

INTISARI

DETEKSI EMOSI PADA TWEET DENGAN MENGGABUNGKAN CONTEXTUALIZED WORD EMBEDDING DAN CONVOLUTIONAL NEURAL NETWORK (CNN)

Oleh

Muhammad Fikri Heldiansyah

17/409436/PA/17743

Twitter merupakan salah satu sosial media dimana menjadi tempat para penggunanya untuk saling berbagi informasi melalui teks yang memiliki batasan 280 karakter yang disebut *tweet*. *Tweet* biasanya juga memiliki emosi para penggunanya dan sudah banyak penelitian dalam mendeteksi emosi pada *tweet* seperti contohnya dengan menggunakan kombinasi *deep learning* seperti CNN dengan *word embedding* seperti *Word2vec*, namun penelitian terkait deteksi emosi yang menggunakan *deep learning* dan *contextualized word embedding* pada *tweet* berbahasa Indonesia masih sangat terbatas.

Pada penelitian ini dilakukan deteksi emosi pada *tweet* berbahasa Indonesia dengan menggabungkan model *Convolutional Neural Network*(CNN) dan *contextualized word embedding* seperti BERT dan ELMo serta *traditional word embedding* *Word2vec* sebagai ekstraksi fitur. Terdapat lima emosi yang akan dideteksi yaitu emosi marah (*anger*), cinta (*love*), takut (*fear*), bahagia (*happy*), dan sedih (*sadness*). Pengujian deteksi emosi pada ketiga model yaitu; BERT-CNN, ELMo-CNN, dan *Word2vec*-CNN, memberikan hasil paling baik pada model BERT-CNN dengan nilai *macro-averaged precision* sebesar 75,40, *macro-average recall* sebesar 71,62, dan *macro-averaged f1-score* sebesar 72,83. Deteksi emosi yang dilakukan pada data dengan *stemming* menunjukkan bahwa model *Word2vec*-CNN dan BERT-CNN tidak lebih baik daripada model yang menggunakan data tanpa *stemming*, sedangkan pada ELMo-CNN menunjukkan *macro-averaged f1-score* yang lebih baik pada data dengan *stemming*.

Kata kunci : Deteksi emosi, *tweet*, *Convolutional Neural Network*, *contextualized word embedding*, BERT, ELMo, *Word2vec*.

ABSTRACT

EMOTION DETECTION ON TWEET BY COMBINED CONTEXTUALIZED WORD EMBEDDING AND CONVOLUTIONAL NEURAL NETWORK (CNN)

By

Muhammad Fikri Heldiansyah

17/409436/PA/17743

Twitter is one of the social media which is a place for users to share information with each other through text that has a limit of 280 characters called tweets. Tweets usually also have the emotions of their users and there have been many studies in detecting emotions in tweets, for example by using a combination of deep learning such as CNN with word embedding such as Word2vec, but research related to emotion detection uses deep learning and contextualized word embedding in tweets with Indonesian language is still very limited.

In this study, emotion detection in Indonesian-language tweets be carried out by combining Convolutional Neural Network (CNN) models and contextualized word embedding such as BERT and ELMo as well as traditional word embedding Word2vec as feature extraction. There are five emotions that will be detected, namely anger, love, fear, happy, and sadness. Testing of emotion detection on the three models, namely; BERT-CNN, ELMo-CNN, and Word2vec-CNN, gave the best results on the BERT-CNN model with a macro-averaged precision value of 75.40, a macro-average recall of 71.62, and a macro-averaged f1-score of 72.83. The emotion detection performed on stemming data shows that the Word2vec-CNN and BERT-CNN models are not better than the model using data without stemming, while the ELMo-CNN shows a better macro-averaged f1-score on the stemming data.

Kata kunci : Emotion Detection, *tweet*, *Convolutional Neural Network*, *contextualized word embedding*, BERT, ELMo, *Word2vec*.