Abstract

Determination of the class quota was often been quite confusing to be done on each time of IT Telkom's registration activities. One of the main factors that affecting the difficulty in determining class quota is its difficult to predict the number of students who will repeat the certain courses. In this final project, will be created a system that will be able to assist in the prediction of the number of students who repeat some courses. The method that will be used is a decision tree C4.5 method, a classification method which will produce a decision tree that can be converted into a set of rules that can later be used as a basis for prediction of repeat students. However, sometimes decision tree methods had a overfit condition, where a decision tree does not have good accuracy for data to be entered into the system. Therefore, there will be pruning, a method to cut the tree into simpler decisions in order to improve the accuracy, by using Reduced Error Pruning algorithm. Then, there will be comparisons between the decision tree that haven't pruning process and the decision tree that have been in pruning in terms of accuracy produced, number of rules are formed, as well as the time required to classify the data to see the effect of pruning on the resulting decision tree.

Keywords: prediction of repeated courses, C4.5 decision tree, overfit, pruning, Reduced Error Pruning