

Abstract

Hadith is a way of life that contains the actions or words of the Prophet Muhammad SAW, so that many Muslims want to learn it. The problem in studying hadith is that the number of hadiths is very large and has a category for each hadith. So we need a system that is able to classify the corpus of hadith, so that hadith is easy to learn according to its class. Therefore, a system is built that is able to classify hadith into recommendations, prohibitions, and information classes by using a combination of preprocessing methods; word weighting using TF-IDF; Information Gain feature selection; Multinomial Naïve Bayes classification; and evaluation of the Confusion Matrix. The results of the research on the system showed that the best performance was obtained using preprocessing without stopword removal techniques and without the application of Information Gain feature selection with an accuracy of 88%. The combination without using stopword removal results in higher accuracy because the model can have more features that can be classified.

Keywords: hadith, *information gain*, *multinomial*, *multi label*