

INTISARI

DETEKSI HATE SPEECH BAHASA INDONESIA MENGGUNAKAN WORD EMBEDDING DAN GATED RECURRENT UNIT

JUNANDA PATIHULLAH
15/388484/PPA/04923

Penyebaran ujaran kebencian tidak hanya dapat disebar lewat ucapan ataupun tindakan, tetapi juga dapat disebar lewat tulisan dan media internet. Semakin meningkatnya aktifitas pengguna di media sosial dan penyebaran informasi yang sangat cepat membuat deteksi ujaran menjadi sulit, terlebih lagi penggunaan *bag of word model* sebagai ekstraksi fitur pada metode *machine learning* tidak cukup memadai untuk mendeteksi ujaran kebencian dalam bahasa Indonesia. Dikarenakan *bag of word model* hanya melihat frekuensi kemunculan kata tanpa memperdulikan hubungan antara kata.

Pada penelitian ini menggunakan metode *word2vec* untuk melakukan ekstraksi fitur karena kemampuan *word2vec* yang dapat mempelajari hubungan atau semantik kata satu dengan kata lainnya. Algoritme *machine learning* yang digunakan adalah *gated recurrent unit*. Algoritme *gated recurrent unit* (GRU) jenis metode *deep learning* yang melakukan proses klasifikasi dan mampu mempelajari hubungan informasi dari waktu sebelumnya dengan waktu sekarang sehingga memiliki kinerja yang cukup baik untuk masalah klasifikasi teks.

Penggunaan *word2vec* sebagai ekstraksi fitur dari algoritme *gated recurrent unit* dengan arsitektur GRU 2 layer, learning rate 0,001 dan 200 neuron pada layer tersembunyi menghasilkan nilai akurasi 95,00% dan *bag of word model* sebagai ekstraksi fitur dengan algoritme random forest decision tree menghasilkan akurasi 82,69%. Penggunaan *word2vec* sebagai ekstraksi fitur dengan algoritme GRU memiliki kinerja yang cukup baik dalam mendeteksi ujaran kebencian.

Kata kunci : gated recurrent unit, ujaran kebencian, deep learning, word2vec, word embedding

ABSTRACT

HATE SPEECH DETECTION IN THE INDONESIAN LANGUAGE USING WORD EMBEDDING AND GATED RECURRENT UNIT

JUNANDA PATIHULLAH
15/388484/PPA/04923

The spread of hate speech can not only be spread through speech or action, but can also be disseminated through writing and internet media. The increasing activity of users in social media and the rapid dissemination of information make speech detection difficult, moreover the use of bag of word models as feature extraction in the machine learning method is not sufficient to detect speech hate in Indonesian. Because the bag of word model only sees the frequency of occurrence of words regardless of the relationship between words.

In this study, the word2vec method was used to extract features because of word2vec's ability to learn the relationship or semantics of one word with other words. The machine learning algorithm used is gated recurrent unit. The gated recurrent unit (GRU) algorithm is a type of deep learning method that performs the classification process and is able to learn information relations from the previous time with the current time so that it has a pretty good performance for text classification problems.

The use of word2vec as feature extraction from the gated recurrent unit algorithm with GRU 2 layer architecture, learning rate 0.001 and 200 neurons in hidden layers produces an accuracy value of 95.00% and bag of word model as feature extraction with a decision tree random algorithm resulting in 82.69% accuracy. The use of word2vec as feature extraction with the GRU algorithm has a pretty good performance in detecting hate speech.

Keywords : gated recurrent unit, hate speech, deep learning, word2vec, word embedding