

Abstract

News occurs because there is information or news related to facts that are happening to be later conveyed to the public. Along with technological developments, information dissemination is now carried out through social media, namely websites that can be accessed using a desktop or mobile media. The selection of news to be included in certain categories if done by humans can cause human error, especially news that is used very much can cause inefficiency. Therefore, an automatic classification system will be the solution to this problem. In classification, feature extraction is a basic process in categorization that is important to do and know. The feature will then be represented in vector form, the vector value is obtained from the weighting of terms (words). This study compares the weighting of Term Frequency - Inverse Document Frequency (TF.IDF) and Term Frequency Absolute (TF.ABS) combined with the unigram extraction feature with the Support Vector Machine (SVM) classification method. From the results of the study, the weighting of the TF-IDF got an accuracy of 96.63% with the result that the f1-score got 97.06%. Meanwhile, the TF-ABS weighting has an accuracy of 89.66% with an f1-score of 96.63%. Using TF-IDF weighting can increase accuracy by 6.97%. instead of using TF-ABS.

Keywords : News, Classification, Support Vector Machine, Feature Extraction, N-gram, Kernel