

1. Pendahuluan

Latar Belakang

Membaca suatu ulasan produk dapat membuat masyarakat lebih tertarik dengan produk yang memberikan ulasan baik berdasarkan standar deviasi sebesar 5,06 daripada informasi yang berasal dari tim marketing hanya sebesar 4,36 [9]. Produk kecantikan membutuhkan banyak pertimbangan saat akan membeli, karena setiap kulit memiliki kondisi yang berbeda. Sehingga, masyarakat harus teliti dalam membaca ulasannya. Namun, ulasan yang ditulis pada situs online jumlahnya sangat banyak dan beragam. Hal ini menyebabkan ulasan menjadi bias. Berdasarkan kondisi tersebut, dibutuhkan analisis sentimen berbasis aspek karena tiap kalimat dapat memuat beberapa sentimen [5]. Analisis sentimen tingkat aspek dapat memberikan ringkasan ulasan yang dilakukan secara keseluruhan dokumen [13]. Hal ini dapat diselesaikan dengan pendekatan machine learning karena terbukti terselesaikan lebih akurat sebesar 70-80% daripada manusia hanya 69% [16]. Pada penelitian [19] membahas teks klasifikasi menggunakan metode K-Nearest Neighbor (KNN) dan Information Gain (IG) dengan dataset WebKB, Reuters-21578, Ling-Spam, WWW Prisoner, WWW Beethoven, dan Usenet menunjukkan bahwa metode KNN bekerja secara maksimal, menghasilkan akurasi sebesar 95,5%. Sedangkan pada penelitian [3] menggunakan metode KNN dengan dataset the Polarity v2.0 from Cornell movie review menghasilkan akurasi sebesar 60%, tetapi dengan menambahkan metode IG akurasi meningkat menjadi 96,8% karena fitur yang tidak relevan dihapus. Namun, pada kedua penelitian tersebut hanya membahas tentang klasifikasi teks dan analisis sentimen tingkat dokumen yang hanya mengelompokkan opini ke dalam kelas sentimen positif atau negatif. Pada proposal ini akan membahas tentang analisis sentimen tingkat aspek dengan metode machine learning, yaitu K-Nearest Neighbor (KNN) dengan pemilihan fitur Information Gain (IG). KNN merupakan metode yang menghasilkan bias namun sederhana dan efisien [19]. Permasalahan tersebut dapat diatasi dengan pemilihan fitur. Menurut penelitian [3, 17, 19], metode terbaik untuk pemilihan fitur yang berhubungan adalah IG. Sehingga, KNN mendapatkan kinerja yang lebih baik [3].

Topik dan Batasannya

Topik pada penelitian ini adalah membuat sistem pada analisis sentimen level aspek untuk mengetahui secara detail hasil sentimen berdasarkan aspek yang telah ditentukan. Batasan masalah pada penelitian ini adalah dataset yang berasal dari *website Female Daily* dengan aspek harga, produk, packaging, dan aroma, yang berisikan sentimen positif, netral, dan negatif. Dalam penelitian ini disebutkan bahwa aspek disebut sebagai kelas dan sentimen dianggap sebagai label.

Tujuan

Tujuan dari penelitian ini adalah untuk menganalisis sentimen berbasis aspek pada ulasan produk kecantikan dengan penggunaan stemming dan normalisasi. Serta menganalisis dan mencari threshold terbaik pada IG serta nilai k yang optimal pada metode KNN. Penelitian ini diterapkan pada dataset yang diambil dari *website female daily*.

Organisasi Tulisan

Bagian selanjutnya yang akan dibahas adalah pada bab 2 membahas studi literatur yang terkait dengan penelitian ini. Pada bab 3 membahas teori dan rancangan sistem penelitian. Hasil dan analisis dibahas pada bab 4. Kemudian, kesimpulan dari hasil penelitian yang sudah dilakukan dibahas pada bab 5.