

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media memiliki peran penting dalam menyebarkan segala informasi. Seiring berjalannya waktu, media berkembang menjadi banyak jenis. Salah satu media yang sangat cepat berkembang adalah media online. Di dalam media *online*, terdapat banyak jenis media yang digunakan dalam menyebarkan informasi. Apalagi, dengan media online, semua orang dapat menggunakannya.

Female Daily adalah sebuah blog pribadi dengan konten fashion dan kecantikan yang dikelola oleh Hanifa pada tahun 2005. Female Daily Network merupakan sebuah situs media informasi bagi wanita yang menyajikan konten seputar dunia wanita di Indonesia. Dalam perjalanannya kini, Female Daily Network telah mengembangkan 4 situs, yaitu Fashionese Daily, Mommies Daily, Clozette Daily dan Travelers Daily.

Banyak dari konten promosi ini yang akhirnya membantu para pengguna yang ingin memakai maupun mengenal produk baru yang diluncurkan ke pasar. Dari berbagai ulasan-ulasan tersebut salah satu kategori yang sangat digemari saat ini terutama oleh kaum wanita adalah mengenai konten kecantikan. Fenomena ini ikut memunculkan pula hadirnya Female Daily yaitu para pembuat konten yang mengkhususkan dirinya untuk berbagi informasi terkait dunia kecantikan.

Sentiment Analysis atau *Opinion Mining* baru-baru ini menjadi topik menarik yang mencoba untuk menggabungkan statistik, kecerdasan buatan dan teknologi. Data Mining dalam kerangka terpadu (Pang, dan Lee, 2008) *opinion mining* adalah informasi tekstual yang berada di dalam *web* dan berisi tentang fakta dan *opini*. *Opini* merupakan pernyataan subjektif yang mencerminkan persepsi seseorang terhadap sesuatu peristiwa, misalnya tentang opini-opini yang berkembang seperti krisis di Libya dan Suriah, perdebatan tentang krisis ekonomi di Yunani, dan downdrating atas kredibilitas Amerika Serikat adalah beberapa topik kontroversial yang dimuat dalam berita sehari-hari. Menganalisa rating movie untuk mengetahui

tingkat pendapatan dari pemutaran suatu movie (Pang, Lee, dan Vaithyanathan, 2002).

Salah satu metode klasifikasi yang sering digunakan adalah *Naïve Bayes Classifier* yang pertama kali dikemukakan oleh Revered Thomas Bayes. Penggunaan *Naïve Bayes Classifier* sudah dikenalkan sejak tahun 1702-1761. Menurut Lewis, Hand dan Yu, *Naïve Bayes Classifier* merupakan pendekatan yang sangat sederhana dan sangat efektif untuk pelatihan klasifikasi (Lewis, 1998) (Hand and Yu, 2001). Sedangkan Kononenko dan Langley menyimpulkan bahwa *Naïve Bayes Classifier* merupakan kemungkinan label kelas data atau bias diasumsikan sebagai atribut kelas yang diberi label (Kononenko, 1990) (Langley, 1994).

Menurut Hamzah *Naïve Bayes* memiliki beberapa kelebihan, yaitu algoritma yang sederhana, lebih cepat dalam penghitungan dan berakurasi tinggi (Hamzah, 2012). Akan tetapi, pada metode *Naïve Bayes* juga memiliki kelemahan dimana sebuah probabilitas tidak bisa mengukur seberapa besar tingkat keakuratan sebuah prediksi. Maka dari itu, metode *Naïve Bayes* perlu dioptimasi dengan cara pemberian bobot menggunakan Gain Ratio. Pemberian bobot pada *Naïve Bayes* menimbulkan permasalahan pada penghitungan probabilitas setiap dokumen. Dimana fitur yang tidak merepresentasikan kelas yang diuji banyak muncul sehingga terjadi kesalahan klasifikasi. Oleh karena itu, pembobotan *Naïve Bayes* masih belum optimal.

Untuk data skala besar sangat dibutuhkan kecepatan dalam proses pencarian data. Sehingga dibutuhkan klasifikasi data terlebih dahulu. *Naive Bayes* merupakan algoritma pembelajaran untuk klasifikasi dengan efisiensi komputasi dan akurasi yang baik, khususnya untuk dimensi dan jumlah data yang besar. Untuk itu dalam penelitian ini akan membuktikan kemampuan *naïve bayes classifier* untuk mengklasifikasikan review produk kecantikan yang diperoleh dari *femaledaily*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka dapat diidentifikasi beberapa rumusan masalah sebagai berikut:

1. Berapa tingkat *Accuracy*, *Precision* dan *Recall Naïve Bayes* tanpa menggunakan pembobotan *Gain Ratio*?
2. Berapa tingkat *Accuracy*, *Precision* dan *Recall Naive Bayes* dengan melakukan optimasi dengan menggunakan pemboboan *Gain Ratio*?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menghitung tingkat *Accuracy*, *Precision* dan *Recall* klasifikasi sentiment menggunakan *Naïve Bayes Classifier*.
2. Menghitung tingkat *Accuracy*, *Precision* dan *Recall* klasifikasi *Naive Bayes Classifier* dengan melakukan optimasi dengan menggunakan pembobotan *Gain Ratio*.

1.4 Batasan Penelitian

Agar penelitian tidak menyimpang dari topik penelitian maka dibuat batasan penelitian seperti berikut:

1. Dataset yang digunakan adalah review produk *White Perfect Clinical Day Cream* yang diperoleh dari <https://reviews.femaledaily.com> yang berjumlah 166 *review* produk dengan masing-masing label data positif 53 dan negatif 113 *review* produk yang di unduh pada periode 19 mei 2019.
2. Hasil klasifikasi yang diperoleh oleh sistem adalah Positif dan Negatif..
3. Hasil penelitian adalah perbandingan *naïve bayes* tanpa pembobotan *Gain Ratio* dan dengan pembobotan *Gain Ratio*.
4. *Accuracy*, *Precision* dan *Recall* dilakukan dengan membandingkan klasifikasi *Naïve Bayes* dan *Weigted Naïve Bayes* dengan klasifikasi system.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

1. Membantu perusahaan dalam mengetahui tingkat kepuasan terhadap produk kecantikan yang dipasarkan.

2. Penulis dapat memahami sentiment analis dengan menggunakan metode *Naïve Bayes* dan *Weigted Naïve Bayes*.

