

HW3 Association Rules

For this homework, we are going to explore the bank data, available on the LMS, and an accompanying description of the attributes and their values. The dataset contains attributes on each person's demographics and banking information in order to determine they will want to obtain the new PEP (Personal Equity Plan).

Your goal is to perform Association Rule discovery on the dataset using Weka.

First perform the necessary preprocessing steps required for association rule mining, specifically the id field needs to be removed and a number of numeric fields need discretization or otherwise converted to nominal.

Next perform association rule discovery on the preprocessed data. Experiment with different parameters and preprocessing so that you get on the order of 20-30 strong rules, e.g. rules with high lift and confidence which at the same time have relatively good support. Don't forget to report in details what you have tried.

Finally, set PEP as the right hand side of the rules, and see what rules are generated.

Select the top 5 most "interesting" rules and for each specify the following:

- Support, Confidence and Lift values
- An explanation of the pattern and why you believe it is interesting based on the business objectives of the company.
- Any recommendations based on the discovered rule that might help the company to better understand behavior of its customers or to develop a business opportunity.

Note that the top 5 most interesting rules are most likely not the top 5 in the strong rules. They are rules, that in addition to having high lift and confidence, also provide some non-trivial, actionable knowledge based on underlying business objectives.

To complete this assignment, write a short report describing your association rule mining process and the resulting 5 interesting rules, each with their three items of explanation and recommendations. For at least one of the rules, discuss the support, confidence and lift values and how they are interpreted in this data set.

You should write the report as if you are working for a client who knows little about data mining. Your report should give your client some insightful and reliable suggestions on what kinds of potential buyers your client should contact, and convince your client that your suggestions are reliable based on the evidence gathered from your experiment results.

In more detail, your report should include:

- Description of preprocessing steps
- Description of parameters and experiments in order to obtain strong rules
- Give the top 5 most interesting rules and the 3 items listed above for each rule
- For one rule, discuss the support, confidence and lift numbers and how they were computed from the data