

ABSTRACT

Education is essentially a patterned effort to create an active learning process for students in generating potential skills and intelligence by gaining knowledge by doing practice questions. The question paper usually contains examples of questions with the level of difficulty and classification based on Revised Bloom's Taxonomy is a case that is not easy if done manually. The researcher was revealed to do automatic question classification with the help of machine learning based on the cognitive level of Revised Bloom's Taxonomy which focused on Indonesian History questions at the SMA level and C1-C3 RBT levels. The dataset was obtained from USBN questions, daily tests, and practice question banks obtained from the internet. The method used is the K-Nearest Neighbor algorithm with the Manhattan distance approach which goes through the data preprocessing stages by carrying out case folding, tokenizing, filtering and stemming processes. Furthermore, the dataset is carried out by the TF-IDF process, namely the weighting of words or terms. The SMOTE oversampling method was used in this study to overcome the imbalance of data from the datasets obtained. Then the use of K-Fold Cross Validation on the dataset with a value of $k = 10$. The results of the accuracy of the research that has been done with the KNN algorithm obtained a fairly high value with 90% precision, 87% recall, 87% F1-Score and 87% accuracy and the oversampling method. SMOTE can be implemented well in handling imbalance data.

Keywords: *Question classification, KNN, Manhattan, RBT, SMOTE, oversampling.*