ABSTRACT

We can find a lot of information about opinions, judgments, views on the Internet. One of them is a film review that contains an audience's opinion of a film. All responses from viewers who review a film cannot be considered as a sentiment, so a technique is needed to analyze the review so as to provide valuable information. Sentiment analysis is a subject in machine learning that aims to extract subjective information from textual reviews. Machine learning Support Vector Machine (SVM) method is often used to perform sentiment analysis. However, the large number of unnecessary words, repeated words, and many attributes cause the classification performance to provide a low level of accuracy. To reduce unnecessary features, it is necessary to implement feature selection, one of which is by using data preprocessing which is an important step for natural language processing (NLP) tasks. This converts the text into a more digestible form so that machine learning algorithms can work better. In this case we perform the following types of preprocessing. With the Imdb Dataset trained using several machine learning classification algorithms such as:

SVM : 89%, Naive Bayes : 76%, SGD Classifier : 89%, Ridge Classifier : 88%, Decision Tree : 71%, Logistic Regression : 88%, Random Forest : 85%, KNN : 81%. Based on the prediction results using several machine learning classification algorithms, there are 2 models that can be used to be applied to the web, namely the SVM Algorithm and the SGDClassifier Algorith.