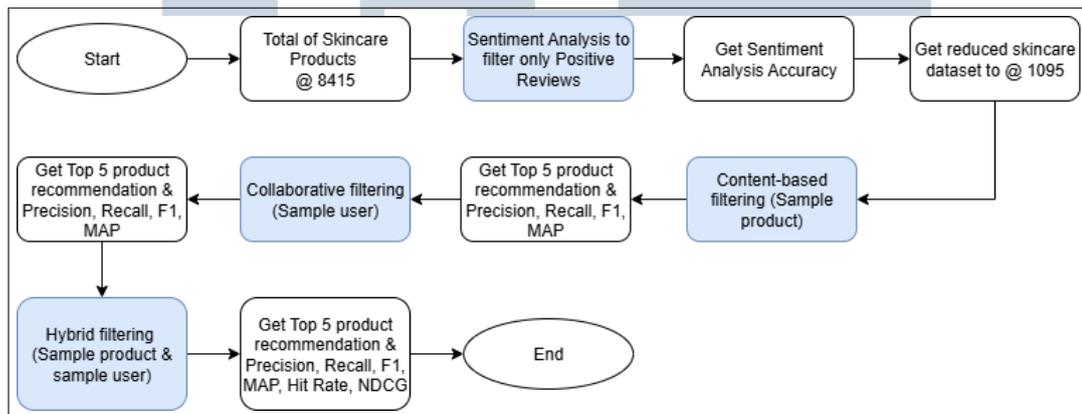


BAB 3 METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

Penelitian dengan judul "Implementasi Hybrid Filtering dan Analisis Sentimen untuk Rekomendasi Produk Skincare" dilakukan secara bertahap. Tahapan ini memberikan gambaran tentang objek penelitian, analisis semua permasalahan yang ada, dan pendekatan yang digunakan dalam penelitian.



Gambar 3.1. Alur Penelitian

Gambar 3.1 menggambarkan alur penelitian yang dilakukan dalam pengembangan sistem rekomendasi produk skincare berbasis analisis sentimen. Proses dimulai dengan dataset awal sebanyak 8.415 data produk skincare. Selanjutnya dilakukan analisis sentimen untuk menyaring hanya ulasan dengan sentimen positif sehingga menghasilkan dataset tereduksi sebanyak 1.095 data. Akurasi model analisis sentimen juga dihitung untuk mengukur kinerja klasifikasi sentimen yang menjadi dasar rekomendasi.

Setelah data dipersiapkan, dilakukan tiga pendekatan sistem rekomendasi: *Content-Based Filtering* (menggunakan sampel produk), *Collaborative Filtering* (menggunakan sampel pengguna), dan *Hybrid Filtering* yang merupakan kombinasi dari keduanya. Evaluasi sistem dilakukan terhadap rekomendasi top-5 produk menggunakan metrik *Precision*, *Recall*, *F1-Score* dan *MAP*. Pada pendekatan *Hybrid Filtering*, digunakan pula metrik tambahan seperti *Hit Rate* dan *Normalized Discounted Cumulative Gain (NDCG)*.

Pada subbab ini juga dilaporkan secara detail rancangan terhadap penelitian

yang dilakukan, baik perancangan secara umum dari sistem yang dibangun, maupun perancangan yang lebih spesifik. Berikut metodologi yang digunakan:

1. Studi Literatur

Penelitian ini disertai dengan adanya studi literatur. Studi literatur yang dilakukan adalah dengan mengumpulkan informasi berdasarkan penelitian yang telah dilakukan sebelumnya. Pada penelitian ini literatur yang digunakan merupakan literatur-literatur yang berhubungan dengan sistem rekomendasi hybrid (*Content-Based dan Collaborative Filtering*), analisis sentimen pada ulasan produk skincare, evaluasi metrik (RMSE, F1-score, *diversity index*), serta algoritma kunci yaitu Support Vector Machine (SVM), Naive Bayes, dan Hybrid Filtering. Literatur-literatur yang digunakan telah dikumpulkan dan menjadi referensi pada penelitian ini.

