

BAB I

PENDAHULUAN

1.1 Latar Belakang

Industri kosmetik di Indonesia saat ini sedang tumbuh sangat pesat dan diprediksi akan terus meningkat nilai pasarnya secara berkelanjutan dari tahun ke tahun [1]. Pertumbuhan tersebut mendorong munculnya berbagai produk skincare dengan fungsi, kandungan bahan, serta klaim manfaat yang semakin beragam. Di sisi lain, perilaku konsumen dalam memilih produk kosmetik juga mengalami perubahan. Konsumen tidak lagi membeli produk hanya berdasarkan popularitas merek, tetapi semakin memperhatikan berbagai informasi yang tersedia sebelum melakukan pembelian. Informasi mengenai kandungan bahan (*ingredients*), manfaat produk, keamanan penggunaan, serta kesesuaian dengan kebutuhan kulit menjadi faktor penting dalam proses pengambilan keputusan [2].

Banyaknya variasi produk yang tersedia menyebabkan konsumen dihadapkan pada jumlah informasi yang sangat besar ketika akan memilih produk skincare. Kondisi ini menimbulkan fenomena *information overload*, yaitu situasi ketika pengguna harus menganalisis terlalu banyak pilihan dan informasi sebelum menentukan produk yang sesuai. Konsumen perlu membandingkan berbagai atribut produk, mulai dari komposisi bahan, fungsi produk, kategori penggunaan, hingga karakteristik visual kemasan. Akibatnya, proses pencarian produk yang sesuai sering kali menjadi lebih sulit dan memerlukan waktu yang lebih lama.

Untuk membantu proses pemilihan produk, berbagai sistem rekomendasi telah dikembangkan. Namun, sistem rekomendasi skincare masih menghadapi tantangan dalam memahami karakteristik produk secara menyeluruh. Sebagian sistem hanya memanfaatkan informasi tekstual, sedangkan sebagian lainnya hanya berfokus pada informasi visual [3]. Padahal, produk skincare memiliki karakteristik multidimensi yang terdiri atas kandungan bahan dan tampilan kemasan yang sama-sama berperan dalam proses identifikasi maupun pencarian produk alternatif. Pemanfaatan satu jenis informasi saja berpotensi menghasilkan rekomendasi yang

kurang mampu merepresentasikan karakteristik produk secara utuh.

Salah satu pendekatan yang dapat digunakan untuk mengatasi keterbatasan tersebut adalah *hybrid multimodal* yang menggabungkan pengolahan data teks dan citra. Pada data teks, metode TF-IDF dapat digunakan untuk memberikan bobot pada kata-kata penting yang terdapat dalam komposisi bahan produk sehingga informasi yang relevan dapat direpresentasikan dengan baik [4]. Pada data citra, arsitektur ResNet-50 mampu mengekstraksi fitur visual yang kompleks dan telah terbukti memiliki kinerja yang baik dalam berbagai tugas pengolahan citra [5]. Kombinasi kedua pendekatan tersebut membuat sistem memperoleh representasi produk yang lebih komprehensif dibandingkan penggunaan satu jenis data saja.

Berdasarkan permasalahan tersebut, penelitian ini mengimplementasikan sistem rekomendasi *hybrid content-based filtering* dengan mengintegrasikan ekstraksi fitur teks menggunakan TF-IDF dan ekstraksi fitur citra menggunakan ResNet-50. Tingkat kemiripan antarproduk dihitung menggunakan *cosine similarity* untuk menghasilkan rekomendasi yang relevan. Sistem yang dikembangkan diharapkan mampu memberikan rekomendasi produk yang memiliki kemiripan baik dari sisi kandungan bahan maupun karakteristik visual sehingga dapat membantu pengguna menemukan alternatif produk yang lebih relevan. Hasil penelitian kemudian direalisasikan dalam bentuk prototype berbasis web sebagai simulasi penyajian rekomendasi produk alternatif bagi pengguna.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang mengenai fenomena kelebihan informasi (information overload) dalam memilih produk skincare, rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan metode *Hybrid Content-Based Filtering* yang mengintegrasikan ekstraksi fitur citra menggunakan ResNet50 dan fitur teks menggunakan TF-IDF pada sistem rekomendasi *skincare*?
2. Bagaimana performa dan tingkat akurasi dari hasil rekomendasi produk skincare yang dihasilkan sistem berdasarkan tingkat kemiripan (*similarity*)

dari penggabungan fitur visual kemasan dan komposisi bahan yang dievaluasi menggunakan *Precision@5*, *Silhouette Score*, dan visualisasi *t-SNE*?

1.3 Batasan Masalah

Pembatasan masalah pada penelitian ini ditetapkan sebagai berikut:

1. Data sekunder bersumber dari dataset publik SkinSafe Kaggle. Populasi data dibatasi 4.800 sampel produk. Pengolahan atribut difokuskan pada citra kemasan (*image*), teks komposisi bahan (*ingredients*). Pendekatan sistem murni mengimplementasikan metode *Content-Based Filtering*. Variabel interaksi pengguna berupa rating, review, riwayat pembelian diabaikan.
2. Arsitektur *ResNet50* mengekstraksi fitur citra. Metode pembobotan *TF-IDF* mengekstraksi fitur teks. Algoritma *Cosine Similarity* mengkalkulasi tingkat kemiripan. Rasio pembobotan hibrida bernilai 20% representasi visual, 80% representasi tekstual.
3. Aplikasi web interaktif berbasis framework *Streamlit* menjadi antarmuka sistem. Fungsionalitas sistem menyajikan rekomendasi produk alternatif. Layanan transaksi e-commerce, fitur pembayaran ditiadakan.
4. Tampilan akhir antarmuka menyajikan 5 (Top-5) rekomendasi produk skincare. Urutan rekomendasi tersusun berdasarkan skor kemiripan tertinggi produk acuan pilihan pengguna.
5. Tiga instrumen pengujian mengevaluasi kinerja sistem. Metrik *Precision@5* mengukur akurasi kategori. Komputasi *Silhouette Score* memvalidasi batas klaster data. Algoritma *t-SNE* memetakan sebaran visual data.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitian ini adalah:

