendasi produk *bodycare*, yang dimana untuk menjalankan aplikasi atau layanan pada server localhost. Hal ini dilakukan guna mempermudah pengguna dalam mencari rekomendasi *bodycare* yang dapat diakses di <https://7ba1-2404-a640-956-461-9ca-20e5-6caf-cf5b.ngrok-free.app> . Pada halaman awal akan menampilkan halaman *Home* dimana pada bagian atas terdapat kolom pencarian yang digunakan untuk melakukan pencarian produk *bodycare* yang dibutuhkan oleh pengguna seperti yang ada pada **Gambar 5.4**.



Gambar 5.4 Tampilan Home

Lalu setelah melakukan pencarian didapatkan hasil lima rekomendasi yang dimana hasil rekomendasi ini dilengkapi dengan informasi mengenai *brand*, nama produk, dan rating seperti yang ada pada **Gambar 5.5**.



Gambar 5.5 Tampilan Hasil Rekomendasi

5.7 Evaluasi Model

Dilakukan sebuah evaluasi model guna mengukur sejauh mana sistem tersebut efektif dalam memberikan rekomendasi yang relevan terhadap pengguna dan dapat menilai kualitas rekomendasi yang diberikan oleh model. Evaluasi model ini juga memungkinkan untuk mengidentifikasi kelemahan dalam model rekomendasi dan dapat memperbaikinya.

Evaluasi Model pada produk *Healthy Bright UV Extra Brightening* yang menghasilkan lima rekomendasi produk yang dapat dilihat pada **Tabel 5.12**

Tabel 5.12 Evaluasi Model 1

No	Brand	Produk	Sesuai/Tidak sesuai	Nilai Kesesuaian
1.	Vaseline	Soft Glow Gluta Vitamin SPF 20 Body Serum	Sesuai	20%
2.	Nivea	Extra White Night Nourish Body Serum	Sesuai	20%
3.	Scarlett	Fragrance Brightening Body Lotion Jolly	Sesuai	20%

4.	Secret Garden	Body Lotion	Sesuai	20%
5.	Vaseline	Healthy Bright Perfect 10	Sesuai	20%
Total				100%

Pada hasil rekomendasi diatas didapatkan nilai akurasi kesesuaian sebesar 100%, menunjukkan bahwa untuk evaluasi pada model pertama sangat efektif dan akurat dalam memberikan saran produk kepada pengguna. Hal-hal yang mendasari kesesuaian antar produk dikarenakan hasil rekomendasi yang keluar memiliki kesamaan dari segi kategori yaitu produk *body lotion* serta karakteristik yang sama pada produk yang dicari.

Evaluasi Model pada produk *Extra White Night Nourish Body Serum* yang menghasilkan lima rekomendasi produk yang dapat dilihat pada **Tabel 5.13**

Tabel 5.13 Evaluasi Model 2

No	Brand	Produk	Sesuai/Tidak sesuai	Nilai Kesesuaian
1.	Nivea	Extra White Care & Protect Body Serum	Sesuai	20%
2.	Nivea	Extra White Night Nourish Body Lotion	Sesuai	20%
3.	The Body Shop	Mango Body Butter	Tidak sesuai	0%
4.	Aveeno	Soothing and Calming Moisturizing Lotion	Sesuai	20%
5.	Natur-E	Daily Nourishing Lotion_Revitalizing	Sesuai	20%
Total				80%

Pada hasil rekomendasi diatas didapatkan nilai akurasi kesesuaian sebesar 80%, hal ini menunjukkan bahwa evaluasi pada model kedua memiliki nilai akurasi yang cukup baik dalam memberikan rekomendasi kepada pengguna. Walaupun nilai yang didapatkan adalah 80% evaluasi pada model kedua ini memiliki potensi yang cukup baik dalam memberikan saran produk kepada pengguna.

Evaluasi Model pada produk *Deodorant Dry Serum Ultra Bright* yang menghasilkan lima rekomendasi produk yang dapat dilihat pada **Tabel 5.14**

Tabel 5.14 Evaluasi Model 3

No	Brand	Produk	Sesuai/Tidak sesuai	Nilai Kesesuaian
1.	Vaseline	Deodorant Dry Serum_Ultra Bright	Sesuai	20%
2.	Kaila	C-You Bright Here Axillary Cream	Sesuai	0%
3.	The Body Culture	Underarm Whitening Balm	Tidak sesuai	0%
4.	Dove	Ultimate Repair Antiperspirant Deodorant Roll On	Sesuai	20%
5.	Everwhite	Chomabright Axillary Cream	Sesuai	0%
Total				60%

Pada hasil rekomendasi diatas didapatkan nilai akurasi kesesuaian sebesar 60%, hal ini menunjukkan bahwa sebagian besar rekomendasi yang diberikan oleh model belum sepenuhnya sesuai tetapi masih cukup baik dalam memberikan rekomendasi.

