

Abstract

Data mining is a data analyze process to find a pattern and rule using software, can analyze big quantities data to information in a pattern that have meanings to decision support. Classification is one of the technique of data mining. Classification purpose to build a model that can differ data class, with an aim to guess class from an unknown label object.

In this final assignment, a software is made to implement a classification method known as decision tree with joining the BOAT and C4.5 algorithm. BOAT uses bootstrap to help fasten the decision tree form process. Bootstrap preferences include the numbag size and bagsize that is used. Bagsize is data sample size used to form decision tree, and numbag is the tree quantity formed from the data sample. Trees that are formed then is being combined until it gets the final tree.

The testing result shows that BOATC4.5 has a better time performance than C45 with a small number of bagsize and numbag, has the same form of tree for data without noise. For data's with an outlier type of noise, the tree form is different but the accuracy value of BOATC4.5 is not lower than C4.5 accuracy.

Keyword: *data mining, decision tree, C4.5, BOAT, bagsize, numbag*