ABSTRACT

Education is the process of changing one's attitudes and behavior. The evaluation used by teachers in our country is in the form of tests such as tests or exams. This study reveals how to make machine learning to classify Indonesian history questions at the cognitive level of high school level Revised bloom's taxonomy on difficulty levels C1 to C3 with the Naive Bayes algorithm. Labeling is done manually to determine questions based on the cognitive level of RBT. To get the accuracy of a machine learning in this study, there are several stages, namely the preprocessing stage where the dataset will be filtered using case folding, tokenizing, filtering, and stemming then the dataset will be weighted with TF-IDF. The researcher used the SMOTE oversampling method to overcome the data imbalance, then tested using K-fold with a total of 10 folds and finally the model was evaluated for performance using a confusion matrix. The classification results obtained that the average score of the K-fold dataset increased by 16% (60% - 76%) after SMOTE was carried out, then the accuracy of the performance evaluation also increased by 21% (61% - 82%) when SMOTE was performed. The results obtained from the implementation of the K-Fold Cross Validation and confusion matrix show that the use of the Naïve Bayes algorithm shows a good accuracy score, and the use of the oversampling method is very helpful in this study to overcome the data imbalance.

Keyword—**question classification, Naïve Bayes, High school Indonesian history, SMOTE, RBT, oversampling, High school questions**