

ABSTRAK

ANALISIS SENTIMEN TOPIK KORUPSI DI INDONESIA PADA SITUS X MENGGUNAKAN ALGORITMA BERNOULLI NAIVE BAYES

Septian Dwi Wicaksono¹, Syarif Hidayatullaah², Miftahur Rahman³

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email : tianfalconers46@gmail.com

Indonesia merupakan negara dengan jumlah penduduk yang banyak sehingga penggunaan internet berkembang sangat pesat. Twitter merupakan salah satu media sosial yang populer di Indonesia maupun di berbagai negara lainnya. Hal ini membuat Twitter potensial sebagai sumber data yang efisien untuk analisis sentimen. Analisis sentimen adalah studi komputasi yang bertujuan untuk memahami, mengekstrak, dan memproses data teks untuk mendapatkan informasi tentang sentimen dalam suatu kalimat. Dari proses penelitian yang dilakukan tentang identifikasi kasus korupsi di Indonesia, maka hasil yang didapat dari klasifikasi komentar masyarakat dibagi dalam dua kelas yaitu positif dan negatif. Penerapan algoritma *Bernoulli Naive Bayes* dalam konteks ini relevan karena model ini mampu mengatasi kelemahan data teks yang sparse, dimana setiap dokumen dianggap sebagai sekumpulan kata yang diwakili oleh variabel biner yang menjelaskan proses pengumpulan data yang diambil dari Twitter untuk tujuan analisis sentimen terhadap topik korupsi. Hasil dari percobaan data dengan menggunakan percobaan pertama menggunakan perbandingan data latih 10% dan data uji 90% dengan nilai *precision* 83%, *recall* 91% dan *f1-score* 87% dari label negatif, sedangkan nilai *precision* 71%, *recall* 64% dan *f1-score* 71% dari label positif, menghasilkan akurasi sebesar 81,88%. Percobaan kedua menggunakan perbandingan data latih 85% dan data uji 15%, menghasilkan akurasi sebesar 77% dengan nilai *precision* 80%, *recall* 89% dan *f1-score* 84% dari label negatif, sedangkan nilai *precision* 72%, *recall* 55% dan *f1-score* 62% dari label positif. Percobaan ketiga data latih 30% dan 70% data uji menghasilkan akurasi 75% dengan nilai *precision* 79%, *recall* 87% dan *f1-score* 83% dari label negatif, sedangkan nilai *precision* 68%, *recall* 55% dan *f1-score* 61% dari label positif. Dari ketiga percobaan yang telah di uji dapat disimpulkan bahwa dengan data latih 10% dan data uji 90% menghasilkan akurasi yang lebih tinggi.

Kata Kunci: *Twitter, Analisis Sentimen, Korupsi, Bernoulli Naive Bayes*

ABSTRACT

SENTIMENT ANALYSIS OF CORRUPTION TOPICS IN INDONESIA ON WEBSITE X USING BERNOULLI NAIVE BAYES ALGORITHM

Septian Dwi Wicaksono¹, Syarif Hidayatullaah², Miftahur Rahman³

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email : tianfalconers46@gmail.com

Indonesia is a country with a large population so internet usage is growing very rapidly. Twitter is one of the most popular social media in Indonesia and other countries. This makes Twitter a potential source of efficient data for sentiment analysis. Sentiment analysis is a computational study that aims to understand, extract, and process text data to obtain information about the sentiment in a sentence. From the research process conducted on the identification of corruption cases in Indonesia, the results obtained from the classification of public comments are divided into two classes, namely positive and negative. The application of the Bernoulli Naive Bayes algorithm in this context is relevant because this model is able to overcome the weaknesses of sparse text data. where each document is considered as a set of words represented by binary variables that explain the process of collecting data taken from Twitter for the purpose of sentiment analysis on the topic of corruption. Results from data experiments using the first experiment using a 10% training data ratio and 90% test data with a precision value of 83%, recall 91% and f1-score 87% of negative labels, while the precision value is 71%, recall 64% and f1-score 71% of positive labels, resulting in an accuracy of 81.88%. The second experiment used a ratio of 85% training data and 15% test data, resulting in an accuracy of 77% with a precision value of 80%, recall 89% and f1-score 84% from negative labels, while the precision value was 72%, recall 55% and f1-score 62% from positive labels. The third trial of 30% training data and 70% test data resulted in 75% accuracy with 79% precision, 87% recall and 83% f1-score from negative labels, while 68% precision, 55% recall and 61% f1-score from positive labels. From the three experiments that have been tested, it can be concluded that with 10% training data and 90% test data produces higher accuracy.

Keywords: *Twitter, Sentiment Analysis, Corruption, Bernoulli Naive Bayes*