2. Pengumpulan Data

Dataset utama diperoleh dari platform Kaggle, mencakup metadata lebih dari 8000 produk skincare dan sekitar 1 juta ulasan pengguna dari situs Sephora. Data dikumpulkan melalui proses *web scraping* menggunakan Python pada Maret 2023 dan diperoleh dari platform Kaggle.

3. Perancangan Model

Tahapan ini merupakan proses perancangan sistem rekomendasi produk perawatan kulit yang mencakup pengolahan data, pemilihan fitur, serta penerapan pendekatan pemodelan yang sesuai. Sistem dirancang untuk menggabungkan tiga pendekatan utama, yaitu *analisis sentimen*, *content-based filtering* dan *collaborative filtering*, yang kemudian digabungkan dalam metode *hybrid filtering* untuk menghasilkan rekomendasi yang lebih relevan.

4. Implementasi Model

Tahapan ini dilakukan untuk merealisasikan rancangan sistem rekomendasi produk perawatan kulit menjadi sebuah sistem yang dapat dijalankan. Proses diawali dengan pemuatan dataset, yang kemudian dilanjutkan dengan tahap data preprocessing untuk membersihkan dan menyiapkan data agar siap digunakan. Selanjutnya, dilakukan proses *feature engineering* dan *exploratory data analysis* (EDA) untuk memahami karakteristik data secara menyeluruh dan mengekstraksi informasi penting. Tahapan berikutnya adalah *sentiment analysis* untuk memahami opini pengguna melalui ulasan produk positif, yang kemudian menjadi pintu pertama dalam sistem

rekomendasi. Sistem ini menggabungkan pendekatan *content-based filtering* dan *collaborative filtering*, yang selanjutnya disatukan melalui metode *hybrid filtering* untuk memberikan rekomendasi yang lebih relevan.

5. Pengujian dan Evaluasi Model

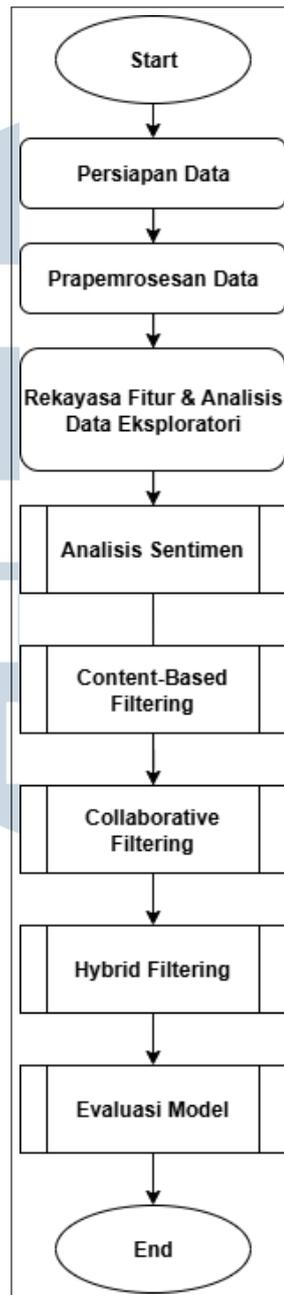
Pada tahapan ini, sistem rekomendasi yang telah dibangun akan diuji untuk mengetahui kinerja dari masing-masing komponen, baik dari hasil analisis sentimen maupun metode rekomendasi yang digunakan. Setelah pengujian selesai, dilakukan evaluasi terhadap hasil yang diperoleh untuk melihat sejauh mana sistem mampu memberikan rekomendasi produk yang sesuai dengan preferensi pengguna. Evaluasi ini juga menjadi dasar dalam menentukan keunggulan pendekatan hybrid dibandingkan metode lainnya yang digunakan dalam penelitian ini.

