ABSTRACT

In today's digital era, the utilization of technology in education is important to support the learning process. The use of an automated system in question classification helps reduce the time and effort required to group questions based on topics, considering that it will be difficult and consume a lot of time if done manually. This research develops a classification model for junior high school mathematics questions using the Naïve Bayes algorithm because of its simplicity and ability to process data. It is then enhanced with the SMOTE method to overcome data imbalance. The purpose of this research is to evaluate the performance of the Naïve Bayes model in the classification of Mathematics questions and identify the effect of using Mathematical symbols on classification results. The results showed that the resulting model achieved 69% accuracy with the application of SMOTE. Evaluation using cross validation showed that classification without math symbols achieved 89.35% accuracy, while with math symbols, it reached 88.79% when using SMOTE. Although the difference in cross validation results is only 0.56%, it shows that either with or without math symbols, the Naïve Bayes model can produce relatively equivalent performance with 89% accuracy in classifying Mathematics questions. This study shows that the use of the SMOTE method is effective in overcoming the problem of data imbalance. In contrast, the use of mathematical symbols has a relatively small impact on model performance.

Keywords— classification, cross validation, naïve bayes, SMOTE, symbols