USING DATA MINING METHODS
FOR DEMOGRAPHIC SURVEY DATA PROCESSING

Abstract

The goal of the thesis was to describe and demonstrate principles of the process of knowledge discovery in databases - data mining (DM). In the theoretical part of the thesis, selected methods for data mining processes are described as well as basic principles of those DM techniques. In the second part of the thesis a DM task is realized in accordance to CRISP-DM methodology. Practical part of the thesis is divided into two parts and data from the survey of American Community Survey served as the basic data for the practical part of the thesis. First part contains a classification task which goal was to determinate whether the selected DM techniques can be used to solve missing data in the surveys. The success rate of classifications and following data value prediction in selected attributes was in 55–80 % range. The second part of the practical part of the thesis was then focused of determining knowledge of interest using associating rules and the GUHA method.