

Abstract

Preprocessing data is a process to cleaning data from noise, outlier, missing value, irrelevant feature, redundant feature etc, so that the data can be more efficient when the next mining process implemented. One of preprocessing techniques is by reduction the dimension of data. Data reduction method that can be implemented is feature selection (choosing only the important features). In this TA, feature selection method that being implemented is MMIMRSC (Maintaining mutual information and minimizing redundancy-synergy coefficient) that purpose to discard irrelevant and redundant feature by calculating the information value and also decreasing the redundant feature by calculating redundancy-synergy coefficient. The performance of reduces data after preprocessing shows that precision and recall value increase and time to build classification model become faster than unreduced data.

Key words: *feature selection, mutual information, data reduction, preprocessing*