

Abstract

Preprocessing is the initial stage in process input data that has not been structured into structured data before the main process is carried out such as classification or sentiment analysis. The pre-processing process is one of the important processes in the case of sentiment analysis. Pre-processing needs to be done so that the sentiment analysis results are more accurate. However, in several studies, the selection of pre-processing methods was not considered so that the results obtained were not optimal. In this study sentiment classification was conducted on product opinion cases taken from conversations in social media twitter by doing 7 combinations of pre-processing methods. The methods used are casefolding, expressive lengthening, emoticons handling, removing URLs, slang handling, punctuations handling, stopwords removal and stemming. The results obtained are a combination of 5 methods they are Removal URL, Emoticon Handling, Case folding, Expressive Lengthening and Stemming is the most optimal method with an accuracy 70.88%.

Keywords: Sentiment analysis, pre-processing, twitter