

Abstract

World Wide Web or WWW is one of the technological phenomenon that is growing very rapidly at this time. WWW provides a range of information services about news, advertising, education, e-commerce and etc. The information available on the WWW has a very large size and distributed globally in the world. Web also contains a wealth of information seen from the structure and *web usage*. Web is a collection of data and information that is potentially to mining in order to generate knowledge that can be useful for the community as well as certain parties.

C5.0 algorithm is an algorithm to generate *decision tree* to classify. Selection of attributes that will be processed using a measure of *information gain*. Size *information gain* is used to select the test attribute at each node in the tree. Attributes with highest *information gain* value will be selected as the parent for the next node. These algorithms form the *decision tree* by way of division and mastering the sample recursively from top to bottom. The users get interpreting information on the *C5.0* classification results are presented in two forms, using a *decision tree* and a set of IF-T HEN rules that are easier to understand.

Based on the results of the analysis has been done can be seen that accuracy for the tree that are generated from training data produces poor accuracy on testing data, therefore a process of *pruning*. Rule generated after *pruning* process has better accuracy on testing data and have the low simplicity of rules, so that the resulting rules are much simpler than the previous tree. Access patterns of users who have been classified not give a significant difference it is because the user accessed the web has an interest in almost the same information. From the results of this classification can be obtained feedback on the web admin for increased performance in terms of web navigation.

Keywords: *web usage, C5.0, decision tree, information gain, pruning*