ABSTRACT

Education is the main component in building quality human resources. Exams are part of the education evaluation process to measure students' ability to understand the material learned. The online exam process requires facilities regarding question management, so classification is needed to group questions according to their topics. Multi-label classification is the process of grouping data into several classes based on the similarity of data characteristics, where each question can have more than one topic. This research focuses on classifying Indonesian language subject questions at the junior high school level using the Problem Transformation method and Random Forest and K-Nearest Neighbor (K-NN) algorithms. The Problem Transformation methods used are Binary Relevance, Classifier Chain, and Label Powerset. The evaluation metric to determine the best performance is based on F1-Score with K-Fold Cross Validation. The results show that the Random Forest algorithm provides the best performance compared to K-NN with the best F1-Score value in all Problem Transformation methods. The best F1-Score value with Label Powerset method on Random Forest algorithm is 69%, and K-NN is 44%. Based on these results, the question classification model using the Random Forest algorithm and the Label Powerset method is more effective in classifying multi-label questions. It is hoped that this research can contribute to improving the efficiency of managing exam questions in online learning systems such as the Learning Management System (LMS).

Keywords— ***K-NN, machine learning, multilabel classification, problem transformation, random forest***