



Multilabel Classification in Indonesian Translation of Religious Text Using Word Centrality Term Weighting

Muhammad Pascal Dewantara¹, Kemas Muslim Lhaksana^{2,*}

^{1,2}School of Computing, Telkom University, Bandung, Indonesia

Email: ¹mpascal Dewantara@student.telkomuniversity.ac.id, ^{2,*}kemasmuslim@telkomuniversity.ac.id

Correspondence Author Email: ²kemasmuslim@telkomuniversity.ac.id

Abstract—This research focuses on enhancing the understanding of the Quran in the Indonesian translation dataset by employing a word centrality that feeds into a classifier model. The primary goal is to compare the hamming loss score from the TF-IDF and TW-IDF feature extraction methods in the Indonesia translation case study. The TF-IDF is commonly used in prior research. It has a higher hamming loss (which is worse in accuracy) than the TW-IDF incorporating centrality measurement more specifically in degree and closeness centrality. This research adds eigenvector centrality for a new compartment from the other methods. We used SVM, Random Forest (Bagging), and AdaBoost (Boosting) for the classifier model, with Mutual Information as the feature selection method. In evaluating the classifier, Hamming Loss is used given that the method is suitable for multilabel classification. Results indicate that the centrality measurement value for the term weighting method offers a significant improvement over regular TF-IDF. Each centrality method gives the best Hamming Loss score in each classifier model. Degree centrality gets 0.1275 in SVM, closeness centrality gets 0.1367 in AdaBoost, and eigenvector centrality gets 0.1204 in Random Forest. However, eigenvector centrality still can be a strong measurement method to lower the Hamming Loss score. Random Forest and AdaBoost give a significance better over SVM.

Keywords: Centrality; Classification; Multilabel; Quran; Weighting