

1. PENDAHULUAN

1.1 Latar Belakang

Saat ini masyarakat dunia berada dalam era perkembangan teknologi yang semakin canggih. Perkembangan teknologi ini membawa beberapa keuntungan dan perubahan, diantaranya perubahan gaya hidup masyarakat tentang cara berjualan dan cara berbelanja [1]. Di Indonesia, data menunjukkan bahwa 96% pengguna internet telah mencari produk atau layanan untuk dibeli secara online [2]. Saat melakukan aktivitas belanja online, review produk sangat penting bagi penjual dan calon pembeli.

Saat calon pembeli berbelanja secara online, barang yang diinginkan tidak bisa disentuh dan dilihat langsung untuk meyakinkan mereka. Review produk bisa menjadi pertimbangan bagi calon pembeli karena memuat pengalaman pembeli lain saat menggunakan produk. Salah satu situs portal yang menyediakan review produk kecantikan di Indonesia adalah Female Daily. Situs ini merupakan wadah bagi perempuan Indonesia untuk berbagi konten tentang dunia perempuan, seperti kecantikan, parenting, fashion, dll. [3]. Di dunia kecantikan, Female Daily memiliki fitur review yang memungkinkan perempuan menulis review di forum. Fitur review produk kecantikan dikategorikan menjadi beberapa bagian, seperti moisturizer, treatment, mask, dan cleanser. Ragam produk kecantikan tersebut kemudian diuraikan kembali sesuai dengan mereknya sehingga pengguna dapat menulis dan mencari review dengan mudah sesuai jenis dan mereknya. Pada tahun 2014, situs web Female Daily memiliki 7,8 juta tampilan halaman bulanan, dan terus bertambah hingga saat ini [4].

Berdasarkan penelitian yang telah dilakukan, banyaknya review menunjukkan bahwa produk tersebut diminati, dan meningkatkan minat beli konsumen [5]. Namun, dalam satu produk, seringkali terdapat ribuan review, dan setiap review memiliki konten yang berbeda. Jumlah ulasan suatu produk mungkin menjadi masalah karena pembeli potensial dan penjual memiliki waktu terbatas untuk menganalisis semua ulasan. Berdasarkan masalah yang telah diuraikan maka diperlukan suatu sistem yang dapat mengolah dan meringkas data review. Data ulasan produk berisi sentimen atau emosi yang diungkapkan oleh calon pembeli. Sentimen pembeli terhadap suatu produk akan terbagi menjadi tiga jenis, yaitu positif, netral, dan negatif. Analisis sentimen akan digunakan untuk membangun sistem. Berdasarkan

levelnya, analisis sentimen dibagi menjadi tiga level, yaitu level dokumen, level kalimat, dan level aspek [6].

Penelitian ini membangun sistem yang menerapkan analisis sentimen pada tingkat aspek agar diperoleh hasil sentimen yang lebih spesifik berdasarkan aspek tersebut. Data yang digunakan adalah kumpulan review produk kecantikan dari situs Female Daily. Pengguna produk menulis review dalam bahasa Indonesia untuk menilai produk dari berbagai aspek berdasarkan pengalaman mereka. Dalam penelitian ini aspek-aspek yang dinilai telah ditentukan oleh peneliti terlebih dahulu yaitu ciri-ciri umum produk kecantikan seperti harga, kemasan, kualitas, dan aroma. Masing-masing aspek yang dipilih memiliki 3 sentimen, yakni positif, negatif, dan netral. Dalam penelitian ini aspek-aspek tersebut dinyatakan sebagai kelas, dan sentimen dianggap sebagai label.

Sistem dibangun dengan menggunakan metode ekstraksi ciri TF-IDF yang dikombinasikan dengan kata N-gram untuk mendapatkan hasil yang maksimal. Model N-gram yang digunakan adalah kata unigram dan kata bigram. Data review dibagi menjadi data train dan data test dengan perbandingan 80:20. Kemudian model dibentuk dari data latih dengan menggunakan k-fold Cross-Validation. Ketika TF-IDF dikombinasikan dengan algoritma SVM, penelitian [7] memperoleh akurasi sebesar 91,24%. Dengan demikian, proses klasifikasi sentimen dalam penelitian ini dilakukan dengan menggunakan algoritma SVM. Berikut adalah kontribusi utama kami dalam penelitian ini.