6. Penulisan Laporan

Tahapan ini merupakan tahap akhir dalam proses penelitian, di mana seluruh proses yang telah dilakukan mulai dari perancangan hingga implementasi sistem direkam dalam bentuk dokumentasi tertulis. Laporan ini disusun secara sistematis dan terstruktur. Penulisan laporan ini bertujuan untuk memberikan gambaran menyeluruh mengenai proses penelitian dan hasil yang telah dicapai.

3.2 Perancangan Model

Gambar 3.2 merupakan *flowchart* model secara keseluruhan. Pada bagian ini terdapat proses yang dilakukan ketika melakukan perancangan terhadap sistem rekomendasi skincare menggunakan analisis sentimen dan hybrid filtering sampai dengan evaluasi terhadap model. Pada tahapan ini terdapat *flowchart* yang menjadi dasar dari perancangan model ini.



Gambar 3.2. Flowchart Perancangan Model Keseluruhan

1. *Data Loading*

Pada tahap awal ini, peneliti melakukan proses pengumpulan dan persiapan data yang menjadi fondasi pengembangan sistem rekomendasi. Dataset utama diperoleh dari platform Kaggle dan terdiri dari dua komponen penting, yaitu:

- (a) **Informasi Produk:** Berisi data lebih dari 8000 produk skincare yang

tersedia di platform Sephora, seperti nama produk, merek, harga, komposisi bahan aktif, kategori produk, dan rating keseluruhan.

- (b) **Ulasan Pengguna:** Berisi sekitar 1.000.000 ulasan berbahasa Inggris, mencakup teks *review*, rating individu, tipe kulit, warna kulit, indikator rekomendasi produk, dan waktu pengiriman ulasan.

Dataset tersebut dikumpulkan melalui proses *web scraping* menggunakan Python pada bulan Maret 2023. Setelah dikumpulkan, kedua komponen dataset digabungkan berdasarkan identitas produk untuk membentuk satu dataset komprehensif yang digunakan dalam proses analisis dan pengembangan model rekomendasi hybrid berbasis konten produk dan preferensi pengguna. Selain itu, studi literatur pendukung turut dilakukan guna memvalidasi metodologi pengolahan data, khususnya terkait teknik ekstraksi bahan aktif kosmetik, klasifikasi warna kulit (misalnya Fitzpatrick Scale), dan pendekatan integrasi data ulasan dalam sistem rekomendasi.

2. *Data Preprocessing*

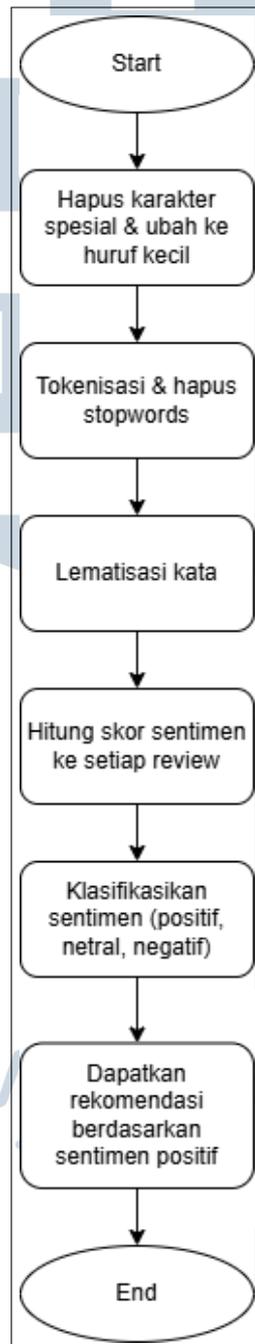
Data preprocessing merupakan tahap penting untuk memastikan kualitas data sebelum dilakukan pemodelan. Proses ini mencakup penghapusan duplikat agar setiap entri unik, penanganan nilai kosong melalui penghapusan atau imputasi sesuai konteks atribut, serta identifikasi outlier untuk mengurangi potensi distorsi dalam pemodelan, khususnya pada variabel kunci seperti harga. Dengan pembersihan data yang sistematis, sistem rekomendasi dapat dibangun di atas dataset yang bersih dan representatif.

