

ABSTRACT

Questions classification based on topic questions is a way of grouping questions based on topic categories, which aims to measure and determine questions to fit the indicators. Questions classification helps students and teachers determine questions based on the topic category. In this study, the researcher intends to create a classification model for Biology questions for grade 11 high school, which is grouped into nine topic categories, namely Cells, Plant and Animal Tissues, Human Movement systems, Circulatory systems, Digestive systems, Respiratory systems, Excretion System, Coordination System, and Human Reproductive System. The questions and topics were obtained from the "Siap Pintar Belajar Mandiri" question bank book.

This study compares the value of accuracy and performance evaluation of two classification algorithms, Naive Bayes and C4.5. To evaluate the performance, the researcher uses cross validation and looks for precision, recall, and f1-score values using a confusion matrix.

From the classification results, the results of the Naive Bayes accuracy are 72.72%, and the accuracy value of performance evaluation using cross validation is 73.09% and the precision value is 73%, recall is 73%, and the F1-Score is 70%. While C4.5 got an accuracy value of 54.54%, and the accuracy value of performance evaluation using cross validation was 54.09% and the precision value was 58%, recall was 56%, and F1-Score was 55%.

Keywords— *Questions Classification, Biology, Naive Bayes, C4.5, Cross Validation*