

Abstract

Hate speech on social media, especially Twitter, often takes the form of racism, sexism, or political interests aimed at certain individuals or groups. These actions can trigger crime, riots, violence and even resistance to individuals or groups. Therefore, we need a process of classifying a tweet whether it is hate speech or not to reduce the abuse that occurs on Twitter. The technology used in the classification of hate speech that is most commonly used is neural networks that require user data and meta data. In previous studies, the Naïve Bayes (NB) method has been used using the bigram, unigram and feature selection features with an accuracy of 80-85%. The k-Nearest Neighbor (kNN) method has also been used which has an accuracy of 70-85% on the clarification of hate speech by political figures. Meanwhile, the most widely used method is the Support Vector Machine (SVM) method with an accuracy of 70 to the highest 95%. To get a higher accuracy in the classification of hate speech, this study will perform a Hybrid Classifier on the Hate Speech Hashtag Classification process using a combination method of MLP, kNN, NB. The data used in this study are Twitter Tweets from November 2021 to June 2022 regarding trending hashtags. The average accuracy performance results obtained using MLP, kNN, NB were 72%, 63%, 73% respectively. To improve the accuracy of the classification results of the three methods, a combination of methods using the Hybrid Classifier is carried out. Experimental results show Hybrid Classifier with voting method can increase accuracy up to 74%. It was found that the use of a hybrid can provide a better system performance than the 3 classifiers in its composition, namely kNN, NB and MLP.