

1. Pendahuluan

Latar Belakang

Hotel adalah salah satu produk pariwisata yang sangat penting bagi wisatawan, perencanaan dan pemesanan hotel di web telah menjadi salah satu kegunaan komersial yang paling penting [1]. Para wisatawan biasanya mencari hotel dengan mempertimbangkan segi pelayanan, fasilitas, maupun jarak tempuh dari wisata yang akan dikunjungi. Karena saat ini teknologi semakin berkembang, para wisatawan bisa memesan hotel melalui website. Biasanya, website tersebut sudah menyediakan fasilitas untuk memberikan *review* atau ulasan tentang pengalaman pribadi seseorang ketika menggunakan hotel. Review inilah yang dapat memberikan informasi baik atau buruknya hotel, agar memudahkan wisatawan untuk memilih hotel, maka dibutuhkan Analisis Sentimen karena layanan pemesanan hotel sering hanya menerbitkan peringkat skalar, misalnya rating antara 1 dan 5 dengan menggunakan bintang. Skor tersebut sangat tidak membantu bagi manajer hotel karena rating tidak memberikan informasi tentang apa yang tamu anggap positif atau tidak menyenangkan [2]. Juga, rating tidak sebanding misalnya ketika sebuah hotel bintang tiga menerima rating yang lebih tinggi daripada hotel bintang empat, itu tidak berarti bahwa yang satu lebih baik daripada yang lain. Bagi manajer hotel komentar pengguna teks akan jauh lebih signifikan daripada rating karena mereka akan tertarik untuk mengetahui apa yang telah dilakukan oleh para pengunjung [2].

Pada riset sebelumnya dilakukan klasifikasi sentimen menggunakan Supervised Learning yaitu SVM dengan feature ekstraksi TF-IDF menghasilkan akurasi 87,2% [3]. Oleh karena itu, pada tugas akhir ini penulis melakukan penelitian terhadap klasifikasi sentimen untuk ulasan hotel yang diberikan oleh tamu hotel dengan menggunakan feature yang lain, yaitu feature selection *chi-square* dan dibandingkan dengan *feature TF-IDF*. Dataset diambil dari beberapa website yang menyediakan fitur ulasan hotel. Penelitian ini memilih metode SVM dengan feature selection Chi-Square pada ulasan hotel dengan mengharapkan akurasi yang lebih bisa dicapai.

Permasalahan yang dibahas pada tugas akhir ini akan melakukan Analisis Sentimen menggunakan metode *Support Vector Machine* pada *review* hotel bahasa Indonesia untuk mengklasifikasikan *review* hotel yang baik atau tidak untuk wisatawan dalam memilih hotel. Metode SVM dipilih karena dapat menghasilkan hasil yang optimal dalam klasifikasi [4]. Selain itu, pada penelitian ini juga menggunakan *feature selection Chi-Square* dan *feature TF-IDF*

sebagai pembandingan, kemudian penggunaan tanpa dan dengan *stemming* juga sebagai pembandingan.

Topik dan Batasannya

Masalah yang dibahas dalam Tugas Akhir ini adalah bagaimana performansi metode *support vector machine* dengan seleksi fitur *chi-square* dan TF-IDF pada data *review* hotel, dan bagaimana pengaruh *Stemming* terhadap tahap *Preprocessing*.

Batasan pekerjaan dalam Tugas Akhir ini, yaitu dataset yang digunakan adalah dataset *review* hotel bahasa Indonesia dan jumlah dataset adalah 200 *review* hotel. Penelitian ini dilakukan pelabelan oleh tiga orang dewasa menggunakan vote dengan pelabelan positif dan negatif. *Preprocessing* yang digunakan adalah *Case Folding*, *Cleansing*, *Tokenization*, *Stemming* dan *Stopword Removal*. *Feature Selection* yang digunakan adalah *Chi-Square* dengan *feature* pembandingan yaitu TF-IDF dan penggunaan tanpa dan dengan *stemming* juga sebagai pembandingan, metode yang digunakan adalah *Support Vector Machine*.

Tujuan

Tujuan dari Tugas Akhir ini adalah untuk mengukur performansi dan akurasi dari metode *Support Vector Machine* dengan membandingkan *feature selection Chi-Square* dengan *feature TF-IDF*. Selain itu juga untuk mengukur performansi penggunaan dengan dan tanpa *stemming*.

Organisasi Tulisan

Selanjutnya pada bab 2 akan dibahas mengenai studi terkait pada penelitian yang dilakukan, bab 3 yang membahas teori dan perancangan sistem penelitian, bab 4 yang membahas evaluasi model penelitian, dan bab 5 yang membahas kesimpulan.