

Abstract

With the large number of categories that can be owned by a data text, manual categorization will be a new problem for the user. With data text that have been classified, it will allow searcher to find information easily in data text cause data text have been grouped according to categories that reflect the contents of the data text. Nowadays, most of the data text have to be a lot of categories in one data or multi-label.

Decision tree is one of the most popular classification method because it is easy to interpreted by humans. Tree generated from a data modeling will form a rules or the rules "if - then".

One of the most popular tree algorithm is C4.5 algorithm. C4.5 algorithm is widely known and used for data classification with numerical attributes and categorical. Multi-label data is a problem in the classification process with the C4.5 algorithm, therefore it will be classifier modification in C4.5 algorithm where is used to treat single-label classification process so that it can perform a multi-label data text classification process. Modifications made is to modify the formula to calculate the value of entropy.

The result show that entropy modification in C4.5 algorithm for multi-label data text can be used. Beside that, it has good performance by looked from accuracy and time evaluation.

Keywords: *classifier* modification, entropy, data text, multi-label, performance