
Abstrak

Publik figur sering mendapat sorotan dari pengguna media sosial, baik karena ucapannya atau bahkan karena peranannya di dalam suatu serial televisi. Umumnya publik figur mengunggah sesuatu di akun sosial media mereka membantu untuk membentuk citra mereka. Namun tidak semua orang yang melihatnya senang, bahkan ada yang tidak suka dengan unggahan tersebut. Penelitian ini bertujuan untuk mengetahui sentimen publik terhadap publik figur Anya Geraldine yang disampaikan di Twitter dalam bahasa Indonesia. Proses klasifikasi pada penelitian ini menggunakan metode klasifikasi *Adaptive Boosting* (AdaBoost) dan *Extreme Gradient Boosting* (XGBoost) dengan *preprocessing* teks menggunakan *cleaning*, *case folding*, *tokenizing* dan *filtering*. Data yang digunakan adalah tweet dalam bahasa Indonesia dengan kata kunci “@anyaselalubelar”, dengan jumlah dataset sebanyak 7.475 tweet dibagi menjadi 6.887 tweet positif dan 588 tweet negatif. Dari hasil label tersebut menggunakan *oversampling* untuk menghindari masalah *overfitting* yang berlebihan. Fitur yang digunakan adalah pembobotan TF-IDF. Untuk memvalidasi keefektifan model yang digunakan, dilakukan 4 skenario eksperimental yaitu pertama kinerja model tanpa *oversampling*, kedua kinerja model dengan *oversampling*, ketiga kinerja model dengan *under sampling* dan keempat kinerja model dengan *hyperparameter tune*. Hasil eksperimen menunjukkan XGBoost+SMOTE+Hyperparameter mencapai 95% dibandingkan AdaBoost+SMOTE+Hyperparameter sebesar 87%. Penerapan SMOTE dan *hyperparameter tune* terbukti dapat mengatasi masalah ketidakseimbangan data dan mendapatkan hasil klasifikasi yang lebih baik.

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