Nick Videtti

nvidetti@syr.edul

Assignment 3

Association Rule Mining

Data Discretization, Rule Mining, and Analysis

7/30/2022

| To Contact, or Not To Contact? | |
| --- | --- |
| **Introduction** | The world has become more and more “paperless” over the last two decades, and for good reason. At a local financial firm, they are looking to do the same for their marketing promotions. After sending out physical pieces of mail for advertising Personal Equity Plans (PEPs), the marketing department now has data on the results of those promotions.  The marketing department can use those results to send mass emails to specific customers based on customer demographics. The results can also help determine whether or not to pursue leads that are given to them from the sales department. This way, paper is saved, as well as time and resources, which can be allocated to targeting only the “best” leads.  All that said, who are the “best” leads, and how can future leads be detected as one of the “best” leads? Luckily, the marketing department has supplied the results of their physical mail promotions. With a big PEP marketing promotion coming up soon, the financial firm is directing the marketing department to only target the best leads and the current customers most likely to accept a PEP. How should the marketing department approach this? |
| **Analysis** | The following metrics have been provided:    Using these, Association Rule Mining will be performed, using the Apriori algorithm. Since the goal is to find what characteristics correlate to customers that have bought a PEP, only rules with a rhs (right hand side) of “pep = YES” will be used. In order to find useful rules among these, only those with a lift greater than 1 will be used.  The first step before this is to make sure all variables are categorical, and that R stores them as a “factor” type. This is how the variables are structured from the raw data after pulling them into R.    No variables seem to be a factor, but that can be handled using R’s as.factor() function. Two variables will need a bit more, however. Since income and age have far too many values to be useful as factors, they will need to first be binned into groups based on ranges of values. One method for doing that is creating a new column in the data and using a “for loop” to assign character values based on the numerical value. After doing that for age and income, changing the age groups and income groups to factors, and removing the original numerical columns for age and income, the structure of the variables are now as follows.    Now the data are finally ready for the Association Rule Mining for rhs of “pep = YES” and lift > 1. |
| **results** | The results of the Association Rule Mining were as follows. The rules were sorted by lift in descending order, but also seem to have the same effect when it comes to confidence.    These rules are all high confidence and low support, indicating that the lhs and rhs do not occur very often in the data, but when they do, they occur together a large majority of the time. These will give a small amount of customers with a very high potential for buying a PEP, which is exactly what will help the marketing department accomplish their goal.  Translating these 6 rules into some results more easily usable by the marketing department, the following can be said about the “best” leads/customers.   1. Has a current account, a savings account, and exactly one child 2. Has no mortgage and does have exactly one child 3. Has a savings account and exactly one child 4. Has a current account and exactly one child 5. Is married and has exactly one child 6. Has exactly one child   Since having exactly one child not only appears in all of the rules, but is one of the rules themselves, it may be appropriate to simplify things and reach out to all customers/leads with exactly one child. |
| **conclusions** | After using the data provided by the marketing team, it was found that the “best” lead/customer is likely to be somebody that has a current account, has a savings account, and has exactly one child. Another significant finding was that the 6th “best” customer/lead simply has exactly one child. As a matter of fact, having exactly one child is part of the criteria for all of the top 6 “best” customers/leads.  Given this information, the marketing department should see how many potential customers there are with exactly one child. If there is a reasonable number based on their own expertise, they could simply start with that approach. If there are too many to contact when only using that one criterion, they could expand that to any of the following, in this exact order. Keep in mind these are sub-groups of the group with exactly one child.   1. Has a current account and a savings account 2. Has no mortgage 3. Has a savings account 4. Has a current account 5. Is married   It is important to remember that these are based on the results from previous marketing promotions. It is paramount to repeat the process that was done in this paper after every marketing promotion, as this will improve the accuracy of predictions each time. |