

Abstract

Information spread through the internet is widely used by people to find anything. One of the most searched information on the internet is information related to Islamic religious knowledge. However, the large amount of information available from various sources makes it difficult for people to find the right information. Research related to this topic has been conducted by previous researchers, but the dataset used only comes from one source. Therefore, in this study a classification system for Islamic question and answer topics was built using the Naïve Bayes and TF-IDF methods. In this study using 1000 question and answer article data taken from Islamic consultation websites, namely rumahfiqih.com and islamqa.info. The multi-class classification uses five categories which are manually-labeled using the category classes on the website. From several test scenarios in this study, the Naïve Bayes classification method using TF-IDF (n-gram level) with a maximum feature of 1000 at a data separation ratio of 70:30 produces the highest accuracy of 81%. The 81% accuracy value was also generated by the SVM classification method, but the difference was in the SVM the highest accuracy value using TF-IDF (word level). It is expected that in the next research will be used more website sources and the use of other classification and feature extraction methods with more optimal value than previous research.

Keywords: question and answer, topic classification, multi-class, naïve bayes, tf-idf