

ABSTRAK

Industri pariwisata mengalami perkembangan yang pesat. Dengan berkembangnya industri dalam bidang pariwisata pemerintah daerah meningkatkan sektor pariwisata dengan mengoptimalkan citra pariwisata yang lebih baik dengan tujuan menarik wisatawan. Maka dari itu, dibutuhkan informasi melalui media sosial untuk mengetahui opini pelanggan mengenai wisata yang ada di Kabupaten Malang untuk bahan evaluasi dalam pengembangan dan kemajuan sektor pariwisata, salah satunya komentar pada media sosial twitter dengan jumlah data 1553. Banyaknya ulasan dan komentar yang diberikan masyarakat sehingga dibutuhkan sebuah cara untuk menganalisis yaitu dengan menerapkan analisis sentimen. Oleh karena itu, dibutuhkan metode sentiment analysis untuk mengetahui sentiment pelanggan wisata di kabupaten malang salah satunya dengan menggunakan algoritma SVM. Algoritma SVM adalah algoritma klasifikasi dengan tingkat akurasi yang tinggi dalam melakukan klasifikasi pada proses analisis sentimen. Walaupun memiliki akurasi yang tinggi algoritma SVM memiliki kelemahan ketika data yang digunakan berdimensi yang tinggi. Oleh karena itu dibutuhkan seleksi fitur untuk mengurangi fitur agar memperingan kerja komputasi metode SVM sehingga mempercepat dalam proses klasifikasi dengan diperlukan proses ekstraksi fitur menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk pembobotan kata dalam dokumen dan menghasilkan fitur sebanyak 2400 fitur. Seleksi fitur *information gain* dapat menurunkan jumlah fitur menjadi 1554 fitur. Hasil evaluasi dari pemodelan SVM mendapatkan nilai akurasi 74%, nilai presisi 76%, nilai recall 74% dan Skor F-1 sebesar 73% dengan menggunakan kernel linear. Hal ini menunjukkan bahwa algoritma *Support Vector Machine* cukup baik dalam model pengklasifikasian teks.

Kata Kunci: *Pariwisata, Crawling data, Analisis Sentimen, TF-IDF, SVM, Information gain*

ABSTRAC

The tourism industry is experiencing rapid growth. With the development of the industry in the field of tourism, local governments improve the tourism sector by optimizing a better tourism image with the aim of attracting tourists. Therefore, information through social media is needed to find out customer opinions about tourism in Malang Regency for evaluation material in the development and progress of the tourism sector, one of which is comments on Twitter social media with a total of 1553 data. There are many reviews and comments given by the public so that It takes a way to analyze that is by applying sentiment analysis. Therefore, a sentiment analysis method is needed to find out tourist customer sentiment in Malang regency, one of which is by using the SVM algorithm. SVM algorithm is a classification algorithm with a high degree of accuracy in classifying the sentiment analysis process. Even though it has high accuracy, the SVM algorithm has a weakness when the data used is of high dimension. Therefore, feature selection is needed to reduce features in order to ease the computational work of the SVM method so as to speed up the classification process by requiring a feature extraction process using TF-IDF (Term Frequency-Inverse Document Frequency) for weighting words in documents and producing as many as 2400 features. Information gain feature selection can reduce the number of features to 1554 features. The evaluation results from the SVM modeling get an accuracy value of 74%, a precision value of 76%, a recall value of 74% and an F-1 score of 73% by using a linear kernel. This shows that the Support Vector Machine algorithm is quite good in text classification models.

Keywords : *Tourism, Crawling data, Sentiment Analysis, TF-IDF, SVM, Information gain.*