

BAB I PENDAHULUAN

1.1. Latar Belakang

Penampilan fisik dapat mempengaruhi kepercayaan diri seseorang, sehingga banyak orang yang menggunakan kosmetik untuk menunjang penampilan [1]. Di Amerika Serikat menunjukkan bahwa 85% perempuan menganggap penggunaan kosmetik sebagai kebutuhan [2]. Meskipun minatnya tinggi, sebagian besar konsumen masih kesulitan untuk memahami komposisi beserta fungsi dari bahan kosmetik. Sekitar 30% pengguna di Indonesia tidak tahu apa yang terkandung dalam produk yang mereka gunakan [3]. Akibatnya, ada 51% konsumen mengalami efek samping seperti jerawat, iritasi, atau kemerahan [4].

Setiap produk kosmetik mengandung kombinasi beberapa bahan yang memiliki banyak manfaat sekaligus, seperti mencerahkan, melembapkan, dan *anti-aging*, tetapi juga memiliki banyak efek samping, seperti alergi dan iritasi [5]. Artinya, satu produk kosmetik tidak hanya memiliki satu kategori efek saja, tetapi beberapa kategori efek sekaligus. Situasi ini membuat proses kategorisasi tidak bisa dilakukan dengan klasifikasi tunggal, sehingga diperlukan pendekatan klasifikasi multi-label.

Untuk menangani klasifikasi multi-label tersebut dapat menggunakan pendekatan metode *problem transformation*, seperti *Binary Relevance*, *Classifier Chains*, dan *Label Powerset* [6]. Ketiganya memiliki keunggulan yang berbeda dalam mengubah multi-label menjadi bentuk yang dapat diprediksi oleh model klasifikasi. *Label Powerset* memiliki keunggulan karena dapat mempertimbangkan hubungan antar label atau kelas. *Binary Relevance* memiliki keunggulan konseptual sederhana, serta proses komputasi yang lebih cepat dan efisien. Sedangkan, *Classifier Chain* memiliki keunggulan yang dapat meningkatkan kinerja prediktif [6].

Kinerja metode *Problem Transformation* telah diuji pada berbagai penelitian. Seperti penelitian membandingkan *Binary Relevance*, *Classifier Chains*,

dan *Label Powerset* dengan *K-Nearest Neighbor* (KNN) sebagai *base classifier*. Hasilnya menunjukkan bahwa metode *Problem Transformation Label Powerset* dengan *K-Nearest Neighbor* menghasilkan nilai akurasi, *precision*, *recall*, dan *F1-score* terbaik, yaitu 86%, 92%, 86%, dan 87% [6]. Sebuah penelitian juga membandingkan *Binary Relevance*, *Classifier Chains*, dan *Label Powerset* dengan *Random Forest* sebagai *base classifier*. Hasilnya menunjukkan bahwa *Binary Relevance* paling stabil, dengan *F1-Score* 0,76 pada skenario 70:30 dan 0,74 pada skenario 60:40 [7]. *Classifier Chain* digunakan pada penelitian [8] dengan KNN dan SVM sebagai *base classifier*. Pada model SVM, *F1-Score* tertinggi diperoleh pada skenario 90:10 sebesar 98% dengan *Hamming Loss* 0,017. Pada model KNN, performa terbaik dicapai pada skenario 90:10 dan 80:20, keduanya menghasilkan *F1-Score* 98% dengan *Hamming Loss* 0,017. Beberapa temuan ini menunjukkan bahwa kinerja data multi-label sangat dipengaruhi oleh kombinasi metode *problem transformation* dan algoritma *classifier*.

Kinerja metode *problem transformation* terbukti sangat bergantung pada jenis *base classifier*nya yang digunakan, sebagaimana ditunjukkan oleh penelitian sebelumnya yang menghasilkan performa berbeda ketika KNN, *Random Forest*, dan SVM dijadikan *base classifier*. Kondisi ini memperlihatkan pentingnya pemilihan model dasar yang mampu mengelola karakteristik data multi-label secara efektif. XGBoost menjadi kandidat yang relevan karena mampu menangani fitur yang kompleks dan tetap stabil pada kondisi ketidakseimbangan data [9]. Namun, penelitian terdahulu belum ada yang secara bersamaan menguji tiga metode *problem transformation* seperti *Binary Relevance*, *Classifier Chains*, dan *Label Powerset* dengan menggunakan XGBoost sebagai *base classifier*. Berdasarkan kekosongan tersebut, penelitian ini mengimplementasikan XGBoost untuk membandingkan ketiga metode tersebut pada data manfaat dan efek samping bahan kosmetik. Dengan demikian, tujuan penelitian ini adalah menemukan kombinasi metode yang memiliki kinerja terbaik yang didasarkan pada metrik evaluasi multi-label untuk menghasilkan klasifikasi label yang lebih akurat.

