ABSTRACT

The task of extracting knowledge from databases is quite often performed by machine learning algorithms. The majority of these algorithms can be applied only to data described by discrete numerical or nominal attributes (features).

Discretization is a process to transform a continuous attribute's value into a finite number of intervals and associate with each interval a numerical, discrete value. For mixed-mode (*continuous* and discrete) data, discretization is usually performed prior to the learning process, called pre-processing.

CAIM (Class-Attribute Interdependence Maximization) is one of discretization algorithm design for supervised learning. It maximizes the class-attribute interdependence and to generate a possibly minimal number of discrete intervals. The algorithm does not require the user to predefine the number of intervals. It considered as CAIM's superiority against other discretization algorithms for supervised learning.

This final project implements CAIM discretization methode for supervised learning to several datasets. C5.o algorithm is used to generate classification rules from data discretized by CAIM. The test performed using CAIM and six other state-of-the-art discretization algorithms show that the accuracy of generated rules is – on average - higher and the number of rules is lower for data discretized by CAIM when compared to data discretized using six other discretization algorithms.

Key word : *CAIM, class-attribute interdependence maximization, discretization*