ABSTRACT

Hadith is the second most fundamental source of Islamic Law after the Holy Qur'an. The Hadith needs to be categorized to ease people's understanding of the purpose of the Hadith. Indonesian-translated Sahih Al-Bukhari Hadith text was used in this study. This text is unique because it provides Hadith in two languages (Indonesian and Arabic), which exist in a single document. That problem needs specifically text processing to know the meaning of every bag of words. Text classification was used to categorize the Hadith text. The text classification in this research was conducted in two main phases: first, data preprocessing involving word embedding with Skip-Gram Negative-Sample (SGNS) algorithm was done followed by classification using the comparison of Backpropagation (BP) and Support Vector Machine (SVM). Based on the results obtained, the system was able to perform text classification well with BP, returning the highest F1-Score at 82.05%, while SVM only got the highest performance at 76.06%.