

Implementasi Metode LSTM Sebagai Pendeteksi Bahasa Kasar Pada Media Sosial

Dira Afwa Mardanela¹, Dr. Moch. Arif Bijaksana, Ir. M. Tech.², Widi Astuti, S.T.,M.Kom.³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹diraafwam@students.telkomuniversity.ac.id, ²arifbijaksana@telkomuniversity.ac.id,

³widiwdu@telkomuniversity.ac.id

Abstrak

Kebebasan berekspresi di media sosial sekarang ini tidak berjalan seiring dengan kesiapan literasi masyarakat, serta pengetahuan yang rendah tentang pemanfaatan media sosial sebagai alat komunikasi menyebabkan pelanggaran tindak tutur kesantunan dalam berkomunikasi di media sosial. Kebebasan berekspresi dengan mengutarakan kebencian menggunakan bahasa kasar ini sering ditemukan di media sosial. Oleh karena itu pada penelitian ini dilakukan pendeteksi bahasa kasar berdasarkan komentar di media sosial menggunakan *Long Short Term Memory* (LSTM). LSTM merupakan salah satu model kelanjutan dari *Recurrent Neural Network* (RNN). agar dapat mendeteksi Dataset yang digunakan terdiri dari tiga label yaitu *Not Abusive*, *Abusive and Offensive*, dan *Abusive but Not Offensive*. Hasil pengujian arsitektur LSTM didapatkan hasil bahwa arsitektur LSTM tidak mampu untuk memprediksi kelas minoritas, tetapi didapatkan kombinasi pembagian dataset berupa 70% data latih dan 30% data uji serta ukuran *batch size* sebesar 32, sehingga menghasilkan F1 Score terbaik sebesar 94% untuk kelas mayoritas *Not Abusive*.

Kata kunci : bahasa kasar, LSTM, *f1 score*, pendeteksi bahasa kasar

Abstract

Freedom of expression on social media today does not go hand in hand with public literacy readiness, and low knowledge about the use of social media as a communication tool causes violations of polite speech acts in communicating on social media. Freedom of expression by expressing hatred using abusive language is often found on social media. Therefore, in this study, abusive language detection was carried out based on comments on social media using Long Short Term Memory (LSTM). LSTM is a continuation model of the Recurrent Neural Network (RNN). in order to detect the dataset used consists of three labels, namely Not Abusive, Abusive and Offensive, and Abusive but Not Offensive. The results of the LSTM architecture testing show that the LSTM architecture is not able to predict the minority class, but a combination of dataset distribution in the form of 70% training data and 30% test data and batch size of 32, resulting in the best F1 Score of 94% for the majority class Not Abusive.

Keywords: abusive language, LSTM, *f1 score*, abusive language detector
