Abstract

Twitter is one of the social media, which has a lot of users, with the amount of money Twiter earns everyday, and it could be classified as a trending topic depicting the current warming news. But there is a problem because many of the tweets are using a hashtag on a warm hashtag with a different tweet content from the hashtag intended, it's allowing users to read the content from each tweet. The research aims to address the problem by dividing into categories on Twitter. This research is using technology classification Support Vector Machine (SVM), did in addition to sustaining Term Frequency Inverse Document Frequency (TF-IDF). Trial results gain a highest accuracy of 81.4% by using dataset scenario and ratio 90:10, and states that the highest accuracy is 81.47% at ratio 90:10, with an n-gram scenario at unigram + bigram + trigram.