

INTISARI

DETEKSI SARKASME UNTUK ANALISIS SENTIMEN PADA *TWEET* BERBAHASA INDONESIA

Oleh

Yessi Yunitasari
16/403723/PPA/05240

Penggunaan media sosial dapat mengubah pola komunikasi seseorang. Salah satu jenis media sosial yang banyak digunakan adalah twitter. Percakapan di tweet dapat diklasifikasikan berdasarkan sentimennya. Adanya sarkasme yang terkandung di dalam suatu tweet mengakibatkan penentuan sentimen suatu tweet menjadi kurang tepat karena sarkasme sulit dianalisis secara otomatis, bahkan oleh manusia sekalipun.

Deteksi sarkasme perlu dilakukan yang diharapkan dapat meningkatkan hasil analisis sentimen. Pengaruh deteksi sarkasme pada analisis sentimen dapat dilihat dari sisi akurasi, presisi, dan recall. Fitur yang digunakan pada deteksi sarkasme adalah unigram dan 4 set fitur Boazizi yang terdiri dari fitur *sentiment-relate*, fitur *punctuation-relate*, fitur *lexical and syntactic*, dan fitur *top word*. Proses klasifikasi sarkasme menggunakan *Random Forest*. Pada analisis sentimen, fitur yang digunakan adalah TF-IDF dan metode klasifikasinya menggunakan Naïve Bayes.

Hasil pengujian menggunakan data twitter untuk analisis sentimen dengan deteksi sarkasme menunjukkan peningkatan rata-rata akurasi sebesar 5,5% dengan nilai akurasi sebesar 80,4%, presisi sebesar 83,2%, dan recall sebesar 91,3%. Sedangkan pengujian deteksi sarkasme yang dilakukan menggunakan *Cross Validation*, menunjukkan akurasi sebesar 72.2%

Kata Kunci : *Naïve bayes*, sarkasme, *tweet*, analisis sentimen, *random forest*

ABSTRACT

SARCASM DETECTION FOR SENTIMENT ANALYSIS IN INDONESIAN LANGUAGE TWEETS

Oleh

Yessi Yunitasari

16/403723/PPA/05240

The use of social media can change person's communication patterns. One type of social media that is widely used is Twitter. Conversations tweeted can be classified based on their sentiments. The existence of sarcasm contained in a tweet resulted in the determination of the sentiment of a tweet to be incorrect because sarcasm is difficult to analyze automatic, even by humans.

Sarcasm detection needs to be done which is expected to improve the results of sentiment analysis. The effect of sarcasm detection on sentiment analysis can be seen in terms of accuracy, precision, and recall. The features used in sarcasm detection are unigram and 4 Boazizi feature sets consisting of sentiment-relate features, punctuation-relate features, lexical and syntactic features, and top word features. The sarcasm classification process uses Random Forest. In the sentiment analysis, the features used are TF-IDF and its classification method using Naïve Bayes.

The test results using twitter data for sentiment analysis with sarcasm detection showed an increase in accuracy of 5.5% with an accurate value of 80.4%, precision of 83.2%, and recall of 91.3%. While the sarcasm detection test carried out using Cross Validation, shows an accuracy of 72.2%

Keywords: Naïve bayes, sarcasm, tweet, sentiment analysis, *random forest*