Abstract

Information is one of the most important things for humans to get. The rapid development of technology makes the process of finding information easier to obtain. Social media is one of the media that makes it easy to get information nowadays. Therefore, today's social media cannot be separated from human daily life. This makes social media users very quickly get information that is very widely spread on the internet. Twitter is a computer-based social media platform for online communication. The social media Twitter has about 1.3 billion accounts and there are around 336 million active users which includes about 500 million tweets per day. This paper uses the k-Nearest Neighbor classification, the author makes an analysis of the classification of tweets on Twitter into argument or non-argument groups. The reason for choosing the k-Nearest Neighbors method is because the k-NN method does not require too specific parameters. The parameters used are not exact parameters, but are from a large number of training data. By using 12 test schemes combined with feature extraction using TF-IDF and Bag of Words as well as distance measures on k-NN namely Euclidean, Manhattan, and Minkowski. The best model obtained in this study is a scheme that uses TF-IDF feature extraction and distance measurement using Euclidean with a value of n = 3 (three), the scheme uses TF-IDF feature extraction and distance measurement using Manhattan with a value of n = 3 (three), and the last is the scheme using TF-IDF feature extraction and distance measurement using Manhattan with a value of n = 4 (four). The results of the three schemes obtained the value of accuracy with a value of 91.5%, precision with a value of 85.7%, recall with a value of 27.3%, and finally F1-Score with a value of 41.4%.

Keywords : ***k-nearest neighbor, twitter, argument, classification, social media***