3. *Feature Engineering & EDA*

Tahap Feature Engineering dan Exploratory Data Analysis (EDA) dilakukan untuk memastikan data yang digunakan relevan dan informatif bagi model. Feature Engineering mencakup pembuatan fitur baru seperti klasifikasi tipe kulit dan segmentasi harga untuk memperkaya informasi yang tersedia. Tahap EDA dilakukan untuk memahami karakteristik distribusi dan hubungan antar variabel dalam dataset. Analisis data dilakukan untuk mengidentifikasi interaksi kompleks antar fitur melalui visualisasi seperti *heatmap* dan *pairplot*. Temuan-temuan dari proses ini menjadi dasar dalam pengambilan keputusan untuk pemodelan sistem rekomendasi yang lebih tepat sasaran.

4. *Sentiment Analysis (VADER)*

Flowchart pada Gambar 3.3 menunjukkan alur proses analisis sentimen menggunakan pendekatan leksikal VADER (Valence Aware Dictionary and sEntiment Reasoner). Tujuan utama proses ini adalah untuk mengategorikan ulasan produk ke dalam sentimen positif, netral, atau negatif, serta memberikan rekomendasi produk berdasarkan ulasan positif pengguna.



Gambar 3.3. Flowchart Proses Sentimen Analisis

Setelah teks ulasan diproses, setiap review diberi skor sentimen menggunakan metode VADER. Skor ini menunjukkan kekuatan dan arah sentimen ulasan, dengan rentang dari -1 (sangat negatif) hingga +1 (sangat positif). Berdasarkan skor *compound* yang dihasilkan, sistem kemudian mengklasifikasikan ulasan ke dalam kategori sentimen positif, netral, atau negatif.

Skor di atas 0.05 dianggap sebagai sentimen positif. Ulasan-ulasan yang dikategorikan sebagai positif inilah yang selanjutnya dijadikan dasar untuk menghasilkan rekomendasi produk. Dengan demikian, sistem dapat menyarankan produk-produk yang telah mendapatkan ulasan baik dan dinilai memuaskan oleh pengguna lain.

Perhitungan Manual Sentimen Analisis VADER

VADER (Valence Aware Dictionary and sEntiment Reasoner) merupakan metode berbasis leksikon yang menghitung skor sentimen berdasarkan intensitas kata, pengingkalan, dan pola konteks. Berikut perhitungan manual untuk tiga contoh ulasan:

Skor sentimen dihitung menggunakan fungsi:

$$\text{compound} = \frac{s}{\sqrt{s^2 + \alpha}} \quad \text{dimana} \quad \alpha = 15$$

dengan s adalah skor mentah sebelum normalisasi.

$$\text{Kategori} = \begin{cases} \text{Positif} & \text{jika } \text{compound} \geq 0.05 \\ \text{Netral} & \text{jika } -0.05 < \text{compound} < 0.05 \\ \text{Negatif} & \text{jika } \text{compound} \leq -0.05 \end{cases}$$

(a) Contoh Ulasan Positif

Teks: *"Love it! It works great, and smells delicious!"*

i. Tokenisasi dan Pencocokan Leksikon:

- Love = +3.1
- great = +3.0

- delicious = +2.3
- ! = +0.292 (intensifier)

ii. Skor Mentah:

$$s = (3.1 + 3.0 + 2.3) \times (1 + 0.292) = 8.4 \times 1.292 = 10.8528$$

iii. Normalisasi:

$$\text{compound} = \frac{10.8528}{\sqrt{10.8528^2 + 15}} = \frac{10.8528}{\sqrt{117.78 + 15}} = \frac{10.8528}{11.52} \approx 0.942$$

iv. Klasifikasi:

$$0.942 > 0.05 \implies \text{Positif}$$

(b) Contoh Ulasan Negatif

Teks: *"Overpriced and overhyped—did nothing for my very dry skin."*

i. Tokenisasi dan Pencocokan Leksikon:

