Abstract

Aside from the Holy Qur'an, Hadith is indeed a life guide that every Muslims in this world must follow. The technology for classifying texts and sentences, including categorizing hadiths, is evolving in tandem with the advancement of the times. The model used to perform classification has also been developed and optimized such as the use of the XGBoost algorithm which is more optimized than the previous tree algorithm. This can also make it easier for us as Muslims to study hadiths by categorizing them according to recommendations, prohibitions, and information. This study conducted text classification of Indonesian translations of hadith texts based on recommendations, prohibitions, and information using the XGBoost algorithm, TF-IDF for its feature extraction, and Chi-Square for its feature selection. In this study, experiments were carried out by changing the order of the preprocessing process for the stopword removal and stemming parts, performing the classification process with and without using chi-square as a feature selection, and adding parameter value during the modeling process with XGBoost and manage to get the final result of 75%, 75%, 74% and 74% for accuracy, precision, recall, and F1-Score.

Keywords: Text classification, Indonesian translated Hadith, XGBoost, TF-IDF, Chi-Square