

INTISARI

METODE *BIDIRECTIONAL GATED RECURRENT UNIT* DAN *FASTTEXT EMBEDDING* UNTUK DETEKSI UJARAN KEBENCIAN BAHASA INDONESIA

Oleh

Reza Pahlevi Sofyan
19/448728/PPA/05811

Data teks yang digunakan dalam menganalisa deteksi ujaran kebencian sering kali terdapat kata-kata yang tidak dapat diambil informasinya dikarenakan kata tersebut tidak dikenali atau out-of-vocabulary (OOV) yang ditunjukkan pada penelitian sebelumnya yang menggunakan *word2vec* sebagai *word embedding*. Kata tersebut dapat berupa kata provokatif yang mengindikasikan ujaran kebencian sehingga dapat menjadi masalah dalam mendeteksi adanya ujaran kebencian. Penelitian ini mengusulkan *Fasttext* sebagai *word embedding* yang dapat menangani permasalahan OOV yang dikombinasikan dengan *Bidirectional gated recurrent unit* (BiGRU) sebagai metode klasifikasi. Penelitian ini bertujuan untuk menganalisa performa kombinasi kedua metode dalam mendeteksi ujaran kebencian yang dibandingkan dengan metode lainnya seperti *word2vec embedding* dan metode *machine learning*. Pengujian dilakukan dengan menerapkan *cross validation* yang diujikan berdasarkan parameter uji pada penelitian dengan banyak data yang diolah sebesar 2185 data. Hasil eksperimen menunjukkan kombinasi *fasttext embedding* dan BiGRU mencapai performa terbaik f1-score sebesar 80,6%. Kontribusi dalam penelitian ini adalah menunjukkan perbandingan kinerja *word embedding* antara *word2vec* dan *fasttext*, perbandingan performa metode usulan dengan metode *machine learning*, serta analisa peningkatan hasil performa dari metode yang diusulkan dengan *tuning hyperparameter* pada BiGRU.

Kata Kunci : Ujaran Kebencian, *Bidirectional Gated Recurrent Unit*, *Fasttext*, *Word Embedding*, Deteksi

ABSTRACT

BIDIRECTIONAL GATED RECURRENT UNIT METHOD AND EMBEDDING FASTTEXT FOR HATE SPEECH DETECTION IN INDONESIAN LANGUAGE

By

Reza Pahlevi Sofyan
19/448728/PPA/05811

The text data used in analyzing hate speech detection often contain words that cannot be retrieved because the word is not recognized or out-of-vocabulary (OOV), as shown in previous research using word2vec as word embedding. The word can be a provocative word that indicates hate speech, so that it can be a problem in detecting hate speech. This research proposes Fasttext as word embedding that can handle OOV problems combined with a Bidirectional gated recurrent unit (BiGRU) as a classification method. This research aims to analyze the performance of the combination of the two methods in detecting hate speech compared to other methods such as word2vec embedding and machine learning methods. The testing was done by applying cross-validation, which was tested based on the test parameters in the research with a lot of processed data of 2185 data. The experimental results show that the combination of fasttext and BiGRU achieves the best performance f1-score of 80.6%. The contribution in this study is to show a comparison of the performance of word embedding between word2vec and fasttext, comparing the performance of the proposed method with the machine learning method, and analysis of improving the performance results of the proposed method by tuning hyperparameters on BiGRU.

Keywords : Hate speech, bidirectional gated recurrent unit, fasttext, word embedding, detection