Tabel 5.15 Total Akurasi Kesesuaian

Evaluasi Model	Nilai Kesesuaian
1	100%
2	80%
3	60%
Total	240%

$$Total\ Evaluasi\ Model = \frac{240}{3} = 80$$

Hasil evaluasi dari tiga model yang telah dilakukan menunjukkan bahwa keseluruhan kinerja pada model-model tersebut memberikan rekomendasi yang cukup baik, dengan total nilai akurasi mencapai 80%. Evaluasi model yang dilakukan ini menggunakan tiga sampel produk dari keseluruhan produk yang ada. Hal ini mengindikasikan bahwa model-model tersebut berhasil dalam memberikan rekomendasi atau prediksi yang sebagian besar sesuai dengan data yang ada atau berdasarkan preferensi pengguna.

Evaluasi ini merupakan langkah awal yang baik dalam mengukur efektivitas model-model tersebut, dan dapat digunakan sebagai dasar untuk terus memperbaiki serta mengoptimalkan kinerja dari sistem rekomendasi sehingga dapat memberikan rekomendasi yang lebih akurat dan relevan bagi pengguna.

BAB 6 PENUTUP

6.1 Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan pada penelitian ini, didapatkan kesimpulan sebagai berikut:

1. Diketahui terdapat 10 brand *bodycare* yang populer pada *beauty platform Female Daily* yaitu brand dari The Body Shop, Nivea, Vaseline, Herborist, Shinzui, Oriflame, Citra, Dove, Scarlett, dan yang terakhir adalah Roro Mendut.
2. Pada penelitian ini hasil rekomendasi yang didapatkan berdasarkan review memberikan lima rekomendasi produk secara berturut-turut sebagai contoh pada produk *Extra White Care & Protect Body Serum* menghasilkan rekomendasi produk *bodycare* yaitu *Healthy Bright Aloe Soothe* dari Vaseline, *Body Lotion* dari Secret Garden, *Murumuru Butter & Rose Body Lotion* dari Love Beauty and Planet, *Fragrance Brightening Body Lotion Jolly* dari Scarlett, dan *Sensational Body Lotion Vanilla & Almond Oil* dari Nivea. Dengan nilai *cosine similarity* berturut-turut sebesar 0.897903, 0.885389, 0.883158, 0.857375, dan 0.804388.

6.2 Saran

Beberapa saran yang peneliti berikan berdasarkan penelitian ini adalah sebagai berikut:

1. Pada penelitian selanjutnya disarankan dapat membandingkan dengan jenis metode lain seperti metode *Collaborative Filtering*, *Hybrid Recommendation System*.
2. Pada penelitian selanjutnya diharapkan dapat menambahkan *stopwords* yang tidak tercantum pada *library*.
3. Untuk penelitian selanjutnya apabila menggunakan 10 *review* pada 1 produk diharapkan 10 *review* itu digabungkan semua pada 1 produk yang sama.

DAFTAR PUSTAKA

- umnadmin. (2023, April 18). *Female Daily Goes To Campus With Innisfree: Mempersiapkan Mindvesment dan Skinvesment di Dunia Kerja*. Retrieved from Universitas Multimedia Nusantara: <https://www.umn.ac.id/female-daily-goes-to-campus-with-innisfree-mempersiapkan-mindvesment-dan-skinvesment-di-dunia-kerja-2/#:~:text=Female%20Daily%20merupakan%20digital%20platform,yang%20telah%20berdiri%20dari%202005>.
- A.S, R., & Shalahuddin, M. (2018). *Rekayasa perangkat lunak terstruktur dan berorientasi objek edisi revisi*. Informatika Bandung.
- Achroni, K., & Famadiani, C. (2017). *Semua Rahasia Kulit Cantik & Sehat Ada di Sini*. Yogyakarta: Javalitera.
- Afrilawati, R. (2022). ANALISIS PELANGGARAN PENJUALAN BODYCARE NON BPOM PADA SHOPPE.
- Apriani, Zakiyudin, H., & Marzuki, K. (2021). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Barupada Kampus Swasta. *Bumigora Information Technology (BITe)*, 19-27.
- Ariantini, D. A., Lumenta, A. S., & Jacobus, A. (2016). PENGUKURAN KEMIRIPAN DOKUMEN TEKS BAHASA INDONESIA MENGGUNAKAN METODE COSINE SIMILARITY . *E-Journal Teknik Informatika*, 2.
- armeliafarah. (2023, Agustus 30). *Dukung Beauty Products Favoritmu, Ini Cara Voting di Female Daily Best of Beauty Awards 2023!* Retrieved from Female Daily: <https://editorial.femaledaily.com/blog/2023/08/30/ini-cara-voting-di-female-daily-best-of-beauty-awards-2023>
- Aulia, G. N., & Patriya, E. (2019). IMPLEMENTASI LEXICON BASED DAN NAIVE BAYES PADA ANALISIS SENTIMEN PENGGUNA TWITTER TOPIK PEMILIHAN PRESIDEN 2019. *Jurnal Ilmiah Informatika Komputer*.