1. Mengimplementasikan metode *Hybrid Content-Based Filtering* pada sistem rekomendasi skincare dengan menggabungkan dua ekstraksi fitur

sekaligus, yaitu fitur visual kemasan menggunakan *ResNet50* dan fitur teks komposisi bahan menggunakan *TF-IDF*, yang diimplementasikan ke dalam antarmuka aplikasi web interaktif berbasis *Streamlit*.

2. Menghasilkan daftar 5 (Top-5) rekomendasi produk skincare yang relevan dengan menghitung skor kemiripan (similarity score) antar produk berdasarkan integrasi data citra dan teks menggunakan rasio pembobotan hibrida (20% citra dan 80% teks) melalui algoritma Cosine Similarity, serta mengevaluasi performa dan tingkat akurasi menggunakan metrik *Precision@5*, *Silhouette Score*, dan visualisasi t-SNE.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi positif baik dari segi pengembangan ilmu pengetahuan maupun penerapan teknologi di masyarakat. Adapun manfaat yang ingin dicapai dijabarkan sebagai berikut:

1. Manfaat Teoritis
 - a. Menambah literatur akademik di bidang kecerdasan buatan, khususnya mengenai efektivitas integrasi fitur multimodal berbasis deep learning (*ResNet-50*) dan pembobotan teks (*TF-IDF*) dalam mengenali karakteristik produk kosmetik.
 - b. Memberikan pemahaman ilmiah yang lebih mendalam mengenai pengaruh penggabungan representasi visual kemasan dan data tekstual bahan aktif terhadap peningkatan relevansi sistem rekomendasi berbasis content-based filtering.
2. Manfaat Praktis
 - a. Menjadi panduan atau referensi bagi pengembang platform e-commerce kosmetik dalam membangun sistem pencarian alternatif produk skincare cerdas yang tidak bergantung pada riwayat interaksi atau data pribadi pengguna.
 - b. Membantu konsumen atau masyarakat luas dalam mengatasi fenomena kelebihan informasi (*information overload*) dengan menyediakan prototype antarmuka web yang mampu menyajikan

referensi produk substitusi secara cepat dan akurat berdasarkan kemiripan fisik maupun kandungan bahan.

- c. Menjadi acuan bagi peneliti selanjutnya untuk mengembangkan sistem rekomendasi lanjutan, khususnya dalam penentuan rasio bobot hibrida optimal serta pemanfaatan kamus stopword klinis pada domain dermatologi.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran menyeluruh mengenai alur pembahasan yang terdapat dalam setiap bab. Uraian singkat mengenai isi dari masing-masing bab adalah sebagai berikut:

BAB I PENDAHULUAN, berisi latar belakang masalah mengenai fenomena *information overload* pada produk *skincare*, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan laporan. Bab ini menjadi landasan awal mengapa penelitian sistem rekomendasi *hybrid* ini perlu dilakukan.

BAB II TINJAUAN PUSTAKA, berisi tinjauan pustaka dari penelitian-penelitian terdahulu yang relevan, serta landasan teori yang menjadi acuan dalam pengembangan sistem. Teori-teori yang dibahas meliputi konsep Sistem Rekomendasi, *Text Mining*, metode *Term Frequency-Inverse Document Frequency* (TF-IDF), *Deep Learning* dengan *Convolutional Neural Network* (CNN), arsitektur ResNet-50, perhitungan kemiripan *Cosine Similarity*, serta metode evaluasi sistem.

BAB III METODE PENELITIAN, di dalamnya terdapat tinjauan umum tentang objek penelitian, analisis masalah, serta tahapan perancangan sistem yang diusulkan. Bab ini menjabarkan alur kerja penelitian secara sistematis, mulai dari pengumpulan *dataset* sekunder, skenario pengujian, hingga desain arsitektur model *Hybrid Multimodal* yang akan dibangun.

BAB IV HASIL DAN PEMBAHASAN, bab ini merupakan inti dari penelitian yang berisi laporan mengenai proses implementasi metode *Hybrid Multimodal* ke dalam sistem. Di dalamnya dijelaskan secara rinci hasil dari setiap tahapan mulai

dari *preprocessing* data citra dan teks, ekstraksi fitur visual menggunakan arsitektur ResNet-50, pembobotan nilai TF-IDF, hingga perhitungan *Cosine Similarity* untuk menghasilkan rekomendasi. Bab ini juga menyajikan hasil evaluasi performa sistem berdasarkan nilai *Precision@K*, *Silhouette Score*, dan visualisasi *t-SNE*.

BAB V PENUTUP, berisi kesimpulan dari hasil penelitian yang telah dilakukan, serta saran-saran yang konstruktif bagi peneliti selanjutnya untuk pengembangan atau perbaikan sistem rekomendasi di masa mendatang.

