

Abstract

Hadith is a source of law for Muslims after the al-qur'an, in which there are instructions in the form of words, actions, attitudes, and others. Hadith must be studied and practiced by Muslims, then used as a way of life after the al-qur'an. Classifying hadith is a way to make it easier for Muslims to learn hadith by looking at the text pattern in the translation of Bukhari hadith based on three classes or categories based on suggestions, prohibitions, and information. The classification carried out is a multi-label classification. The classification process uses N-gram and TF-IDF as feature extraction, CART and bagging as classification methods, and hamming loss as evaluation methods. Bagging is used to cover the shortcomings of CART, namely, the CART model is less stable, which, if there is a slight change in the training data, will have a significant effect on the resulting learning model. Several testing methods were carried out to obtain the best hamming loss value in this study. Based on several tests that have been carried out, the best hamming loss value is 0.1914 or 80.86%. These results indicate that the use of bagging can help increase accuracy by 5%.

Keyword : Classification, hadith Bukhari, preprocessing, feature extraction, CART, Bagging