

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Analisis sentimen, juga disebut penggalian opini, adalah bidang studi yang menganalisis pendapat, sentimen, penilaian, sikap, dan emosi orang terhadap entitas dan atributnya yang diungkapkan dalam teks tertulis dan mengelompokkan opini tersebut menjadi kategori positif dan negatif. Entitas dapat berupa produk, layanan, organisasi, individu, acara, masalah, atau topik[1]. Pada penelitian ini menggunakan analisis sentimen berbasis aspek. Dimana bukan hanya menentukan bagaimana sentimennya, tetapi juga pada aspek apa sentimen tersebut ditujukan. Dari segi kelengkapan analisis sentiment level aspek bisa dibilang lebih lengkap bila dibandingkan dengan level dokumen atau pun level kalimat[2]. Hal ini dikarenakan analisis sentimen level kalimat hanya difokuskan pada penentuan sentimen positif dan negatif saja, tanpa aspek yang terkandung di dalamnya[2]. Untuk analisis yang lengkap tersebut, maka perlu diketahui sentimen positif atau negatif pada masing-masing aspek sehingga yang perlu dilakukan adalah analisis sentimen berbasis aspek. Seiring dengan berkembangnya teknologi dalam melakukan pembelian sebuah produk pun bisa dilakukan secara *online* salah satunya yaitu melalui *marketplace* dimana pada *marketplace* tersebut terdapat sebuah ulasan mengenai sebuah produk, ulasan mengenai produk tersebut dapat membantu apakah produk tersebut layak untuk dibeli, salah satu jenis produk yang dapat terbantu dengan ulasan yaitu produk kecantikan. dimana ulasan dapat membantu pembeli dalam menentukan produk yang akan dibeli untuk mengetahui pendapat masyarakat yaitu positif dan negatif mengenai aspek-nya, seperti harga, packaging, efektifitas, dan aroma[3].

Salah satu metode yang sering digunakan yaitu metode *K-Nearest Neighbor* (K-NN), salah satu penelitian dalam penggunaan metode K-NN yaitu dalam klasifikasi berita[4] yang dibagi menjadi empat kategori berita menggunakan K-NN yang menghasilkan akurasi tertinggi 87%. Metode K-NN merupakan metode klasifikasi yang mudah digunakan dan banyak digunakan

(populer), namun memiliki kekurangan dalam menangani data yang tidak seimbang[5]. Pada penelitian Analisis Sentimen berbasis Aspek terhadap produk Kecantikan[3] menghasilkan akurasi rata-rata sebesar 74.21% namun rata-rata performansi *F-Measure* yang didapatkan sebesar 45.19% hal ini dikarenakan penggunaan polaritas yang tidak seimbang untuk masing-masing aspek.

Salah satu metode yang dapat menangani permasalahan data yang tidak seimbang yaitu Metode *Neighbor Weighted K-Nearest Neighbor* (NWKNN) merupakan pengembangan dari metode K-NN yang akan meningkatkan akurasi dari metode K-NN[6]. Pada Penelitian untuk analisis sentimen pada level kalimat, dengan kasus review aplikasi mobile menggunakan metode NWKNN[1] untuk data seimbang dengan data positif 100 dan data negatif 100 menghasilkan akurasi tertinggi 90% untuk kedua metode dengan nilai $K=15$, sedangkan untuk jumlah data tidak seimbang menggunakan metode K-NN mendapatkan akurasi 48% sedangkan untuk metode NWKNN mendapatkan akurasi 80% untuk $K=45$. Namun belum diketahui hasil metode NWKNN untuk kasus analisis sentimen berbasis aspek. Pada penelitian lain, mengklasifikasi penyimpangan tubuh menggunakan NWKNN[6] menghasilkan nilai akurasi tertinggi sebesar 95% sedangkan untuk menggunakan metode K-NN menghasilkan akurasi tertinggi sebesar 90%..

Dari beberapa penelitian sebelumnya yang telah diuraikan, diketahui metode KNN memiliki kekurangan yaitu dalam penggunaan data yang tidak seimbang[5], dimana pada penelitian menggunakan K-NN dan SMOTE untuk menangani data yang tidak seimbang dapat meningkatkan akurasi dari metode K-NN[5]. Dari penelitian menggunakan metode NWKNN yang merupakan pengembangan metode K-NN dapat menangani atau meningkatkan akurasi untuk data yang tidak seimbang[1]. Berdasarkan penelitian sebelumnya yang telah diuraikan dan permasalahan yang terjadi pada penelitian sebelumnya, maka pada penelitian ini akan dilakukan analisis sentimen berbasis aspek terhadap produk kecantikan menggunakan *Neighbor Weighted K-Nearest Neighbor* (NWKNN).

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan maka dirumuskan sebuah permasalahan yaitu Bagaimana performansi dari algoritma Neighbor Weighted K-Nearest Neighbor dalam melakukan analisis sentimen berbasis aspek terhadap ulasan produk kecantikan dalam menggunakan data yang tidak seimbang.