- Overpriced = -2.5
- overhyped = -2.0
- nothing = -1.0
- very dry = -1.5 (intensifier + kata dasar)
- | = +0.18 (penekanan)

ii. Skor Mentah:

$$s = (-2.5) + (-2.0) + (-1.0) + (-1.5) + 0.18 = -6.82$$

iii. Normalisasi:

$$\text{compound} = \frac{-6.82}{\sqrt{(-6.82)^2 + 15}} = \frac{-6.82}{\sqrt{46.51 + 15}} = \frac{-6.82}{7.84} \approx -0.870$$

iv. Klasifikasi:

$$-0.870 < -0.05 \implies \text{Negatif}$$

Tabel 3.1. Hasil Perhitungan Manual VADER

Ulasan	Skor Mentah (s)	Compound	Kategori
Positif	10.8528	0.942	Positif
Negatif	-6.82	-0.870	Negatif

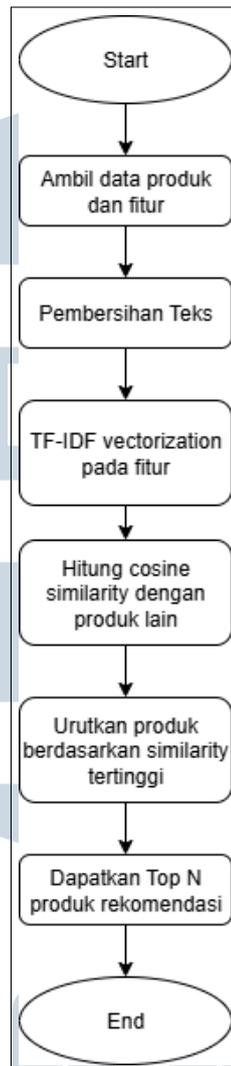
Evaluasi Sentimen Analisis

Proses evaluasi model dilakukan setelah model LSTM selesai dilatih menggunakan data pelatihan. Evaluasi ini bertujuan untuk mengetahui sejauh mana model LSTM mampu mengklasifikasikan sentimen ulasan dengan akurat. Model yang telah dibangun dievaluasi menggunakan data uji yang telah dipisahkan sebelumnya pada tahap *split dataset*. Hasil evaluasi ini ditampilkan dalam bentuk nilai akurasi yang mencerminkan tingkat ketepatan prediksi model terhadap data uji. Proses ini menjadi indikator penting untuk menilai kemampuan generalisasi model terhadap data baru, serta membantu dalam mengidentifikasi potensi *overfitting* atau *underfitting* selama pelatihan.

5. Content-Based Filtering

Gambar 3.4 menggambarkan alur sistem rekomendasi berbasis konten (*Content-Based Filtering*) yang digunakan untuk menyarankan produk-produk perawatan kulit serupa berdasarkan kemiripan fitur, khususnya kandungan bahan produk. Pendekatan ini memungkinkan sistem memberikan rekomendasi personal berdasarkan karakteristik produk yang telah disukai pengguna.

U M M N
 U N I V E R S I T A S
 M U L T I M E D I A
 N U S A N T A R A



Gambar 3.4. Flowchart Content-Based Filtering

Proses *content-based filtering* dimulai dengan mengambil data produk beserta fitur deskriptifnya, khususnya daftar kandungan (*ingredients*), sebagai dasar pengukuran kemiripan antar produk. Data teks kemudian dibersihkan dan diproses melalui tahap *text cleaning* untuk menghilangkan elemen yang tidak relevan. Selanjutnya, dilakukan vektorisasi menggunakan metode TF-IDF untuk mengubah informasi kandungan menjadi bentuk numerik yang dapat dihitung.

