

ABSTRAK

Pemilihan produk *skincare* yang sesuai dengan kebutuhan pengguna sering menjadi permasalahan akibat banyaknya variasi produk serta perbedaan penulisan kandungan bahan. Kondisi ini menyebabkan pengguna kesulitan dalam menemukan produk *skincare* yang memiliki kandungan sesuai dan relevan.

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi *skincare* berbasis kandungan (*ingredients*) dengan menerapkan metode *content-based filtering*. Sistem dibangun menggunakan pemodelan *Term Frequency–Inverse Document Frequency* (TF-IDF) dan algoritma *K-Nearest Neighbor* (KNN), disertai proses *preprocessing* meliputi *case folding*, pembersihan karakter, tokenisasi, dan normalisasi sinonim kandungan.

Data *skincare* direpresentasikan dalam bentuk vektor TF-IDF dan digunakan untuk menghitung tingkat kemiripan antara *query* pengguna dan produk *skincare* menggunakan metode *cosine similarity* dan *euclidean distance*. Evaluasi sistem dilakukan menggunakan metrik *precision*, *Mean Reciprocal Rank* (MRR), dan *Mean Average Precision* (MAP) pada skenario Top-5, Top-7, dan Top-10 rekomendasi.

Hasil penelitian menunjukkan bahwa metode *cosine similarity* memiliki performa yang lebih baik dibandingkan *euclidean distance* dengan nilai *precision* tertinggi sebesar 0,95, nilai MRR sebesar 0,95, dan nilai MAP sebesar 0,88. Penelitian ini menghasilkan sistem rekomendasi *skincare* berbasis kandungan yang mampu memberikan rekomendasi produk secara relevan.

ABSTRACT

Selecting skincare products that suit users' needs is often problematic due to the wide variety of products available and differences in the labelling of ingredients. This makes it difficult for users to find skincare products with suitable and relevant ingredients.

This study aims to develop an ingredient-based skincare recommendation system by applying the content-based filtering method. The system is built using Term Frequency–Inverse Document Frequency (TF-IDF) modelling and the K-Nearest Neighbor (KNN) algorithm, accompanied by a preprocessing process that includes case folding, character cleaning, tokenization, and synonym normalization of ingredients.

Skincare data is represented in the form of TF-IDF vectors and used to calculate the similarity between user queries and skincare products using the cosine similarity and Euclidean distance methods. The system is evaluated using precision, Mean Reciprocal Rank (MRR), and Mean Average Precision (MAP) metrics in Top-5, Top-7, and Top-10 recommendation scenarios.

The results show that the cosine similarity method performs better than Euclidean distance with the highest precision value of 0.95, an MRR value of 0.95, and an MAP value of 0.88. This research has produced an ingredient-based skincare recommendation system that is capable of providing relevant product recommendations.

