Forrest Fallon  
Bank Data Analysis  
IST 707  
10/25/21

In order to properly train our data to agree with the parameters of associative rule mining, a few quick changes to data classes and value-representation are necessary. We begin this process by importing the data, checking for any null values, and separating integer value columns (age, income) into “buckets” that the rule mining programs can better work with. After that is complete, we can then double check that every column value is considered a factor, along with changing our yes/no answers to a full label of what column they exist within, along with the answer to the column (“Mortgage – Yes”).

Graphical user interface, text, application, chat or text message

Description automatically generated

Next, we will begin the process of honing our Apriori rules software; the assignment asks for “20-30 strong rules,” so in this example we will aim for the median of that boundary. Keep in mind that the rules needing to be “strong” is shorthand for having respectable values within the parameters known as support (proportion of transactions of which an itemset appears), confidence (how likely item 2 is included given the existence of item 1), and lift (same as confidence, however, also considers the support of item 2).

Ideally, we want to keep the support above 0.25, meaning that each item that is considered for a rule should appear at least 25% of the time within the rule searching. Anything lower would likely be too “vague” for the program and would yield a plethora of rules that are too infrequent for any meaningful analysis. The confidence parameter can also be tricky; too high of a value can yield very few results, and too low of value is not worth considering in any further analysis. For this exercise we will be sticking to the “sweet-spot” of 0.7.   
  
 Let’s see how these parameter choices will play against our assignment’s preferred amount of rules:   
  
  
  
This was a good first attempt! 42 rules is too many for this exercise, let’s adjust:  
  
  
  
Closer! Let’s try one more adjustment:  
  
  
  
Perfect! We are exactly within the requirement of the assignment and can proceed with the rules analysis.

Now that our ruleset is established, let’s take a look at the first 10 entries within the list sorted by lift:  
Text

Description automatically generated  
  
 As seen above, some useful information for the bank’s decision making process can already be found. The bank’s efforts are currently focused on PEP account creation, therefore, it is a great sign that the top 4 strongest rules created involve the customers that are non-PEP.

Next, we will point the rules analysis program directly at our PEP columns, and determine which attributes are most associated (or not associated).  
  
Text

Description automatically generated

Excellent, our program has now separated the answers into Pep – Yes and Pep – No. We can further dissect these and perform further analysis:  
Text

Description automatically generated  
Text

Description automatically generated

What just took place above was the separation of our pep ruleset into the Yes’ and No’s. After that took place, we then told the program to display our results according to their respective Lift measurement. This will detail our strongest rules, thus, presenting the best data to make assumptions on at this time.

**Rules Analysis** Bank data is always a great avenue for data analysis practices; the amount of variables at play is always enough for more advanced analysis techniques, yet the outside factors of a person’s financial choices are always at play. We must look further than what each of the above rules is telling us about the numbers and consider why that rule could be happening in the first place.   
  
The first rule we will discuss is our strongest rule from the first set of rules created:  


The existence of this rule, along with its overall strength provides some food for thought regarding the PEP program. The fact that the largest group of people that do not own a PEP account are those that are married shows us that something needs changing regarding how a PEP account treats married couples. Perhaps the name of the product is enough to steer some people away, “Personal Equity Plan.” The word Personal sounds inclusive and could be deterring those that are married from considering it. Unless there is a marriage-calibrated Equity Plan in existence, the PEP should be adjusted to appeal married people.

Rule number 2:   
  
 Our Apriori settings were able to tell us that with 85% confidence, if the account holder has one child and no mortgage payment, they will have a PEP. This rule is particularly interesting because of how it potentially offers us some information as to what exactly a PEP is.

We must consider why a person with one child and no mortgage payment would consider a product involving their financial plan. To go further, we must assume that the lack of a mortgage payment means that said person is renting property instead. Anyone who has lived in an apartment complex knows that they are not always ideal for raising a child. Many people often have a change in world-view after their first child is born, and an “enlightenment” can happen which introduces many changes in how a person lives their life. This bank clearly has enough accounts of people that are tired of raising a child in a rental setting, and they subscribed to a product that should help them find a way out.

Rule number 3:  


This rule’s confidence alone is enough to catch our attention, however, the support being as low as it is teaches an important lesson in rule association. When considering a rule for analysis, all parameter values should be considered…not just the highest one. What the rule parameters above are telling us is that the three variables listed on our LHS did not appear often, however, when they did appear they were 90% likely (confidence) to not have a PEP. Our lift value here suggests a strong correlation between the LHS and RHS, more specifically, a lift greater than 1 tells us that the presence of said variables actually has a positive effect on how many times our RHS occurs.

The bank could still benefit from this finding though! A good next step would be to further analyze this grouping of people and determine their mean age. One could assume that people within this group are young and newly married, perhaps not thinking ahead far enough to warrant the need of a PEP at this time. To attract people within this grouping, the bank should consider a better marketing strategy that details how a PEP could benefit newlyweds.

Rule number 4:  


Similar to our example above, the support level of this next rule implies we should not make concrete assumptions on the entire dataset based on this one rule. Instead, consider what this rule may be a symptom of. Think back to our rule number 2 analysis, the people that have a child yet no mortgage often had a PEP; rule number 4 could be implying the results of these people subscribing to the PEP program.

Although we are currently working with a slice of past-tense data within this exercise, bank data is constantly changing. The PEP program is still running, and new accounts are converted frequently. Rule number 4 could be pointing us in the direction of checking in on our current PEP users and determining which other groupings in our data they were previously a part of. If a timeline is able to be developed for those that were once a PEP – No and now a PEP – Yes, the bank could use that information to determine where their current efforts are working best/worst.

Rule number 5:  


This rule helps to remind us to stay vigilant when considering our LHS and RHS. The above is telling us that many savings account holders are not PEP users. The support and confidence are high in this rule finding, so why the middling lift? A lift of 1 tells us that a person not having a PEP account does not have any positive or negative impact on whether they have a savings account.

The bank benefits from having this information because of the interaction (or lack thereof) between PEP’s and savings accounts is now visible. People are well aware of the fact that having a savings account is a good idea, and it seems that many of them do not need a PEP to remind them. Perhaps the bank could adjust their gameplan with those already in possession of a savings account, and detail to them how a PEP would further organize their finances.

In conclusion, this exercise does a wonderful job of highlighting the need to gather the full scope of the product being studied. The available data on the bank’s customers was enough to still make useful comparisons, however, knowing more about what exactly a PEP does would further help with suggestions to the bank. The individual rules we discussed above are still great information to be handing over to the bank at this time; it would be wise to consider this a first draft and set up another meeting to discuss how we could go further with our analysis once all details regarding the PEP program are known.