

1. Pendahuluan

Kosmetik atau produk kecantikan adalah produk yang digunakan untuk menjaga kebersihan tubuh dan meningkatkan daya tarik [1]. Penggunaan produk kosmetik saat ini telah menjadi bagian dari gaya hidup banyak orang. Hal ini berdampak dengan semakin banyak pula konsumen kosmetik yang ingin memberikan ulasan pada produk yang digunakan. Pemberian ulasan ini membantu konsumen lain yang sedang mempertimbangkan membeli suatu produk untuk dapat melihat opini langsung dari konsumen yang telah menggunakan produk tersebut. Saat ini telah banyak ulasan yang diberikan untuk beragam produk dari beragam *brand* dan tersebar di banyak platform baik lokal maupun non-lokal.

Situs Female Daily merupakan *platform* yang karena menjadi wadah untuk para konsumen produk kecantikan berbagi pengalaman mereka menggunakan produk kecantikan. Konsumen dapat memberikan ulasan berupa teks terhadap suatu produk dan juga dapat memberikan *rating* dengan skala 1 sampai 5 pada produk tersebut. Namun penelitian ini tidak menggunakan *rating* karena akan dilakukan analisis sentimen dari ulasan teks yang diberikan.

Sentimen adalah pendapat atau pandangan yang didasarkan pada sesuatu perasaan. Analisis sentimen juga dikenal sebagai *opinion mining*, yaitu penelitian terkait sentimen seseorang terhadap suatu entitas [2]. Dalam beberapa tahun terakhir, terjadi peningkatan pesat jumlah forum diskusi ataupun situs untuk berbagi ulasan dimana ulasan yang berikan berdasarkan sentimen pemberi ulasan [3]. Ulasan produk kecantikan juga meningkat pesat seiring dengan meningkatnya angka penggunaan produk kecantikan. Saat ini banyak ulasan produk kecantikan yang tersedia pada situs lokal, namun penelitian terkait analisis sentimen masih dilakukan untuk menentukan sentimen suatu ulasan tergolong positif atau negatif [4]. Selain penelitian [4], terdapat penelitian yang juga menggunakan ulasan dari situs Female Daily untuk prediksi *rating* produk dari ulasan yang diberikan [5]. Dengan banyaknya ulasan yang tersedia dan penelitian yang masih terbatas, banyaknya ulasan masih berbanding terbalik dengan kejelasan identifikasi aspek yang menjadi pertimbangan dalam penggunaan suatu produk kecantikan.

Analisis sentimen level aspek dianggap lebih mengungkapkan esensi penuh ulasan dibandingkan dengan analisis sentimen biasa. Oleh karena itu, dibutuhkan analisis sentimen level aspek untuk analisis masukan pengguna yang lebih detail. Secara bisnis, keluaran dari analisis sentimen level aspek memungkinkan penyedia layanan atau produk untuk mengetahui pandangan konsumen terhadap aspek dari suatu produk [6]. Kemampuan untuk mengetahui aspek secara lebih terperinci inilah yang membuat analisis sentimen level aspek digunakan untuk analisis ulasan produk atau jasa. Analisis sentimen level aspek dapat diimplementasikan untuk kasus yang memiliki aspek lebih dari satu dan dikenal dengan *multi-aspect sentiment analysis*. Penelitian ini menggunakan *multi-aspect sentiment analysis* untuk analisis yang lebih detail terkait perbandingan antar aspek.

Ulasan produk pada situs Female Daily tidak hanya menggunakan bahasa Indonesia. Banyak ulasan yang ditulis dalam bahasa Inggris ataupun code-mixed (campuran bahasa Indonesia dan Bahasa Inggris), namun penggunaan satu bahasa pada keseluruhan dataset mampu mengurangi kompleksitas dan meningkatkan kualitas dari model yang dibangun [7]. Pemrosesan ribuan data yang setiap datanya terdiri dari satu atau lebih dari satu kalimat membutuhkan waktu komputasi yang tinggi saat proses klasifikasi. Ekstraksi fitur dapat menjadi solusi untuk permasalahan data teks yang terlalu banyak. Ekstraksi fitur yang paling populer digunakan dalam klasifikasi teks adalah TF-IDF dan n-gram karena menghasilkan akurasi yang baik pada model yang dibangun [8].

Model yang dibangun dalam penelitian ini bertujuan untuk melakukan *multi-aspect sentiment analysis* pada ulasan produk kecantikan. Ulasan pada dataset diterjemahkan seluruhnya ke dalam bahasa Indonesia. Metode ekstraksi fitur yang digunakan dalam penelitian ini adalah *Term Frequency-Inverse Document Frequency* (TF-IDF) dan n-gram. Metode klasifikasi yang digunakan adalah Random Forest. Proses klasifikasi Random Forest akan dilakukan dengan melakukan *hyperparameter tuning*. *Hyperparameter tuning* dapat membantu mencari hasil paling maksimal untuk mengisi parameter-parameter di model yang dibangun. Hasil dari nilai yang beragam untuk setiap parameter kemudian dibandingkan untuk mencari kombinasi yang memiliki performansi tinggi.

Batasan masalah pada penelitian ini adalah dari sepuluh kategori produk, hanya empat kategori produk yang ulasannya digunakan. Keempat kategori produk tersebut adalah *toner*, *serum*, *sun protection*, dan *exfoliator*. Terdapat empat aspek yang diberikan pada setiap ulasan, aspek-aspek tersebut adalah harga, pengemasan, produk, dan aroma. Aspek produk berkaitan dengan pendapat umum yang terungkap dalam ulasan seperti menyukai atau tidak merekomendasi. *Labeling* pada aspek dilakukan secara manual dimana terdapat ulasan positif, netral, dan negatif pada setiap aspek. Aspek ulasan akan diberikan nilai *default* 0 jika tidak disebutkan dalam ulasan.

Tujuan penelitian ini adalah untuk mengetahui pengaruh teknik *preprocessing*, n-gram dan *hyperparameter tuning* pada *Random Forest* dalam proses klasifikasi sentimen terhadap aspek pada ulasan produk kecantikan. Bagian selanjutnya dari penelitian ini adalah bagian 2 yang membahas studi terkait penelitian yang telah dilakukan sebelumnya. Pada bagian 3 dibahas teori yang berkaitan dengan teori terkait perancangan sistem yang dibangun. Selanjutnya pada bagian 4 akan dijelaskan evaluasi penelitian terkait hasil pengujian dan analisis hasil pengujian. Pada bagian 5 kemudian akan dibahas kesimpulan dan saran untuk penelitian yang akan datang.