ABSTRACT

Sentiment analysis is a way to extract emotion from a text. The purpose of this sentiment analysis is to determine the positive or negative sentiment in a tweet from Twitter regarding the 2020 American Presidential election. One way to determine this is to classify the text. By doing text classification, we can predict the sentiment of a tweet. However, there is a problem that is the number of attributes a text has. Therefore, feature selection is carried out using the TFIDF (Term Frequency - Inverse Document Frequency) method. TF-IDF is a technique of weighting a word in a document. In this study, the researchers tried to compare 2 classification algorithms, namely naïve Bayes and support vector machines. The evaluation result of the machine learning using cross-validation with the value of K is 10 and using mean approach shows that the model reaches the highest accuracy with 82% using linear kernel. Based on 1000 data tweet about Donald Trump, SVM model with accuracy of 82% succeeded in predicting 3688 neutral label or 36.88% of the people think neutral about Trump, 3078 positive label or 30.78% of the people think positive about trump, and 3234 negative label or 32.34% of the people think negative about Trump. Then, Based on 1000 data tweet about Joe Biden, SVM model succeeded in predicting 4239 neutral label or 42.39% of the people think neutral about Biden, 2962 positive label or 29.62% of the people think positive about Biden, and 2799 negative label or 27.99% of the people think negative about Biden.