Abstract

Text is easier to understand because it uses the language used by humans. Even so, sometimes some texts also have their own meanings that need to be studied more deeply. Sentiment is a view based on certain feelings towards something. We may often encounter sentiment texts when we are in cyberspace, one of which is when we are looking for review information on something. An easy method to process this information is called sentiment analysis. Sentiment analysis is a method for classifying sentiments into positive, negative, or neutral meanings. With a method that helps process information, it makes it easier for us to absorb and get the information we are looking for. This study aims to implement a N-gram feature selection method to increase the accuracy of the model built based on research that has been done by R. Maulana (2020), which shows an increase in the value of accuracy on the model using the feature selection process. The research was conducted by running three differents scenario for testing whether the TF-IDF feature extraction method has an influence on the accuracy value and comparing the stemming method and lemmatization method on preprocessing. In this study, a dataset in the form of an English movie review text was used and uses the support vector machine (SVM) classification method. Based the study from those scenarios, we get the highest accuracy value with result 88,10 % through the TF-IDF feature extraction and N-Gram Unigram + Bigram parameters scenario.

Keywords: Sentiment, SVM, N-gram