Abstract

Data mining is a process of extracting or mining knowledge from large amount of data. There are three tasks in data mining, they are *classification*, association, and clusterization. A classification in data mining is a task to find the pattern of a data group. The aim of finding the pattern is to get the data value which still not known. But in many times it needs a huge resource to do classification. It is because the dataset which to be classified has high dimensionality ,many irrelevant or redundant variables.

Therefore a *pre-proccessing* task is needed before the classification proses. There are many pre-proccessing steps, one of them is Variable Selection. Variable selection is a proccess of identifying and deleting variables which are irrelevant or redundant.

This final assignment spesifically studying about variable selection using Correlation-based method. In selecting variable, Correlation-based method do counting and comparing correlation level between variable and its class variable or variables with other variables . Then the test is done by comparing the accuration, precision and recall value between original dataset and the Correlation-based dataset. From the test result is known that variable selection using correlation-based method can reduce data dimensionality with accuration, precision and recall value close to original dataset.

Keywords : C*lassification, pre-proccessing, Variable Selection, Correlation based*