

ABSTRACT

WEAKLY SUPERVISED EMOTION CLASSIFICATION FOR INDONESIAN TWEETS

By

Pradipta Rakahartyanto

19/438455/PA/18913

Twitter is a very popular platform to gather various types of responses. With its data being used for multiple research despite it needing to be handled in multiple ways in order to train the data. Most tweets also contain the emotion of the users, which has been used many times by many researchers for experiments. But the amount of research relating to emotional classification in Bahasa Indonesia, especially using weak supervision is still minimal.

In this study, emotion classification in Indonesia tweets will be conducted, by using weak supervision to label the datasets such as Rubrix, and modeling them using Long short-term memory (LSTM) with GloVe and Support Vector Machine (SVM) with CountVectorizer as a feature extraction. Five emotions will be used as labels, namely anger, love, happy, fear, and sadness. Testing will be done by using 2 models to represent each machine learning and deep learning methods, namely GloVe-LSTM and SVM. The GloVe-LSTM with the combined dataset, gave out the best result with an accuracy of 81.9% and a validated accuracy of 51%, with a macro-average precision of 54%, a macro-average recall of 53% and a macro-average f1-score of 53%. And for SVM it gave out the best result using the combined weakly supervised dataset together with the manually labeled dataset, with an accuracy of 61%, a macro-average precision of 64%, a macro-average recall of 63%, and a macro-average f1-score of 63%. This shows that weak supervision is a tool that can be used for emotion classification, with SVM benefitting from it the most.

Key words : Emotion Classification, tweet, Weak Supervision, Long short-term memory, Rubrix, Support Vector Machine, GloVe, CountVectorizer