

Abstract

Hadith is the second source of law and guidance for Muslims after the Qur'an. There are many hadiths that have been narrated, but Bukhari's hadith has the highest level of validity according to the Islamic scholar. Along with the development of technology, hadith is very easy to obtain through the digital world. However, learning hadith is not as easy as we imagine. The number of hadiths that exist and also not have been categorized make learning hadith with specific categories very difficult to do. Therefore, the author conducted a research about classification of suggestion, prohibition and information on Bukhari's hadith that was translated to the Indonesian language which is expected to facilitate the public in learning hadith easily. The classification process uses unigram/bigram model with Mutual Information (MI) as feature selection and Support Vector Machine (SVM) as classification method. This research used several scenarios of testing by modifying the term model, preprocessing, selection feature and some kind of classification method to prove that SVM is one of the text classification method that suitable to use. Testing using a unigram model, without stopword removal or stemming, using MI and also SVM gives the best hamming loss value, it is 0.0686. The research also indicates that SVM with MI produces higher accuracy than other text classification methods.

Keyword: Bukhari's Hadith, Classification, Hamming loss, Mutual Information, Preprocessing, Support Vector Machine.