

## 1. Pendahuluan

### Latar Belakang

Tahun 2021, OJK mencatat 88,1 persen pengguna internet di Indonesia telah menggunakan layanan *e-commerce* dalam membeli berbagai produk. Angka tersebut didapatkan dari survei We Are Social. Survei tersebut meunjukkan bahwa negara Indonesia berada di peringkat pertama dari segi layanan *e-commerce* [1]. Website Female Daily adalah Salah satu website yang menyediakan layanan informasi seputar produk kecantikan adalah. Pada website <https://reviews.femaledaily.com/>, lebih dari 30.000 produk kecantikan dan tercatat terdapat 300.000 *reviews*. Fitur *review* merupakan salah satu fitur yang sangat membantu pengguna dalam menentukan produk kecantikan yang akan dibeli untuk digunakan dalam *daily beauty care* pengguna[2]. Karena terlalu banyak ulasan akan membutuhkan waktu lama untuk membacanya, hampir tidak mungkin pengguna bisa membaca keseluruhan informasi yang ada[3]. Sehingga diperlukan proses analisis sentiment.

Analisis sentimen atau dikenal dengan *opinion mining* merupakan bagian ilmu yang menganalisis pendapat orang, sentimen, penilaian, sikap, dan emosi[4]. Klasifikasi terhadap sentimen memiliki tujuan untuk menemukan solusi dari masalah dengan cara mengelompokkan pendapat atau ulasan pengguna menjadi opini positif, netral, atau negatif.

Sebelumnya telah dilakukan penelitian terdahulu yang berkaitan dengan topik serupa. Analisis sentimen pada ulasan produk konsumen bahasa Indonesia menggunakan metode *Naïve Bayes* dan TF-IDF. Hasil dari penelitian tersebut mampu menghasilkan nilai akurasi 82% [5]. Penelitian lainnya adalah analisis sentimen menggunakan *Random Forest Ensemble* untuk review mobile product di Kanada. Dari penelitian tersebut menghasilkan bahwa *Random Forest* merupakan classifier yang bagus untuk mengklasifikasikan multiclass dengan menghasilkan nilai accuracy 72% [6]. Dalam penelitian [7] metode *Chi-square* berfungsi untuk memberikan nilai pada kata atau fitur. Fitur yang terpilih selanjutnya digunakan untuk tahap klasifikasi. Hasil dari penelitian ini mampu meningkatkan performansi dari model yang dibangun. Pada penelitian sebelumnya terdapat pendekatan baru yaitu *Random Forest*. Metode tersebut tidak hanya meningkatkan akurasi tetapi juga kompleksitas waktu [8]. Klasifikasi *Random Forest* menghasilkan performa sistem yang sangat bagus untuk dataset ulasan *movie* berbahasa Inggris. RF dan *Mutual Information* sebagai seleksi fitur dikombinasikan sehingga menghasilkan nilai akurasi dan nilai *f1-scores* dengan nilai masing-masing sebesar 79% dan 75% [9].

Pada penelitian ini penulis menggunakan metode *Random Forest* dan TF-IDF sebagai Classification dan Ekstraksi fitur dalam menganalisis sentimen terhadap ulasan produk kecantikan pada website *Female Daily*. Sebelum melakukan klasifikasi dengan banyak fitur, menyeleksi fitur sangat diperlukan untuk meningkatkan kinerja sistem. TF-IDF dapat meningkatkan kinerja sistem dalam studi sebelumnya[7]. Klasifikasi *Random Forest* untuk mengklasifikasikan data yang imbalanced dan metode tersebut memiliki kinerja yang sangat baik pada penelitian sebelumnya [8] [9]. TF-IDF dipilih menggantikan Word2vec, karena Word2vec tidak terlalu cocok dengan penelitian yang dibuat. Oleh sebab itu TF-IDF digunakan agar bisa meningkatkan akurasi dan relevansi hasil pencarian[17].

### Topik dan Batasannya

Topik dan batasan masalah pada penelitian ini adalah pengujian untuk mengetahui performansi dari model berdasarkan skenario pengujian. Batasan masalah pada penelitian ini adalah dilakukan analisis sentimen menggunakan metode *Random Forest* dan TF-IDF sebagai seleksi fitur. Data yang digunakan adalah data review produk kecantikan berbahasa Indonesia yang berasal dari website *Female Daily*, dengan jumlah 3959 data dengan 4 level aspek, yaitu harga, kemasan, produk dan aroma. Pada setiap aspek memiliki kelas yaitu positif, netral dan negatif.

### Tujuan

Tujuan dari penelitian tugas akhir ini berdasarkan latar belakang dan perumusan masalah yaitu melakukan analisis sistem pada ulasan produk kecantikan dari website female daily menggunakan metode *Random Forest* dan TF-IDF sebagai seleksi fitur, kemudian mengukur hasil performansi dari sistem yang dibangun.

### Organisasi Tulisan

Pada bab 2 membahas studi terkait dengan penelitian ini. Pada bab 3 membahas teori dan rancangan sistem penelitian. Hasil dan analisis dibahas pada bab 4. Kemudian pada bab 5 membahas kesimpulan dari penelitian ini.