- a. Kami menggunakan data baru yang belum pernah dipelajari sebelumnya
- b. Kami membuat kamus baru yang dinormalisasi dengan menambahkan 426 kata
- c. Kami membuat kamus kata-kata tak berarti baru dengan menambahkan 283 kata
- d. Kami menganalisis pengaruh tahap preprocessing, perbandingan unigram dan bigram, dan hasil tuning hyperparameter SVM. tinggi.

1.2 Rumusan Masalah

Rumusan masalah yang akan diangkat pada penelitian ini adalah Bagaimana cara membangun sistem dan mengukur performansi akurasi dari algoritma SVM yang dibantu oleh TF-IDF untuk menentukan fitur dari ulasan produk kecantikan berbahasa Indonesia dari situs *female daily*?

Adapun batasan masalah pada tugas akhir ini adalah sebagai berikut.

1. *Dataset* yang digunakan hanya data yang berbahasa Indonesia dengan format .csv

2. *Dataset* berasal dari situs *Female Daily*
3. Kategori produk yang digunakan pada *dataset* hanya berupa produk *skin care*, yaitu kategori *facial wash, toner, wash-off, serum & essence, sun protection, lip balm & treatments*, dan *scrub & exfoliator*
4. Klasifikasi sentimen berada pada level aspek

1.3 Tujuan

Tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut.

1. Mampu menangani ekstraksi fitur yang terdapat pada ulasan produk dengan metode TF-IDF
2. Mampu menganalisis sentimen pembeli dari suatu ulasan produk dengan algoritma *Support Vector Machine*
3. Mengetahui performansi akurasi dari algoritma SVM yang dibantu oleh TF-IDF

1.4 Rencana Kegiatan

Rencana kegiatan pada pelaksanaan penelitian ini adalah sebagai berikut.

1. Studi Literatur

Penulis mempelajari sumber literatur yang berkaitan dengan topik yang diangkat. Sumber literatur tersebut berupa jurnal, artikel, dan karya ilmiah. Kegiatan ini bertujuan agar penulis dapat memahami permasalahan dan solusi dari topik yang diangkat.

2. Pengumpulan Data

Kegiatan pengumpulan data dilakukan untuk mengumpulkan data yang akan digunakan pada penelitian. Data yang digunakan pada penelitian ini adalah data yang diambil dari situs *Female Daily*. Data ini berisi 15000 ulasan produk kecantikan yang ditulis oleh pembeli.

3. Analisis dan Perancangan Sistem

Pada kegiatan ini, penulis melakukan analisis kebutuhan yang dibutuhkan dalam sistem dan merancang sistem dengan menentukan alur serta tahapan yang akan dilakukan dalam membangun sistem.

4. Implementasi Sistem

Penulis mengimplementasikan rancangan sistem yang telah dibuat dengan menggunakan *dataset* yang telah dikumpulkan. Sistem akan dibangun dengan bahasa pemrograman Python menggunakan TF-IDF dan *Support Vector Machine* (SVM).

5. Analisis Hasil Implementasi

Pada kegiatan ini, penulis melakukan pengujian terhadap sistem yang telah dibangun menggunakan *confusion matrix*. Pengujian ini bertujuan untuk mengetahui seberapa baik hasil implementasi yang telah dilakukan dan memastikan hasil implementasi sesuai dengan tujuan yang ingin dicapai.

6. Penulisan Laporan

Tahap penulisan laporan merupakan tahapan terakhir yang dilakukan oleh penulis. Laporan berisi dokumentasi dari seluruh tahapan yang dilakukan untuk membangun sistem serta hasil dan evaluasinya. Laporan ini dapat digunakan dan dikembangkan untuk penelitian selanjutnya.

1.5 Jadwal Kegiatan

Berikut adalah jadwal kegiatan yang dilakukan dalam penelitian ini.

No	Kegiatan	Bulan ke-																
		1	2	3	4	5	6											
1	Studi Literatur	█	█	█	█	█	█	█	█	█	█	█						
2	Pengumpulan Data	█	█	█														
3	Analisis dan Perancangan Sistem			█	█	█	█	█										
4	Implementasi Sistem					█	█	█	█									
5	Analisa Hasil Implementasi							█	█	█	█	█						
6	Penulisan Laporan			█	█	█	█	█	█	█	█	█	█	█	█	█	█	█

Tabel 1 Jadwal Kegiatan Pengerjaan Tugas Akhir