- BBC. (2014, December 30). *AirAsia QZ8501: Does bad weather cause plane crashes?* Retrieved from BBC: <http://www.bbc.com/news/world-30631968>
- Cahyono, Y. (2017). ANALISIS SENTIMENT PADA SOSIAL MEDIA TWITTER MENGGUNAKAN NAÏVE BAYES CLASSIFIER DENGAN FEATURE SELECTION PARTICLE SWARM OPTIMIZATION DAN TERM FREQUENCY. *JURNAL INFORMATIKA UNIVERSITAS PAMULANG*, 16.
- Deni Rusdian, D. R. (2019). ANALISA SENTIMEN TERHADAP TOKOH PUBLIK MENGGUNAKAN. *Journal of Computer Engineering System and Scienc*, Vol. 4 No. 2.
- Dzumiroh, L., & Saptono, R. (2012). Penerapan Metode Collaborative Filtering Menggunakan Rating Implisit pada Sistem Rekomendasi Pemilihan Film di Rental VCD. *Jurnal ITSMART*, 55.
- Ekaputri, S. D., Karolina, C. M., & Abdullah, N. N. (2020). LITERASI INFORMASI KECANTIKANDI KALANGAN PENGGUNA FEMALE DAILY NETWORK. *JURNAL DOKUMENTASI DAN INFORMASI*, 113.
- Ekaputri1, S. D., Karolina2, C. M., & Abdullah3, N. N. (2020). LITERASI INFORMASI KECANTIKAN DI KALANGAN PENGGUNA FEMALE DAILY NETWORK. *Jurnal Dokumentasi dan Informasi*, 112.
- Fajriansyah, M., Adikara, P. P., & Widodo, A. W. (2021). Sistem Rekomendasi Film Menggunakan Content Based Filtering. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2188-2199.
- Fajriansyah, M., Adikara, P. P., & Widodo, A. W. (2021). Sistem Rekomendasi Film Menggunakan Content Based Filtering. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2188-2199.
- Fajriansyah, M., Fajriansyah, M., & Fajriansyah, M. (2021). Sistem Rekomendasi Film Menggunakan Content Based Filtering. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2188-2199.
- Fauzi, A. R., & Nurmalina, R. (2012). *Merawat Kulit & Wajah*. Jakarta: PT Elex Media Komputindo.

- Filcha, A., & Hayaty, M. (2019). Implementasi Algoritma Rabin-Karp untuk Pendeteksi Plagiarisme pada Dokumen Tugas Mahasiswa (Rabin-Karp Algorithm Implementation to Detect Plagiarism on Student's Assignment Document). *Jurnal Informatika*.
- Hari, S. N., Rostamailis, & Astuti, M. (2015). TUBUH, PENGGUNAAN LULUR ZAITUN TERHADAP PERAWATAN KULIT TUBUH. *Journal of Home Economics and Tourism*, 4.
- Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 306-312.
- Johnson, R. A., & Bhattacharyya, G. K. (2010). *Statistics Principles & Methods*. USA: John Wiley & Sons.
- Kristanto, M. C. (2020). PENENTUAN GAYA DESAIN UNTUK MEMBANGUN BRAND IMAGE PRODUK BODY CARE "REVEUSE". 74.
- Langgeni, Baizal, & Firdaus. (2010). CLUSTERING ARTIKEL BERITA BERBAHASA INDONESIA MENGGUNAKAN UNSUPERVISED FEATURE SELECTION. *Seminar Nasional Informatika, Yogyakarta*.
- Lee, S. Y. (2019). Document vectorization method using network information of words. *Document vectorization method using network information of words*.
- Li, X. (2013). Comparison and Analysis between Holt Exponential Smoothing and Brown Exponential Smoothing Used for Freight Turnover Forecast. *Third International Conference on Intelligent System Design and Engineering Applications* (pp. 453-456). IEEE.
- Maarif, A. A. (2015). PENERAPAN ALGORITMA TF-IDF UNTUK PENCARIAN KARYA ILMIAH.
- Marthasari, G. I., Azhar, Y., & Puspitaningrum, D. K. (2015). SISTEM REKOMENDASI PENYEWAAN PERLENGKAPAN PESTA MENGGUNAKAN COLLABORATIVE FILTERING DAN PENGGALIAN ATURAN ASOSIASI. *Jurnal Simantec*.

