

Abstract

Members on online community forums often ask the same questions as questions that have been asked by other members. Often this question gets responses from other members. As a result, there is duplication of the same questions and answers on online community forums. The next consequence will certainly burden the system, because the system requires more data storage. A system is needed to identify the same questions so that other members do not have to answer the same questions again. In this research, a question identification system was created using the Convolutional Neural Network method with feature extraction using word2vec. There are 4 scenarios in this research with the best results in scenario 4 which uses stemmer and stopwords. The results of this scenario showed a precision of 0.834, recall of 0.761, F1-score of 0.796, and accuracy of 80.42%. This scenario is better than the testing scenario without stemming and stopwords, which yielded a precision of 0.810, recall of 0.735, F1-score of 0.771, and accuracy of 78.11%.

Keywords : question similarity, convolutional neural network, online community forums, word2vec

