ABSTRACT

MULTI-LABEL CLASSIFICATION OF HADITH TEXTS USING K-NEAREST NEIGHBOR AND INTEGRATED WITH POWERSET LABELS

By
Ian Aprilianto Said
16102196

Hadith is one of the foundations of Islam besides the Qur'an, Hadith is also the meaning of all the attitudes, words and deeds of the Prophet Muhammad in teaching Islam. Hadiths have various meanings, Hadiths which contain advice, information, and prohibitions that are intended for all Muslims. In the hadith there are several hadith texts that contain 3 meanings in 1 hadith, or it can also be called *multi-label*. With a very large number of hadith texts, at this time it really takes time to label each one hadith text, therefore a classification model is created that can group hadith texts with multiple labels. This study uses a *dataset* in the form of hadith which has 1064 texts that have been labeled accordingly. In the previous study, classification was carried out using the same *dataset* in the form of authentic bukhari hadith text with an f1-score of 0.8539 by testing using the k-nearest neighbor with binary relevance and a hamming loss result of 0.178403 by testing using a convolutional neural network with a powerset label. Basically KNN is an algorithm that focuses on pattern recognition cases, but it can also be used to carry out single-label classification, therefore in this study the powerset label method is used to classify *multi-label*s in the KNN algorithm in carrying out *multi-label* classification of the hadith texts of sahih bukhari. In this test, the best hamming loss results were obtained from using the powerset label method with a value of k = 4 and an average value of hamming loss 0.0795, and the results of using the binary relevance method with a value of k = 5 and an average value of hamming loss 0.0821, so that it can be concluded that the use of the binary relevance method is not better than the use of powerset labels in this study.

Keyword: *Hadits, KNN, Label Powerset, Binary Relevance, Multi-label Classification*