

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Di era saat ini, teknologi tumbuh dan berkembang begitu pesat. Internet telah mengantarkan perubahan terhadap perilaku masyarakat. Membeli produk melalui *online* semakin diminati dan sudah menjadi kebiasaan untuk membeli berbagai macam produk salah satunya yaitu produk kecantikan. Sebelum membeli produk kecantikan, dibutuhkan banyak pertimbangan saat akan membeli, karena setiap kulit memiliki kondisi yang berbeda. Komentar produk kecantikan bisa menolong pelanggan untuk mengenali mutu pada produk kecantikan tersebut cocok atau tidak cocok untuk digunakan [1]. Namun, komentar yang ada di situs *online* sangat beragam dan jumlahnya sangat banyak. Hal tersebut membuat komentar menjadi bias. Berdasarkan kondisi tersebut, dibutuhkan analisis sentimen berbasis aspek karena tiap kalimat dapat memuat beberapa sentimen. Analisis sentimen tingkat aspek dapat memberikan ringkasan komentar yang dilakukan secara keseluruhan dokumen [2].

Analisis sentimen merupakan sebuah proses untuk menganalisis atau mengidentifikasi sebuah opini seseorang yang menunjukkan sikap terhadap suatu topik atau produk tertentu masuk ke dalam kategori positif, negatif, atau netral [3]. Analisis sentimen dibagi menjadi 3 tingkat yaitu, tingkat dokumen, tingkat kalimat dan tingkat aspek [1]. Pada penelitian [2] yang dilakukan oleh Ekky Yulianti Prastika S., Said Al Faraby dan Mahendra Dwifabri P. menghasilkan akurasi 74,21% dengan nilai $k = 23$. Pada Penelitian [4] yang dilakukan oleh Sumandeep Kaur, Geeta Sikka, dan Lalit Kumar Awasthi, metode *K-Nearest Neighbor* dan *N-Gram* hasil eksperimen menunjukkan 7 persen perbaikan akurasi dari metode SVM pada analisis sentimen. Eksperimennya dilakukan pada data bahasa Inggris dan saran untuk penelitian selanjutnya pengujian dilakukan pada bahasa lain [4]. Lalu penelitian yang dilakukan oleh Retno Sari dengan metode *N-gram* dan *K-Nearest Neighbor*. Menggunakan bigram menghasilkan akurasi sebesar 77.01% dengan

nilai $K = 7$ menggunakan *10 fold cross validation* dari 100 data yang digunakan[5]. Namun, pada penelitian yang dilakukan oleh Retno Sari dan penelitian yang dilakukan oleh Sumandeep Kaur, Geeta Sikka, dan Lalit Kumar Awasthi tersebut hanya membahas analisis sentimen tingkat dokumen yang hanya mengelompokkan opini ke dalam kelas sentimen positif, Non, atau netral. Analisis sentimen pada level aspek menunjukkan performa yang lebih baik [6].

Berbagai metode telah dikembangkan dan diterapkan untuk analisis sentimen. Salah satu metode tersebut adalah *K-Nearest Neighbor* (KNN). KNN merupakan metode yang melakukan klasifikasi berdasarkan kedekatan jarak atau kemiripan ciri atau sifat dari suatu data dengan data yang lain. KNN melakukan klasifikasi suatu data dengan mengidentifikasi sejumlah k tetangga terdekat kemudian mengambil kelas terbanyak dalam menentukan jenis kelas dari data tersebut. KNN termasuk algoritma yang sederhana bila dibandingkan dengan algoritma lain karena KNN hanya mengandalkan memori dan tidak membangun model pada saat melakukan pembelajaran mesin.

Berdasarkan latar belakang yang telah diuraikan dan permasalahan yang terjadi dari beberapa penelitian sebelumnya, maka penelitian ini akan dilakukan analisis sentimen pada ulasan produk kecantikan berbasis aspek menggunakan algoritma *K-Nearest Neighbor* dengan fitur ekstraksi ngram.

1.2. Rumusan Masalah

Berdasarkan uraian yang disampaikan dalam latar belakang masalah, maka rumusan masalahnya adalah Bagaimana performansi dari algoritma *K-Nearest Neighbor* dengan fitur ekstraksi *ngram* dalam melakukan analisis sentimen berbasis aspek terhadap ulasan produk kecantikan.

1.3. Maksud dan Tujuan Penelitian

Maksud dari penelitian ini adalah untuk melakukan analisis sentimen pada ulasan produk kecantikan dengan menggunakan *K-Nearest Neighbor* dan *N-Gram*. Tujuan dari penelitian ini adalah untuk mengetahui performansi dari *K-Nearest Neighbor* dan *N-Gram* dalam melakukan analisis sentimen berbasis aspek terhadap produk kecantikan.

