

INTISARI

ANALISIS SENTIMEN PADA TWITTER BERBAHASA INDONESIA MENGUNAKAN METODE BERBASIS GRAPH DAN SUPPORT VECTOR MACHINE

ABDUL MALIK
15/388453/PPA/04892

Besarnya pengaruh dan manfaat dari analisis sentimen pada data media sosial menyebabkan penelitian di bidang ini berkembang pesat, setiap tahun banyak pendekatan model yang dikembangkan. Salah satu metode yang populer digunakan yaitu dengan pendekatan *machine learning*. Penggunaan *machine learning* untuk analisis sentimen pada dokumen teks membutuhkan proses ekstraksi fitur untuk membangun bobot *vector* yang digunakan pada *machine learning* dalam membangun model klasifikasi. Metode pembobotan fitur yang umum digunakan yaitu dengan pendekatan model TF-IDF dengan cara mentransformasikan data tekstual menggunakan representasi tabel.

Penelitian ini menggunakan pendekatan model berbasis *graph* dan *machine learning* untuk analisis sentimen pada dokumen *tweet* berbahasa Indonesia. Model *graph* digunakan untuk proses ekstraksi fitur dalam proses pembobotan *term* dimana dokumen *tweet* direpresentasikan kedalam model *graph* dan metode *machine learning* yang digunakan yaitu *support vector machine* (SVM). Proses pembobotan fitur menggunakan metode TW-IDF dimana bobot fitur dihitung berdasarkan ukuran *centralitas* node dari model *graph*. Ada dua model pengukuran *centralitas* yang diuji dalam penelitian ini, yaitu *degree centrality* dan *closeeness centrality*.

Hasil pengujian menunjukkan model SVM yang dibangun menggunakan fitur TW-IDF berdasarkan ukuran *centralitas* node dari *graph* dapat memberikan kinerja yang baik. Pada proses pengujian model SVM dengan fitur TW-IDF berdasarkan ukuran *closeeness centrality* dapat memberikan nilai akurasi 88,60%, nilai *f-measure* 89,36%. dan untuk ukuran *degree centrality* dapat memberikan nilai akurasi 87,00%, nilai *f-measure* 87,33%. Sedangkan untuk Model SVM yang dibangun menggunakan fitur TF-IDF memberikan nilai akurasi 87,60%, dan nilai *f-measure* 88,36%.

Kata kunci: analisis sentimen, SVM, *graph-based*, *centrality measures*.

ABSTRACT

ANALISIS SENTIMEN PADA TWITTER BERBAHASA INDONESIA MENGUNAKAN METODE BERBASIS GRAPH DAN SUPPORT VECTOR MACHINE

ABDUL MALIK
15/388453/PPA/04892

The magnitude of the influence and benefits of sentiment analysis on social media data has led to a rapid development of research in this field, each year with many model approaches being developed. One method that is popularly used is the machine learning approach. The use of machine learning for sentiment analysis in text documents requires a feature extraction process to build vector weights used in machine learning in building a classification model. The commonly used feature weighting method is the TF-IDF model approach by transforming textual data using table representation.

This study uses a graph and machine learning based approach to sentiment analysis in Indonesian tweet documents. Graph model is used for feature extraction processes in the term weighting process where the tweet document is represented in the graph model and the machine learning method used is support vector machine (SVM). Feature weighting process uses the TW-IDF method where feature weights are calculated based on the node centrality size of the graph model. There are two centrality measurement models tested in this study, namely degree centrality and closeness centrality.

The test results show that the SVM model built using the TW-IDF feature based on the size of the centrality in the graph model can provide good performance. In the testing process, the SVM model with the TW-IDF feature based on the size of the closeness centrality can provide an accuracy value of 88.60%, f-measure value of 89.36%. and for degree centrality size can give an accuracy value of 87.00%, f-measure is 87.33%. Whereas for the SVM Model which was built using the TF-IDF feature, it gave an accuracy value of 87.60%, and f-measure value of 88.36%.

Keyword: sentiment analysis, SVM, graph-based, centrality measures.