

INTISARI

ANALISIS SENTIMEN KOMENTAR PENGGUNA TERHADAP *THREAD* PADA FORUM BERITA DAN POLITIK DI KASKUS

Oleh
Ema Nur Afifah
14/366061/PA/16207

Kaskus merupakan forum daring terbesar di Indonesia. Salah satu forum yang paling banyak dikunjungi adalah forum Berita dan Politik. Pengguna dapat bebas berkomentar mengenai isu politik terkini melalui situs Kaskus. Situs Kaskus digunakan dalam penelitian ini sebagai sumber data untuk menganalisis sentimen komentar *thread* pada forum Berita dan Politik di Kaskus. Analisis yang digunakan menggunakan metode Multinomial Naïve Bayes (MNB) dan Support Vector Machine (SVM), kemudian dibandingkan metode manakah yang paling sesuai. Pertama, data komentar dikumpulkan dengan metode *scraping* data dari situs Kaskus. Kedua, data yang diperoleh dilabeli secara manual untuk dijadikan data training. Kemudian dilakukan *preprocessing* dengan kamus dan tanpa kamus untuk memudahkan melakukan analisis sekaligus menambahkan daftar kata *slang* untuk penggantian dengan kata baku. Setelah data tersebut melewati *preprocessing*, kemudian dianalisis untuk mengetahui metode manakah yang memiliki akurasi paling baik dan dibandingkan hasilnya antara penggunaan kamus dan tanpa kamus.

Pada percobaan ini, digunakan 1000 komentar *thread* yang telah dilabeli yaitu 400 positif, 400 negatif, dan 200 netral. Pada *preprocessing* menggunakan kamus dihasilkan rata-rata akurasi metode MNB sebesar 50,4%, *precision score* 47,1%, *recall score* 42,6%, *f1 score* 38,9%, dan *run time* 7532s. Sedangkan tanpa kamus dihasilkan akurasi sebesar 50,8%, *precision score* 47,1%, *recall score* 43%, *f1 score* 40%, dan *run time* 6,67s. Pada metode SVM *preprocessing* menggunakan kamus dihasilkan rata-rata akurasi sebesar 48,2%, *precision score* 43,3%, *recall score* 42,5%, *f1 score* 41,5% dan *run time* 7531s. Sedangkan tanpa kamus dihasilkan akurasi sebesar 48,1%, *precision score* 42,9%, *recall score* 42,3%, *f1 score* 41,1%, dan *run time* 6,69s. Dari hasil ini, diketahui bahwa tanpa penggunaan kamus, hasil akurasi dan *evaluation metrics* kedua metode menjadi lebih baik, serta memiliki *run time* yang lebih tinggi. Metode MNB mampu menganalisis komentar *thread* di Kaskus lebih baik daripada SVM baik menggunakan kamus maupun tanpa kamus.

Kata Kunci: Kaskus, Analisis Sentimen, TF-IDF, Support Vectore Macine, Multinomial Naïve Bayes

ABSTRACT

SENTIMENT ANALYSIST OF USER COMMENT ABOUT THREAD OF BERITA AND POLITIK FORUM AT KASKUS

By

Ema Nur Afifah

14/366061/PA/16207

Kaskus is the largest online forum in Indonesia. One of the most visited forums is the News and Politics forum. Users can freely comment on current political issues through the Kaskus website. The Kaskus website is used in this study as a source of data for analyzing sentiment comment threads on the News and Politics forum at Kaskus. The analysis used using the Multinomial Naïve Bayes (MNB) method and Support Vector Machine (SVM), then compare which method is most suitable. First, comment data is collected by scraping data from the Kaskus website. Second, the data obtained is labeled manually to be used as training data. Then preprocessing is done with a dictionary and without a dictionary to make it easy to do the analysis while adding a list of slang words to replace with standard words. After the data has passed through preprocessing, it is then analyzed to find out which method has the best accuracy and compare the results between using a dictionary and without a dictionary.

In this experiment, 1000 thread comments that have been labeled are 400 positive, 400 negative, and 200 neutral. In preprocessing using a dictionary, the average accuracy of MNB method is 50.4%, 47.1% precision score, 42.6% recall score, f1 score 38,9%, and run time 7532s. Whereas without a dictionary the accuracy is 50.8%, the precision score is 47.1%, the recall score is 43%, the f1 score is 40%, and the run time is 6.67s. In the SVM preprocessing method using a dictionary produced an average accuracy of 48.2%, 43.3% precision score, 42.5% recall score, 41.5% f1 score and 7531s run time. Meanwhile, without a dictionary, the accuracy is 48.1%, the precision score is 42.9%, the recall score is 42.3%, the f1 score is 41.1%, and the run time is 6.69s. From these results, it is known that without the use of dictionaries, the results of the accuracy and evaluation metrics of the two methods are better, and have a higher run time. The MNB method is able to analyze thread comments on Kaskus better than SVM using both dictionaries and without dictionaries.

Keyword: Kaskus, Sentiment Analysis, TF-IDF, Support Vector Machine, Multinomial Naïve Bayes