

**ANALISIS SENTIMEN TERKAIT LAYANAN GOFOOD DAN GRABFOOD
PADA MEDIA SOSIAL *TWITTER* MENGGUNAKAN ALGORITMA
*SUPPORT VECTOR MACHINE (SVM)***

Realdy Agsar Dwi Anggoro

ABSTRAK

Kepopuleran media sosial *twitter* di Indonesia menyebabkan cepatnya penyebaran informasi melalui dunia maya ini. Informasi tersebut pada umumnya dapat berupa opini terhadap suatu layanan ataupun suatu produk. Opini tersebut dapat dianalisis sentimennya agar dapat memahami makna dan emosi yang terkandung dalam kalimat opini tersebut. Oleh karena itu, pada penelitian ini akan dilakukan analisis sentimen pada layanan gofood dan grabfood dengan menggunakan data yang diperoleh dari media sosial *twitter* dengan kata kunci pencarian gofood dan grabfood. Data tersebut akan dilabeli secara manual oleh 2 orang penilai menjadi kelas positif dan kelas negatif. Hasil dari pelabelan tersebut akan dievaluasi menggunakan *kappa value*. Dari hasil pelabelan tersebut untuk data gofood dengan total data sebanyak 386 dokumen *tweet* diperoleh sebanyak sebanyak 75,6% dokumen *tweet* positif dan 24,4% dokumen *tweet* negatif. Sedangkan untuk data grabfood dengan total data sebanyak 336 dokumen *tweet* diperoleh sebanyak 78,6% dokumen *tweet* positif dan 21,4% dokumen *tweet* negatif. Data gofood dan data grabfood yang sudah dilabeli tersebut selanjutnya akan dipraproses dan digabung sebelum lanjut ke tahap pembobotan kata dengan menggunakan metode TF-IDF (*Term Frequency - Inverse Document Frequency*). Data yang telah diberi bobot akan dipisahkan menjadi data latih dan data uji menggunakan metode *hold out estimation* dengan proporsi pembagian data latih dan data uji sebesar 70:30 secara acak dan seimbang. Data-data tersebut masing-masing akan digunakan untuk membuat model klasifikasi dengan menggunakan algoritma *support vector machine (SVM)* dan untuk mengevaluasi performanya. Hasil evaluasi model

pertama dengan menggunakan *confusion matrix* memiliki akurasi sebesar 80,18%, *recall* sebesar 100%, dan *specificity* sebesar 14%. Pada evaluasi model pertama, nilai *recall* sangat tinggi yaitu sebesar 100% namun nilai *specificity* sangat kecil yaitu sebesar 14% karena data pada penelitian ini tidak seimbang lebih didominasi oleh data dengan sentimen positif. Oleh karena itu, pada pembuatan model kedua diterapkan metode *undersampling* pada data latih untuk menangani hal tersebut. Hasil evaluasi pada model kedua memiliki akurasi sebesar 79,26%, *recall* sebesar 86,23%, dan *specificity* sebesar 56%. Dari kedua hasil evaluasi model tersebut nilai *specificity* mengalami kenaikan dari 14% menjadi 56%. Berdasarkan hasil tersebut terbukti bahwa metode *undersampling* dapat mengatasi masalah ketidakseimbangan kelas.

Kata kunci: *twitter*, analisis sentimen, pelabelan, tf-idf, *support vector machine*, *undersampling*

**SENTIMENT ANALYSIS ABOUT GOFOOD AND GRABFOOD
SERVICES ON SOCIAL MEDIA TWITTER USING SUPPORT VECTOR
MACHINE (SVM) ALGORITHM**

Realdy Agsar Dwi Anggoro

ABSTRACT

The popularity of social media Twitter in Indonesia has led to the rapid dissemination of information through this virtual world. This information can generally be in the form of an opinion on a service or a product. The sentiment can be analyzed in order to understand the meaning and emotions contained in the opinion sentence. Therefore, in this study, sentiment analysis will be carried out on gofood and grabfood services using data obtained from social media twitter with the search keywords gofood and grabfood. The data will be manually labeled by 2 anotators into a positive class and a negative class. The results of the labeling will be evaluated using kappa value. From the labeling results for gofood data with a total data of 386 tweet documents, 75.6% of positive tweet documents and 24.4% negative tweet documents obtained. Meanwhile, for grabfood data with a total of 336 tweet documents, 78.6% of positive tweet documents and 21.4% negative tweet documents were obtained. The labeled gofood and grabfood data will then be preprocessed and combined before proceeding to the word weighting stage using the TF-IDF (Term Frequency - Inverse Document Frequency) method. Data that has been weighted will be separated into training data and test data using the hold out estimation method with the proportion of training data and test data being divided randomly and balanced at 70:30. Each of these data will be used to create a classification model using the support vector machine (SVM) algorithm and to evaluate its performance. The results of the evaluation of the first model using the confusion matrix have an accuracy of 80.18%, recall of 100%, and specificity of 14%. In the evaluation of the first model, the recall value is very high at 100% but the specificity value is very

small at 14% because the data in this study is not balanced and is dominated by data with positive sentiments. Therefore, in making the second model an undersampling method was applied to the training data to handle this. The evaluation results in the second model have an accuracy of 79.26%, recall of 86.23%, and specificity of 56%. From the results of the two model evaluations, the specificity value increased from 14% to 56%. Based on these results, it is proven that the undersampling method can overcome the class imbalance problem.

Keywords: twitter, sentiment analysis, labeling, tf-idf, support vector machine, undersampling