

1. Pendahuluan

Latar Belakang

Pada Tahun 2021, OJK mencatat 88,1 persen pengguna internet di Indonesia telah menggunakan layanan *e-commerce* dalam membeli sejumlah produk. Angka tersebut didapatkan dari hasil survei We Are Social. Survei tersebut meunjukkan bahwa Indonesia menduduki peringkat pertama dari segi layanan *e-commerce* [1]. Salah satu website yang menyediakan layanan informasi seputar wanita adalah Female Daily. Pada website <https://reviews.femaledaily.com/> tercatat terdapat 300.000 *reviews* dari lebih dari 30.000 produk kecantikan. 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]. Banyaknya ulasan akan membutuhkan waktu lama untuk membaca, hampir tidak mungkin pengguna akan membaca keseluruhan informasi yang ada[3]. Sehingga diperlukan proses analisis sentiment.

Analisis sentimen atau disebut juga *opinion mining*, adalah bidang studi menganalisis opini orang, sentimen, evaluasi, penilaian, sikap, dan emosi[4]. Klasifikasi sentimen bertujuan untuk mengatasi masalah ini dengan secara otomatis mengelompokkan opini atau pada *dataset* penelitian ini yaitu *review* pengguna menjadi opini positif atau negatif.

Sebelumnya telah dilakukan penelitian terdahulu yang berkaitan dengan topik serupa. Analisis sentimen pada *review* produk konsumen bahasa Indonesia dengan 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]. Pada penelitian Chi Square merupakan seleksi fitur yang digunakan untuk menghitung tingkat dependensi sebuah fitur terhadap suatu kelas. Dalam penelitian [7] metode Chi Square memberikan nilai pada fitur yang selanjutnya diurutkan dan diseleksi menurut persentase yang diujikan. Fitur yang terseleksi digunakan untuk proses klasifikasi. Hasil dari penelitian ini mampu meningkatkan performansi dari model yang dibangun. Pada penelitian sebelumnya terdapat pendekatan baru yaitu Modified Balanced Random Forest. Metode tersebut tidak hanya meningkatkan akurasi tetapi juga kompleksitas waktu [8]. Klasifikasi Modified Balanced Random Forest memiliki kinerja yang sangat baik untuk dataset 2 movie review berbahasa Inggris. Modified Balanced Random Forest dan Mutual Information sebagai seleksi fitur mampu menghasilkan nilai akurasi 79% dan nilai f1-scores 75% [9].

Pada penelitian ini penulis menggunakan metode Modified Balanced Random dan Chi Square sebagai seleksi 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. *Chi-square* dapat meningkatkan kinerja sistem dalam studi sebelumnya[7]. Klasifikasi MBRF untuk mengklasifikasikan data yang imbalanced dan metode tersebut memiliki kinerja yang sangat baik pada penelitian sebelumnya [8] [9].

Topik dan Batasannya

Topik dan batasan masalah dalam penelitian tugas akhir ini adalah pengujian untuk mengetahui performansi dari model berdasarkan skenario pengujian. Batasan masalah pada penelitian ini adalah dilakukan analisis sentimen menggunakan metode MBRF dan *Chi-square* 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 MBRF dan *Chi-square* 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