1.2. Rumusan Masalah

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

1. Bagaimana penerapan metode *problem transformation* (*Binary Relevance*, *Classifier Chain*, dan *Label Powerset*) dalam klasifikasi multi-label manfaat dan efek samping produk kosmetik?
2. Metode *problem transformation* mana yang menghasilkan performa terbaik berdasarkan metrik evaluasi multi-label?

1.3. Batasan Masalah

Agar pembahasan penulisan tepat pada sasaran dan memperjelas ruang lingkup, maka penulis membatasi masalah yang akan dibahas yaitu sebagai berikut :

1. Penelitian ini berfokus pada pembangunan dan evaluasi model klasifikasi multi-label untuk bahan kosmetik.
2. Dataset yang digunakan berisi data teks berupa komposisi bahan kosmetik beserta informasi manfaat dan potensi efek samping yang ditimbulkannya.
3. Metode klasifikasi yang dibandingkan mencakup tiga pendekatan *problem transformation*, yaitu *Binary Relevance*, *Classifier Chain*, dan *Label Powerset*, dengan XGBoost sebagai base classifier.
4. Analisis dilakukan menggunakan data sekunder, sehingga kebenaran manfaat dan efek samping bahan kosmetik sepenuhnya bergantung pada sumber dataset yang digunakan.

1.4 Tujuan Penelitian

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

1. Menerapkan metode *problem transformation* (*Binary Relevance*, *Classifier Chain*, dan *Label Powerset*) pada model klasifikasi multi-label untuk menganalisis manfaat dan efek samping produk kosmetik.

2. Membandingkan performa ketiga metode *problem transformation* tersebut berdasarkan metrik evaluasi multi-label untuk menentukan metode dengan hasil paling optimal.

1.5 Manfaat Penelitian

Secara umum, manfaat penelitian ini terbagi menjadi manfaat praktis dan manfaat teoritis sebagai berikut:

Manfaat praktis

1. Hasil evaluasi performa setiap metode dapat membantu peneliti atau praktisi memilih pendekatan klasifikasi multi-label yang paling efisien untuk analisis produk kosmetik berdasarkan komposisi bahan.
2. Memberikan contoh implementasi model multi-label *classification* berbasis teks yang dapat dijadikan acuan dalam penelitian lanjutan di bidang kecantikan dan kesehatan kulit.

Manfaat Teoritis

1. Memberikan kontribusi terhadap pengembangan pengetahuan di bidang *machine learning*, khususnya pada penerapan klasifikasi multi-label berbasis teks.
2. Menambah referensi akademik mengenai perbandingan tiga metode *problem transformation* (*Binary Relevance*, *Classifier Chain*, dan *Label Powerset*) dengan XGBoost sebagai *base classifier*.
3. Menghasilkan temuan empiris yang dapat dijadikan rujukan teoretis oleh penelitian lanjutan di bidang NLP, analisis kosmetik, dan kesehatan kulit

1.6 Sistematika Penulisan

Guna memudahkan pembaca saat mengikuti apa yang telah dipaparkan dalam laporan skripsi ini, Berikut adalah sistematika dalam penulisan penelitian tersusun dari 5 (lima) bab :

BAB I PENDAHULUAN

Pada bab ini berisikan gambaran umum penelitian yang

dilakukan, meliputi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

BAB II TINJAUAN PUSTAKA

Bab ini berisi landasan teori yang relevan dengan penelitian dan tinjauan peneliti terdahulu

BAB III METODE PENELITIAN

Bab ini menjelaskan alur penelitian serta metodologi yang digunakan dalam pembangunan model klasifikasi multi label.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi model yang telah dibangun, mulai dari proses pelatihan, evaluasi, hingga perbandingan performa antar metode problem transformation. Pembahasan juga mencakup interpretasi hasil, penyebab perbedaan performa, serta kesesuaian hasil dengan tujuan penelitian.

BAB V PENUTUP

Bab ini berisi kesimpulan penelitian yang menjawab rumusan masalah serta saran untuk penelitian selanjutnya.