Berdasarkan representasi TF-IDF tersebut, sistem menghitung kemiripan antar produk menggunakan *cosine similarity*. Untuk menjaga kualitas rekomendasi, hanya produk dengan ulasan positif berdasarkan hasil analisis sentimen yang dipertimbangkan. Produk-produk yang memiliki kemiripan

tertinggi akan disusun dalam urutan dan ditampilkan sebagai Top-N rekomendasi kepada pengguna, dilengkapi dengan informasi utama seperti nama, merek, harga, dan komposisi.

6. Collaborative Filtering

Gambar 3.5 menunjukkan alur proses sistem rekomendasi dengan pendekatan Collaborative Filtering berbasis model-based menggunakan teknik Singular Value Decomposition (SVD). Pendekatan ini merekomendasikan produk perawatan kulit kepada pengguna berdasarkan pola interaksi dan preferensi pengguna lain yang memiliki kesamaan.



Gambar 3.5. Flowchart Collaborative Filtering

Proses *Collaborative Filtering* dimulai dengan membangun matriks interaksi antara pengguna dan produk berdasarkan data rating. Matriks ini merepresentasikan hubungan eksplisit yang digunakan untuk mempelajari preferensi pengguna. Untuk menangkap pola tersembunyi dari interaksi tersebut, sistem menerapkan metode *Singular Value Decomposition* (SVD) guna mereduksi dimensi dan mengekstraksi fitur laten dari pengguna maupun produk.

Setelah proses dekomposisi, sistem memproyeksikan pengguna dan produk ke dalam ruang fitur berdimensi rendah dan memprediksi rating menggunakan operasi dot product antar vektor. Berdasarkan hasil prediksi, sistem menyusun daftar rekomendasi Top-N produk yang paling relevan bagi masing-masing pengguna.

7. *Hybrid Filtering*

Gambar 3.6 menunjukkan alur proses sistem rekomendasi dengan pendekatan *Hybrid Filtering* yang menggabungkan hasil dari metode *Content-Based Filtering* (CBF) dan *Collaborative Filtering* (CF). Pendekatan ini bertujuan untuk memanfaatkan kelebihan masing-masing metode: CBF yang mempertimbangkan kemiripan antar produk berdasarkan fitur deskriptif, serta CF yang memperhitungkan preferensi pengguna lain yang serupa.





Gambar 3.6. Flowchart Hybrid Filtering

Pendekatan Hybrid Filtering menggabungkan kekuatan Content-Based Filtering (CBF) dan Collaborative Filtering (CF) untuk menghasilkan rekomendasi yang lebih akurat dan menyeluruh. Sistem menggabungkan hasil rekomendasi dari kedua metode, lalu menyaring duplikasi untuk memastikan setiap produk hanya muncul satu kali dalam daftar akhir.

Selanjutnya, sistem menghitung skor gabungan untuk setiap produk berdasarkan peringkat dari CBF dan CF, dengan menerapkan bobot tertentu agar kontribusi keduanya seimbang sesuai konteks. Daftar rekomendasi akhir disusun berdasarkan skor tertinggi dan ditampilkan sebagai Top-N produk yang paling relevan bagi pengguna.

8. *Evaluate Model*

Evaluasi model dalam penelitian ini bertujuan untuk mengukur sejauh

mana sistem mampu memberikan rekomendasi produk yang relevan bagi pengguna. Evaluasi dilakukan menggunakan beberapa metrik perankingan, yaitu *Precision@K*, *Recall@K*, *Hit Rate@K*, *NDCG@K*, dan *MAP@K* dengan nilai K bervariasi (1, 5, 10, dan 20).

Relevansi produk ditentukan berdasarkan nilai rating; produk dengan rating di atas ambang batas (diatas 3) dianggap relevan. Berdasarkan data tersebut, sistem mengukur presisi dan cakupan rekomendasi, menghitung seberapa sering produk relevan muncul di daftar teratas, dan menilai kualitas urutan melalui bobot posisi (NDCG). Dengan pendekatan ini, performa sistem dapat dievaluasi secara menyeluruh dalam berbagai skenario jumlah rekomendasi.

