

INTISARI

PERBANDINGAN EKSTRAKSI FITUR TF-IDF DAN BAG OF WORDS UNTUK KLASIFIKASI EMOSI PADA DATA TWITTER MENGGUNAKAN SVM DAN KNN

oleh

Abyan Burhanuddin Ahmad
18/427568/PA/18528

Sosial media telah berkembang pesat di seluruh dunia selama satu dekade terakhir ini. Berdasarkan data statistik *Hootsuite* di tahun 2020, mayoritas pengguna internet juga merupakan pengguna sosial media. Salah satu media sosial paling populer di Indonesia adalah twitter. Twitter sering digunakan penggunanya untuk melakukan diskusi dan percakapan dalam bentuk *tweet*. Percakapan dan diskusi dalam bentuk *tweet* tersebut mengandung berbagai data yang dapat dianalisa seperti emosi untuk mengetahui respon, perasaan, atau pendapat orang-orang mengenai topik tertentu.

Penelitian ini melakukan perbandingan performa ekstraksi fitur TF-IDF dan *Bag-of-Words* pada permasalahan emosi klasifikasi menggunakan dataset twitter berbahasa Indonesia. Penelitian ini juga menggunakan *K-Nearest Neighbor* (K-NN) dan *Support Vector Machine* (SVM) sebagai metode klasifikasi.

Setelah dilakukannya penelitian ini, diperoleh kesimpulan bahwa secara keseluruhan, pada model dengan parameter terbaik, ekstraksi fitur TF-IDF memiliki performa yang lebih baik dibandingkan dengan *Bag of Words*. Model terbaik yang ditemukan merupakan model SVM dengan ekstraksi fitur TF-IDF dengan akurasi, presisi, *recall*, dan *F1-measure* secara berturut-turut adalah 64.22%, 67.4%, 64.57%, dan 65.19%.

Kata kunci: *Bag-of-Words*, klasifikasi emosi, KNN, SVM, TF-IDF, Twitter

ABSTRACT

COMPARISON OF FEATURE EXTRACTION TF-IDF AND BAG OF WORDS FOR EMOTION CLASSIFICATION ON TWITTER DATA USING SVM AND KNN

by

Abyan Burhanuddin Ahmad
18/427568/PA/18528

Social media has grown rapidly all over the world over the past decade. Based on Hootsuite statistical data in 2020, the majority of internet users are also social media users. One of the most popular social media in Indonesia is Twitter. Twitter is often used by users to conduct discussions and conversations in the form of tweets. This conversations and discussions contains various data that can be analyzed such as emotions to find out the responses, feelings, or opinions of people on certain topics.

This study compares the performance of TF-IDF and Bag-of-Words feature extraction on emotional classification problems using the Indonesian language twitter dataset. This study also uses K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) as classification methods.

After doing this research, it was concluded that overall, in the models with the best parameter, TF-IDF feature extraction has a better performance than Bag of Words. The best model found is the SVM model with TF-IDF feature extraction with accuracy, precision, recall, and F1-measure are 64.22%, 67.4%, 64.57%, and 65.19%, respectively.

Keywords: *Bag-of-Words*, emotion classification, KNN, SVM, TF-IDF, Twitter