

ANALISIS SENTIMEN PADA ULASAN WISATA TAMAN ADIPURA SUMENEP DENGAN PENERAPAN METODE *RANDOM FOREST* DAN SELEKSI FITUR *INFORMATION GAIN*

ABSTRAK

Sumenep merupakan kabupaten di Jawa Timur yang memiliki banyak destinasi wisata, salah satunya Taman Adipura Sumenep. Banyaknya pengunjung yang datang, tidak sedikit yang memberikan opini salah satunya pada aplikasi *goggle maps*. Penelitian ini dilakukan untuk menganalisis sentimen ulasan wisata Taman Adipura Sumenep pada *google maps* dengan menggunakan metode *Random Forest* dan seleksi fitur *Information Gain*. Kelebihan metode *Random Forest* adalah dapat mengelolah data dengan jumlah yang besar, tidak rentang terhadap *noise* dan *outlier*, serta lebih kecil untuk terjadinya *overfitting*. Namun, metode *Random Forest* memiliki kelemahan ketika bekerja dengan jumlah fitur yang banyak. Adapun tujuan dari penelitian ini adalah menerapkan seleksi fitur *Information Gain* untuk mengurangi jumlah fitur agar dapat meningkatkan performa *Random Forest*. Pengambilan data dilakukan dengan teknik *scrapping* menggunakan *beautifulshoup*. Data yang digunakan berjumlah 1972 ulasan. Selanjutnya dilakukan beberapa *text preprocessing*, yang meliputi *casefolding*, *cleaning*, *normalization*, *tokenizing*, *filtering*, *stemming* serta pembobotan kata menggunakan TF-IDF. Hasil dari pengujian model klasifikasi *Random Forest* terjadi peningkatan performa apabila menggunakan seleksi fitur *Information Gain* dengan *threshold* 0.0008 dan data yang sudah di-*oversampling* yang menghasilkan 552 fitur/kata dengan *accuracy* sebesar 98%, *precission* 100%, *recall* 97%, *f1-socre* 99%, dengan waktu pemrosesan mencapai 0.42 detik dibandingkan tanpa *Information Gain* yang menghasilkan 1287 fitur/kata dengan *accuracy* 98%, *precission* 100%, *recall* 97%, *f1-score* 98% dan waktu pemrosesan yang lebih lama mencapai 1.36 detik.

Kata kunci: *Google Maps*, Analisis Sentimen, *Random Forest*, Seleksi Fitur, *Information Gain*, Taman Adipura Sumenep.

ABSTRACT

Sumenep is a district in East Java that has many tourist destinations, one of which is Taman Adipura Sumenep. Many visitors come and some of them give their opinions on the Google Maps app. This research was conducted to analyze the sentiment of reviews of Taman Adipura Sumenep on Google Maps using the Random Forest method and the Information Gain feature selection method. The advantage of the Random Forest method is that it can handle a large amount of data, is not sensitive to noise and outliers, and is less prone to overfitting. However, the Random Forest method has a disadvantage when working with a large number of features. The purpose of this research is to apply the Information Gain feature selection method to reduce the number of features in order to improve the performance of the Random Forest. The data was collected using the scraping technique using beautifulshoup. The data used was 1972 reviews. Then, several text preprocessing was done, which included casefolding, cleaning, normalization, tokenizing, filtering, stemming and word weighting using TF-IDF. The results of the Random Forest classification model testing showed an improvement in performance when using the Information Gain feature selection method with a threshold of 0.0008 and oversampled data, which resulted in 552 features/words with an accuracy of 98%, precision 100%, recall 97%, f1-score 99%, and a processing time of 0.42 seconds compared to without Information Gain which resulted in 1287 features/words with an accuracy of 98%, precision 100%, recall 97%, f1-score 98% and a longer processing time of 1.36 seconds.

Keywords: *Google Maps, Sentiment Analysis, Random Forest, Feature Selection, Information Gain, Taman Adipura Sumenep.*