- Maulida, N. A., Arsi, A. A., & Alimi, M. Y. (2018). Redefinisi Konsep Maskulinitas Laki-Laki Pengguna Perawatan Kulit Di Klinik Kecantikan Armina Desa Robayan Jepara. *Journal of Education, Society and Culture*, 345.
- Maxmanroe. (2014, Agustus 8). *Female Daily ~ Startup Sukses yang Berawal dari Hobi Blogging*. Retrieved Agustus 6, 2023, from <https://www.maxmanroe.com/>: <https://www.maxmanroe.com/female-daily-startup-sukses-yang-berawal-dari-hobi-blogging.html>
- Mondi, R. H., Wijayanto, A., & Winarno. (2019). RECOMMENDATION SYSTEM WITH CONTENT-BASED FILTERING METHOD FOR CULINARY TOURISM IN MANGAN APPLICATION. *Jurnal Ilmiah Teknologi dan Informasi*.
- Mulianingsih, A. M., & Ambarwati, N. S. (2021). PEMANFAATAN LIDAH BUAYA (ALOE VERA) SEBAGAI BAHAN BAKU PERAWATAN KECANTIKAN KULIT. *Jurnal Tata Rias*.
- Mulyana, D. I., & Lutfianti, N. (2023). Analisis Sentimen Dengan Algoritma SVM Dalam. *JURNAL SISFOTENIKA*, 55.
- Nurjannah, M., Hamdani, & Astuti, I. F. (2013). PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING. *Jurnal Informatika Mulawarman*, 111.
- Prasetyo, E. (2012). *Data mining : konsep dan aplikasi menggunakan MATLAB*. CV Andi Offset.
- Ratino, Hafidz, N., Anggraeni, S., & Gata, W. (2020). Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes. *Jurnal JUPITER*, 1-11.
- Revitasari, D. (2022). *Latar Belakang Bodycare*, 1.
- Riyan, A., Naf'an, M. Z., & Burhanuddin, A. (2019). Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen. *Jurnal Linguistik Komputasional*, 24.
- Santoso, V. I., Virginia, G., & Lukito, Y. (2017). PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE

- SUPPORT VECTOR MACHINE. *JURNAL TRANSFORMATIKA*, Volume 14, Nomor 2.
- Sebastia, L., Garcia, I., Onaindia, E., & Guzman, C. (2009). e-TOURISM : A TOURIST RECOMMENDATION AND PLANNING APPLICATION. *International Journal on Artificial Intelligence Tools*, 717-738.
- Seelam, S. (2021, Mei 26). *Machine Learning Fundamentals: Cosine Similarity and Cosine Distance*. Retrieved from Medium: <https://medium.com/geekculture/cosine-similarity-and-cosine-distance-48eed889a5c4>
- Sholeh hudin, M., Fauzi, M. A., & Adinugroho, S. (2018). Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi (Studi Kasus: Universitas Brawijaya). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5518-5524.
- Simatupang, M. P., & Utomo, D. P. (2019). ANALISA TESTIMONIAL DENGAN MENGGUNAKAN ALGORITMA TEXT MINING DAN TERM FREQUENCY- INVERSE DOCUMENT FREQUENCY (TF-IDF) PADA TOKO ALLMEEART. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 809.
- Siregar, R. R., Sinaga, F. A., & Arianto, R. (2017). APLIKASI PENENTUAN DOSEN PENGUJI SKRIPSI MENGGUNAKAN METODE TF-IDF DAN VECTOR SPACE MODEL. *Journal of Computer Science and Information Systems*, 1, 171-186.
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2011). *Probability & Statistics for Engineers & Scientists 9th Ed*. USA: Pearson.
- Yulian, E. (2018). Text Mining dengan K-Means Clustering pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung. *Jurnal MATEMATIKA "MANTIK"*, 55.

LAMPIRAN

Lampiran 1 Dataset dan *Script Python*

<https://bit.ly/DatasetAvisa>