

ABSTRACT

Increasing the amount of data in sentiment analysis research become a problem that requires a different approach. Traditional machine learning such as Naïve bayes, Support Vector Machine, and Maximum Entropy has limitations in handling large amount of data (big data era). Deep learning approach appears to be one of the approaches that can be used. Deep learning in some research provide some good results, especially in sentiment analysis research. This research will use the LSTM approach to do sentiment analysis. This LSTM method was chosen because it's able to handle large amounts of data, and also data long-term dependencies. The CNN and RNN methods has limitations in handling long data like long-term dependencies.

In this research, the size variations from the dimension of word embedding were added to analyze the effect on the results of the accuracy and training time required by the LSTM method. This word embedding was created using embedding layer from Keras, word2vec Skip-Gram and CBOW. From this research, the LSTM method had the highest accuracy of 88.5% in dimension of 50 and 128 with the word2vec Skip-Gram model. And had the prediction time less than 1 second by trained LSTM model with the dimensions less than 50.

Keywords : *Sentiment Analysis, Deep learning, LSTM, Word embedding*

INTISARI

Peningkatan jumlah data yang diproses dalam kasus analisis sentimen menjadi permasalahan yang memerlukan pendekatan yang berbeda. Pendekatan dengan machine learning tradisional seperti metode Naïve Bayes, Support Vector Machine ataupun Maximum Entropy memiliki keterbatasan dalam menangani data dalam jumlah yang banyak. Deep learning menjadi salah satu pilihan pendekatan yang bisa digunakan. Dalam beberapa penelitian, deep learning mampu memberikan hasil yang bagus, tak terkecuali pada penelitian analisis sentimen. Pada penelitian analisis sentimen ini, metode deep learning yang digunakan adalah metode LSTM. LSTM dipilih karena mampu menangani data dalam jumlah yang banyak dan juga sekaligus yang berukuran panjang (*long-term dependency*). Metode CNN dan RNN memiliki kelemahan dalam menangani data yang panjang.

Dalam penelitian ini, variasi ukuran dari dimensi vektor data ditambahkan untuk menganalisa pengaruhnya terhadap hasil akurasi dan waktu training yang diperlukan oleh metode LSTM. Variasi ukuran dimensi dari vektor data dibuat dengan menggunakan *embedding layer library* Keras, model word2vec Skip-Gram dan CBOW. Hasil yang diperoleh pada penelitian adalah akurasi tertinggi sebesar 88.5% pada ukuran dimensi 50 dan 128 dengan model word2vec Skip-Gram. Serta waktu prediksi kurang dari 1 detik oleh model LSTM terlatih dengan ukuran dimensi kurang dari 50.

Kata kunci – Analisis sentimen, Deep learning, LSTM, word embedding



UNIVERSITAS
GADJAH MADA

Pengaruh Word Embedding Dimension Reduction Terhadap Kinerja LSTM untuk Analisis Sentimen
WIDI WIDAYAT, Teguh Bharata Adji, S.T., M.T., M.Eng., Ph.D. ; Widyawan, S.T., M.Sc., Ph.D.
Universitas Gadjah Mada, 2019 | Diunduh dari <http://etd.repository.ugm.ac.id/>