ABSTRACT

Education at the formal level requires aspects of testing to ensure science is absorbed by students. In an effort to ensure students understand the material taught, teachers hold competency tests that are divided into several terms each semester. Competency tests consist of various types of questions, one of which is multiple choice. The questions collected by teachers in the e-learning platform require sorting so that they can measure the intellectual level of students and include thinking skills.

Evaluation in sorting questions found many unfiltered questions in accordance with the achievement of competence. In this study, the screening process involving Revised Bloom's Taxonomy (RBT) algorithm will be implemented automatically using one of the machine learning methods, Stochastic Gradient Descent (SGD) on the history of the high school level. Due to the imbalance in the dataset, this study will also compare between datasets that apply and those that do not apply the SMOTE oversampling method.

The classification results from this study were obtained through the implementation of the SGD algorithm with python tools. In the dataset without SMOTE, the SGD algorithm has an average K-Fold score of 62%. In the dataset with SMOTE, the SGD algorithm has an average score of 93%. The accuracy results from the confusion matrix show that the SGD algorithm on the dataset without SMOTE has 62% performance, while the SGD algorithm on the dataset with SMOTE gets 94% performance.

The results obtained through the K-Fold Cross Validation process and the confusion matrix show that the data set using oversampling has better results than the data set without using oversampling.

*Keywords—**question classification, RBT, high school history, SGD, SMOTE***