

ABSTRAK

BIDIRECTIONAL LONG SHORT TERM MEMORY METHOD AND WORD2VEC EXTRACTION APPROACH FOR HATE SPEECH ANALYSIS OF INDONESIAN POLITICAL FIGURES

By

Auliya Rahman Isnain
17/418625 / PPA / 05408

Hate speech is communication that looks down on someone or group based on characteristics such as (race, ethnicity, gender, citizenship, religion and organization) Alfina, (2017). Twitter is a social media that allows users to express their feelings and opinions through tweets, including tweets that contain expressions of hatred. Because Twitter has a major influence on the success or destruction of one's image (Buntoro, 2016).

The purpose of this research is to classify hate speech or hate speech using the Bidirectional Long Short Term Memory method and the word2vec feature extraction method with Continuous bag-of-words (CBOW) architecture. For testing the BiLSTM method with the calculation of the value of accuracy, precision, recall, and F-measure. The use of word2vec and the Bidirectional Long Short Term Memory method with CBOW architecture, with epoch 10, learning rate 0.001 and the number of neurons 150 on the hidden layer, produce an accuracy rate of 85.27%, with a precision value of 78.58%, recall 85, respectively 48% and F-measure 80.93%. As for Bidirectional Long Short Term Memory by adding one layer has an accuracy of 86.25%. The addition of one layer to BiLSTM only increased by 0.98%.

Keywords: Hate Speech, LSTM, BiLSTM, Word2vec, CBOW, Skipgram, Twitter.

INTISARI

METODE *BIDIRECTIONAL LONG SHORT TERM MEMORY* DAN PENDEKATAN EKSTRAKSI FITUR *WORD2VEC* UNTUK DETEKSI *HATE SPEECH* TERHADAP TOKOH POLITIK INDONESIA

Oleh

Auliya Rahman Isnain

17/418625/PPA/05408

Perbincangan mengenai ujaran kebencian saat ini sedang hangat di Indonesia, terutama melalui media sosial. Ujaran kebencian merupakan komunikasi yang meremehkan seseorang atau kelompok berdasarkan karakteristik seperti (ras, etnis, jenis kelamin, kewarganegaraan, agama dan organisasi). Twitter salah satu media sosial yang digunakan seseorang untuk mengutarakan perasaan dan opini melalui *tweet*, termasuk *tweet* yang mengandung ujaran kebencian. Karena twitter mempunyai pengaruh besar bagi kesuksesan ataupun kehancuran citra seseorang.

Penelitian ini bertujuan untuk mendeteksi ujaran kebencian atau bukan ujaran kebencian dengan menggunakan metode *Bidirectional Long Short Term Memory* dan metode ekstraksi fitur *word2vec* dengan arsitektur *Continuous bag-of-words* (CBOW). Untuk pengujian metode BiLSTM dengan perhitungan nilai akurasi, presisi, *recall*, dan *F-measure*.

Penggunaan *word2vec* dan metode *Bidirectional Long Short Term Memory* dengan arsitektur CBOW, dengan *epoch* 10, *learning rate* 0.001 dan jumlah neuron 200 pada layer tersembunyi, menghasilkan tingkat akurasi 91,08%, dengan masing-masing nilai presisi 92,00%, *recall* 90,20% dan *F-measure* 91,09%. Sedangkan untuk *Bidirectional Long Short Term Memory* dengan tiga layer memiliki akurasi 93,06%. Penambahan satu layer pada BiLSTM meningkat 1,98%.

Kata Kunci: Ujaran Kebencian, LSTM, BiLSTM, *Word2vec*, CBOW, Skipgram, Twitter.