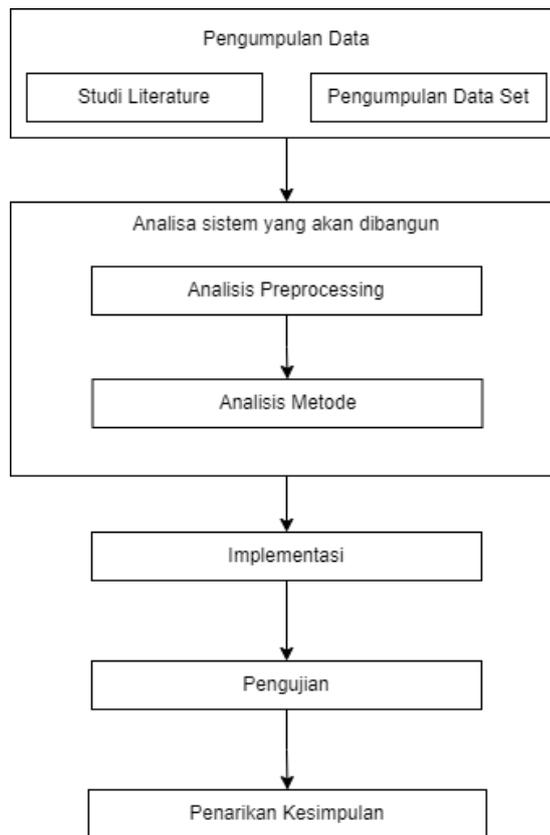
1.4. Batasan Masalah

Adapun Batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Ulasan yang digunakan hanya ulasan yang berbahasa Indonesia pada ulasan produk kecantikan.
2. Data yang digunakan berasal dari situs kaggle berikut <https://www.kaggle.com/datasets/hafidahmusthaanah/skincare-review?select=00.+Review.csv>.
3. Aspek yang akan digunakan adalah harga, efektivitas, kemasan dan aroma.
4. Menggunakan Pembobotan TF-IDF.
5. Polaritas yang telah diolah akan diklasifikasikan ke dalam kelas positif, negatif, dan non sentimen.
6. Pelabelan akan dilakukan secara manual.

1.5. Metodologi Penelitian

Metodologi penelitian yang akan dilakukan pada penelitian dapat dilihat pada Gambar 1.1.



Gambar 1.1. Gambaran Penelitian

1.5.1. Metode Pengumpulan Data

Metode pengumpulan yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Studi Literatur

Studi literatur merupakan teknik pengumpulan data dengan cara mengumpulkan jurnal maupun buku yang berkaitan dengan analisis sentimen berdasarkan aspek.

2. Pengumpulan Dataset

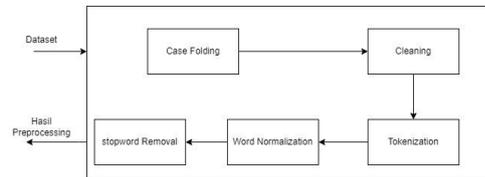
Dataset yang digunakan pada penelitian ini berupa ulasan produk kecantikan yang didapatkan dari situs Kaggle.com.

1.5.2. Metode Tahapan Analisis

Tahapan analisis pada penelitian ini adalah sebagai berikut:

1. Analisis *Preprocessing*, merupakan tahap untuk memproses data agar data menjadi lebih terstruktur dan lebih siap untuk diolah.

Langkah yang dilakukan pada tahap preprocessing terdiri dari *case folding*, *Cleaning*, *Tokenization*, *Word Normalization*, *Stopword Removal*, *Trigram*.



Gambar 1.2. Preprocessing

2. Analisis Metode, pada tahap ini akan dilakukan klasifikasi dengan menggunakan metode *K-Nearest Neighbor*. Meliputi, penggunaan ekstraksi fitur Trigram dan pembobotan TF-IDF.

1.5.3. Implementasi

Pada tahapan implementasi akan dilakukan pembuatan sistem berupa simulator, adapun hal yang bisa dilakukan oleh simulator yaitu memasukkan dan membaca data ulasan yang telah dilabeli secara manual, kemudian melakukan implementasi preprocessing yaitu: *case folding*, *cleaning*, *tokenization*, *word normalization*, *stopword removal*, dan *N-Gram* setelah data masukkan dipraproses, tahap selanjutnya yaitu pembobotan kata dengan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) menggunakan *library sklearn*. Lalu data dari pembobotan kata akan dilakukan pembagian data, yaitu data latih dan data uji. Lalu implementasi dengan menggunakan algoritma *K-Nearest Neighbor*.

1.5.4. Pengujian

Dalam tahapan ini akan dilakukan pengujian dan evaluasi performansi dari analisis sentimen berbasis aspek dengan menerapkan metode *K-Nearest Neighbor* pada ulasan produk kecantikan.

1.5.5. Penarikan Kesimpulan

Pada tahap ini akan dilakukan penarikan kesimpulan berdasarkan hasil dari tahap pengujian dalam menerapkan metode *K-Nearest Neighbor* pada analisis sentimen berbasis aspek.

1.6. Sistematika Penulisan

Sistematika penulisan yang digunakan pada penelitian ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Bab 1 menguraikan tentang latar belakang yang terdiri dari permasalahan yang berkaitan dengan penelitian, rumusan masalah, maksud dan tujuan penelitian, batasan masalah, metode penelitian, dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab 2 membahas dan menjelaskan mengenai landasan teori yang mendukung penulisan skripsi.

BAB 3 ANALISIS DAN PERANCANGAN

Bab 3 berisikan tentang analisis terhadap sistem yang akan dibangun meliputi analisis masalah, analisis perancangan, analisis data masukan, dan analisis pada pemrosesan serta klasifikasi.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisikan hasil implementasi dari analisis dan pengujian terhadap sistem yang sudah dibangun.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan mengenai hasil penelitian dan saran yang diharapkan akan berguna untuk penelitian lebih lanjut pada kasus yang sama.