

ABSTRAK

Media sosial Twitter dengan 586 juta pengguna aktif merupakan sebuah ruang publik yang memiliki peran dalam penyebaran informasi yang sangat cepat. Namun, keberadaan akun palsu atau *social spambots* yang semakin berkembang di Twitter mampu meniru perilaku pengguna asli sehingga menimbulkan tantangan dalam menjaga kualitas interaksi dan kredibilitas informasi. Berbagai pendekatan deteksi akun palsu berbasis satu modalitas saja belum cukup merepresentasikan karakteristik akun secara baik. Penelitian ini mengusulkan model multimodal *deep learning* yang menggabungkan metadata akun dan konten teks *tweet* melalui pendekatan *late fusion* untuk mengklasifikasikan pengguna *fake* dan *real* pada media sosial X. Metadata akun diproses menggunakan *Multi-Layer Perceptron* (MLP), sedangkan konten teks *tweet* diproses menggunakan *Long Short-Term Memory* (LSTM) dengan *word embedding* GloVe Twitter. Dataset yang digunakan adalah Cresci-2017 dengan penanganan ketidakseimbangan kelas menggunakan teknik *random undersampling*. Hasil penelitian menunjukkan bahwa pada pendekatan berbasis metadata, XGBoost memperoleh *F1-score* 0,9950 dan *F1-score* sebesar MLP 0,9842, sedangkan pada pendekatan teks *tweet*, SVM mencapai *F1-score* sebesar 0,9210 dan LSTM mencapai *F1-score* 0,9392. Pendekatan multimodal menggabungkan kedua modalitas melalui *late fusion* dengan menghasilkan performa terbaik yaitu *accuracy* sebesar 0,9920 dan *weighted F1-score* sebesar 0,9921, baik pada multimodal *deep learning* maupun *machine learning*. Hasil menunjukkan bahwa peningkatan hasil evaluasi pada pemodelan multimodal terhadap pemodelan unimodal menunjukkan bahwa model mampu memberikan representasi yang lebih baik melalui penggabungan informasi dari teks *tweet* dan metadata akun, serta mampu menjaga keseimbangan performa antara kedua kelas pengguna.

Kata Kunci: Akun Palsu, *Social Spambots*, Media Sosial X, Multimodal *Deep Learning*, LSTM, MLP, *Late Fusion*, Cresci-2017