

Abstract

Twitter is one of the popular social media to channel opinions in the form of criticism and suggestions. Criticism can be categorized as hate speech if the criticism attacks something targeted. With the limit of 280 characters in a tweet, there is often a mismatch of vocabulary due to abbreviations. This can be solved with *word embedding*. This study utilizes feature expansion to reduce vocabulary mismatches in hate speech on Twitter containing Indonesian by using *Global Vectors (GloVe)*. Feature selection related to the best model is carried out using the *Logistic Regression (LR)*, *Random Forest (RF)*, and *Artificial Neural Network (ANN)* algorithms. The results show that feature expansion is quite effective in improving performance with an accuracy rate of 88.59% on the RF model of 5,000 features with a combination of TF-IDF Tweet corpus. Accuracy increase reached 2.37% by ANN classifier with Top 10 from Tweet, IndoNews corpus compared to the predetermined *baseline*.

Keywords: hate speech, GloVe, feature expansion