1.3. Maksud dan Tujuan Penelitian

Maksud dari penelitian ini adalah untuk melakukan analisis sentimen pada ulasan kosmetik di “Female Daily” dengan menggunakan NWKNN. Adapun tujuan pada penelitian ini adalah mengukur performansi dari metode NWKNN pada kasus analisis sentimen berbasis aspek terhadap ulasan produk kecantikan berdasarkan aspek-aspek yang terkait dengan menggunakan data yang tidak seimbang.

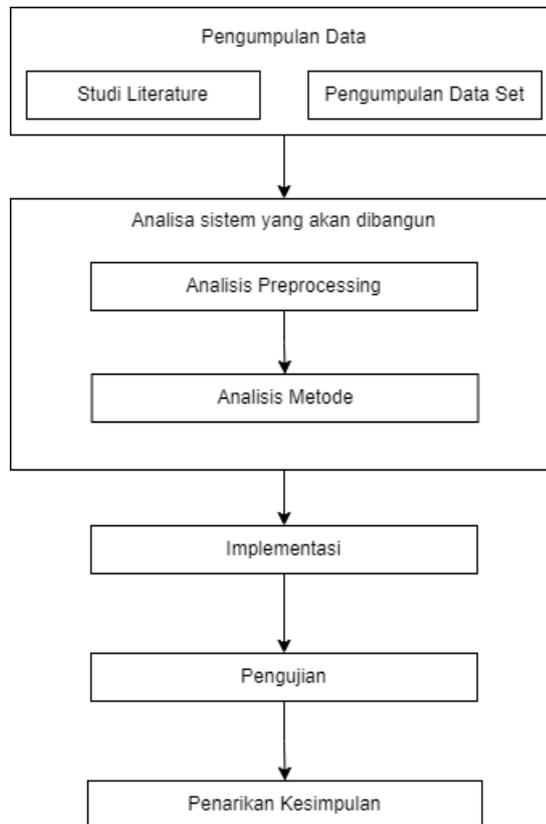
1.4. Batasan Masalah

Adapun Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Sumber dataset diambil dari situs Kaggle.com dengan url <https://www.kaggle.com/datasets/hafidahmusthaanah/skincare-review?select=00.+Review.csv>.
2. Aspek yang akan digunakan adalah harga, kemasan, efektifitas, aroma.
3. Menghitung jarak tetangga menggunakan Euclidean Distance.
4. Menggunakan ekstraksi fitur TF-IDF.
5. Data yang telah diolah akan diklasifikasikan ke dalam kelas positif, negatif, non sentimen.
6. Pelabelan sentimen dilakukan secara manual.
7. *Stemming* menggunakan *library* Sastrawi.
8. Jenis multilabel yang digunakan *Binary Classification* dengan menggunakan *One-Versus-Rest*.

1.5. Metodologi Penelitian

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



Gambar 1. 1. Metodologi Penelitian

1.5.1. Metode Pengumpulan Data

Berikut ini merupakan metode pengumpulan data yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Studi Literatur

Studi literatur merupakan proses pengumpulan data atau informasi yang berkaitan dengan penelitian yang sedang dijalankan. Pengumpulan data dilakukan dengan cara membaca buku ataupun hasil penelitian yang sudah dilakukan sebelumnya seperti jurnal, paper.

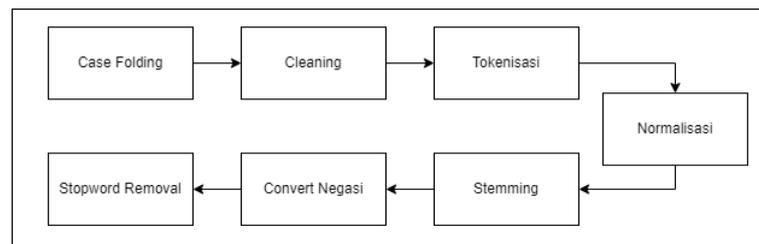
2. Pengumpulan Dataset

Dataset yang digunakan pada penelitian ini berupa ulasan produk kecantikan pada situs Female Daily 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*, *spelling Normalization*, *Stopword Removal*, *Stemming*.



Gambar 1. 2. Tahapan Preprocessing

2. Analisis Metode, pada tahap ini akan dilakukan klasifikasi dengan menggunakan metode NWKNN.

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*, *normalization*, *stemming*, *convert negation* dan *stopword removal*, 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 algoritma dengan menggunakan NWKNN yang akan menghasilkan kelas.

1.5.4. Pengujian

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

1.5.5. Penarikan Kesimpulan

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

1.6. Sistematika Penulisan

Sistematika penulisan yang 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

Bab 3 berisikan tentang analisis terhadap sistem yang akan dibangun meliputi analisis masalah, 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.