

BAB I PENDAHULUAN

1.1. Latar Belakang

Di zaman sekarang produk kecantikan sudah menjadi kebutuhan bagi manusia terutama oleh kaum wanita, namun tidak semua produk kecantikan memiliki kualitas yang baik atau cocok untuk semua orang. Pastiya sebelum membeli produk, konsumen akan mencari tahu mengenai kualitas produk kecantikan yang sesuai dengan keinginan konsumen. Untuk mencari tahu mengenai hal tersebut, konsumen dapat membaca ulasan produk agar mendapatkan gambaran tentang produk yang akan dibeli. Terdapat *website Female Daily* yang berisi konten ulasan seputar dunia kecantikan seperti *skincare, makeup, nails*, dll. Ulasan produk pun bermacam-macam, ada yang mengandung kalimat positif dan kalimat negatif.

Ulasan dinilai dapat membantu konsumen dalam memberikan gambaran mengenai produk tersebut [1], yang mana konsumen dapat mengetahui umpan balik mengenai suatu produk dari pengguna produk yang telah memberikan ulasan pada laman yang disediakan [2]. Namun *website Female Daily* memiliki ratusan bahkan ribuan ulasan untuk setiap produk, akibatnya membutuhkan waktu yang lebih lama untuk membaca ulasan satu persatu secara manual.

Oleh karena itu, diperlukan pemrosesan teks berupa klasifikasi sentimen secara otomatis yang dapat mengkategorikan sebuah ulasan produk ke dalam kelas positif atau negatif [3][4]. Klasifikasi sentimen adalah proses *text mining* untuk mengetahui pendapat seseorang atau opini baik berupa isu, atau pandangan terhadap suatu objek yang dilakukan dengan mengumpulkan data [5]. Dalam klasifikasi teks, *preprocessing* memiliki peran penting. *Preprocessing* bertujuan untuk mengolah data mentah menjadi data yang siap digunakan untuk klasifikasi, karena data yang akan digunakan mengandung struktur kata yang kurang baik. Salah satunya seperti mengandung bahasa *slang*. Bahasa *slang* merupakan kata singkatan yang tidak sesuai dengan KBBI [6]. Tahap *preprocessing* yang digunakan pada Tugas Akhir ini adalah data *cleaning, case folding, normalisasi, stopwords removal, dan stemming*. Tugas Akhir ini menggunakan ulasan berbahasa Indonesia yang banyak mengandung kata tidak baku, sehingga dibutuhkan normalisasi untuk mengatasi permasalahan tersebut [7]. Selain itu untuk mengetahui pengaruh *preprocessing* terhadap klasifikasi, Tugas Akhir ini akan melakukan

percobaan *preprocessing* tanpa *stemming*, *preprocessing* tanpa *stopwords*, serta *preprocessing* menggunakan *stemming* dan *stopwords*.

Setelah *preprocessing*, dilakukan pembobotan kata. Pembobotan kata yaitu mengubah data hasil *preprocessing* menjadi numerik agar dapat di proses di *classifier*. Terdapat penelitian sebelumnya mengenai dampak ekstraksi fitur terhadap analisis sentimen [8]. Dilakukan pengujian ekstraksi fitur menggunakan N-Gram dan TF-IDF, hasil yang didapatkan dari penelitian ini menunjukkan TF-IDF memberikan hasil akurasi 3-4% lebih baik dibandingkan dengan N-Gram [8]. Selain itu, TF-IDF termasuk algoritma yang efisien dan mudah dipahami [9].

Terdapat penelitian mengenai perbandingan metode klasifikasi menggunakan *Multinomial Naïve Bayes* dengan *K-Nearest Neighbor*. Dari penelitian ini metode *Multinomial Naïve Bayes* memiliki performa yang lebih baik dibandingkan dengan *K-Nearest Neighbor*, hal ini dibuktikan dengan *Multinomial Naïve Bayes* menghasilkan akurasi 93.33%, sedangkan *K-Nearest Neighbor* hanya menghasilkan akurasi sebesar 81.67% [10]. *Multinomial Naïve Bayes* merupakan metode perkembangan dari *Naïve Bayes Classifier*, yang mana metode ini memiliki waktu komputasi yang rendah, mudah diimplementasikan dan sejauh ini teorema *bayes* cukup baik digunakan untuk masalah klasifikasi teks [11][12]. Berdasarkan beberapa penelitian tersebut, Tugas Akhir ini akan menggunakan metode klasifikasi *Multinomial Naïve Bayes* dan ekstraksi fitur TF-IDF.

1.2. Rumusan Masalah

Rumusan masalah dari Tugas Akhir ini adalah sebagai berikut:

1. Bagaimana pengaruh tahap *preprocessing* jika dilakukan *preprocessing* tanpa *stemming*, tanpa *stopwords*, dan *preprocessing* menggunakan *stemming* dan *stopwords* terhadap hasil F1-Score?
2. Bagaimana pengaruh *balancing* data terhadap tahap *preprocessing* jika dilakukan *preprocessing* tanpa *stemming*, tanpa *stopwords*, dan *preprocessing* menggunakan *stemming* dan *stopwords* terhadap hasil F1-Score?

1.3. Batasan Masalah

Batasan masalah pada Tugas Akhir ini adalah:

1. Dataset yang digunakan berupa ulasan produk kecantikan yang diambil dari *website Female Daily* Menggunakan metode *Multinomial Naïve Bayes* dengan seleksi fitur *Information Gain*.
2. Mengklasifikasikan ulasan ke dalam kelas positif dan negatif.
3. Menggunakan metode *Multinomial Naïve Bayes* dan ekstraksi fitur TF-IDF.

1.4. Tujuan Penelitian

Tujuan dari Tugas Akhir ini sebagai berikut:

1. Mengetahui pengaruh *preprocessing* tanpa *stemming*, tanpa *stopwords*, dan *preprocessing* menggunakan *stemming* dan *stopwords* terhadap hasil F1-Score.
2. Mengetahui pengaruh *balancing* data dengan pengaruh *preprocessing* tanpa *stemming*, tanpa *stopwords*, dan *preprocessing* menggunakan *stemming* dan *stopwords* terhadap hasil F1-Score.

1.5. Metodologi Penelitian

1. Kajian Pustaka

Pada tahap kajian pustaka digunakan untuk melakukan pengumpulan jurnal atau buku yang terkait untuk mendapatkan informasi, memperdalam materi serta membandingkan atau mengembangkan jurnal.

2. Pengumpulan dan Analisis Data

Dataset yang digunakan dalam klasifikasi sentiment Tugas Akhir ini didapatkan dari *Female Daily* dengan ulasan berbahasa indonesia.

3. Analisis dan Perancangan Sistem

Untuk membantu memahami sistem yang akan dibuat, maka dalam tahap ini terdapat *flowchart* yang menggambarkan urutan kerja setiap proses pada sistem. Tugas Akhir ini melakukan klasifikasi menggunakan metode *Multinomial Naïve Bayes*, sedangkan proses seleksi fitur menggunakan *TF-IDF*.

4. Implementasi Model

Perancangan sistem yang telah dirancang sebelumnya, akan diimplementasikan sehingga tujuan penelitian dapat tercapai. Bahasa pemrograman yang digunakan dalam Tugas Akhir ini yaitu python.

