
Abstract

Public figures are often scrutinized by social media users, either because of what they say or even because of their role in a television series. Generally, public figures upload something on their social media accounts to help shape their image. But not everyone who sees it is happy, some even dislike the upload. This study aims to determine public sentiment towards public figure Anya Geraldine conveyed on Twitter in Indonesian. The classification process in this study uses the Adaptive Boosting (AdaBoost) and Extreme Gradient Boosting (XGBoost) classification methods with text preprocessing using cleaning, case folding, tokenization and filtering. The data used are tweets in Indonesian with the keyword ”@anyaselalubenaar”, with a total dataset of 7,475 tweets divided into 6,887 positive tweets and 588 negative tweets. From the label results using oversampling to avoid excessive overfitting problems. The feature used is TF-IDF weighting. To validate the effectiveness of the model used, 4 experimental scenarios were carried out, namely first model performance without oversampling, second model performance with oversampling, third model performance with undersampling and fourth model performance with Hyperparameter tune. The experimental results show that XGBoost+SMOTE+Hyperparameter achieved 95% compared to AdaBoost+SMOTE+Hyperparameter of 87%. The application of SMOTE and Hyperparameter tune is proven to overcome the problem of data imbalance and get better classification results.

Keywords: Twitter, Hate Speech, Sentiment, Analysis, AdaBoost, XGBoost
