Jeff Levesque

Professor Gates

IST 565

Homework #3

**Introduction**

Data mining techniques including association rule mining, can be found useful across many fields. Some examples vary from genetic research, market basket analysis, as well as various forms of marketing and financial analysis. More generally, banking institutions can use mining techniques, by attempting to predict whether a customer would buy a bond, IRA, PEP, using historical data from previous and current customers.

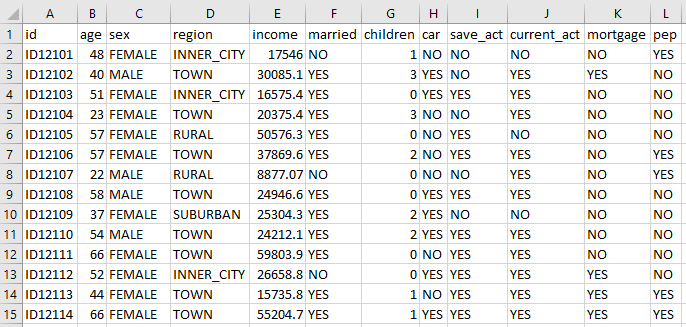
Association rule mining could additionally be implemented against attributes of customers in a banking institution. Using the apriori technique, some customer attributes can be determined to be associated with another, along with indicative measures of how good these rules are. For example, *support* is a fractional measure of how frequent a rule occurs, while *confidence* is the probability that the rule is correct. On the other hand, *lift*, indicates whether the given rule exceeds the anticipated confidence level. When the lift is larger than 1.0, then the antecedent and consequent is more closely associated, than when the two are independent (lift <= 1.0).

Using established measures to determine how good an association rule is, allows institutions to understand which set of attributes may or may not be closely related. These facts allow necessary decision makers to properly target current, and future customers. For example, if banking institutions found that customers within a given age range, with chosen attributes more frequent in delinquency, then it would be appropriate for the banking institution to make necessary policies to counteract these limitations.

**Analysis**

Data Preparation:

A csv dataset representing banking clients, each containing 11 additional attributes:

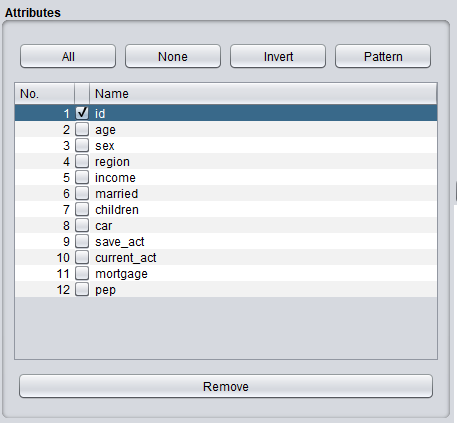


The first column represents a unique identification number for the corresponding customer. The remaining columns are indicative whether the customer possess the respective attribute. However, some attributes require some definition. For example, pep indicates whether the customer bought PEP after the last mailing. Additionally, region, represents four possible values:

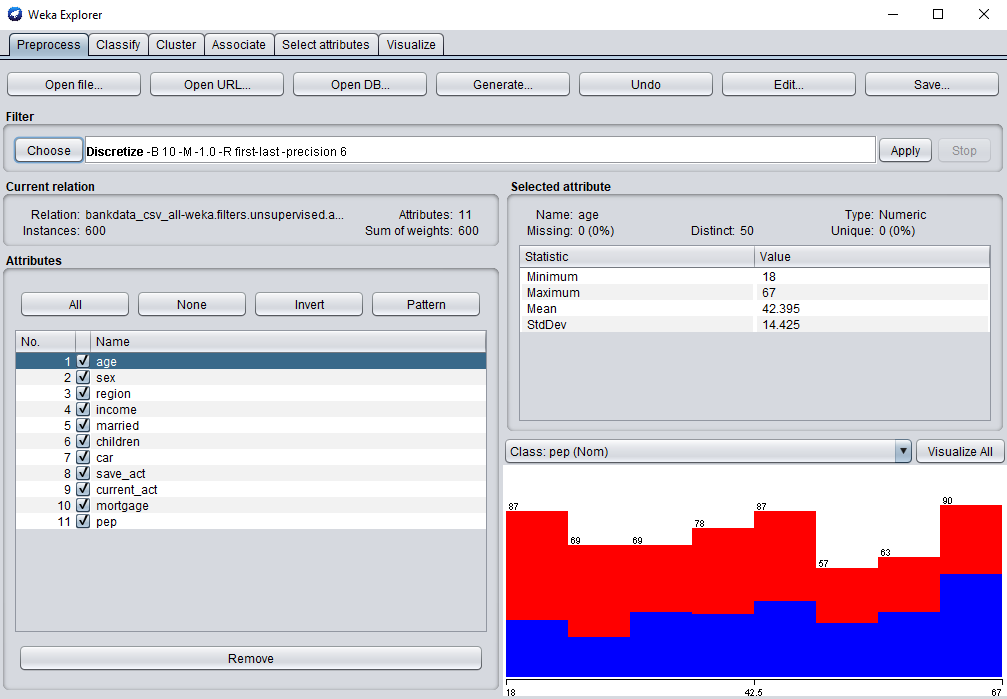
* inner\_city
* rural
* suburban
* town

Processing:

The provided dataset was processed using Weka. Specifically, after the dataset was loaded, the ID column was removed:



Then, all numerical fields were discretized by applying the unsupervised discretize filter to the remaining columns in the weka explorer:



This allowed a general apriori association rule mining to be implemented (left), as well as class association rules (right), to be configured in attempt to predict PEP:

|  |  |
| --- | --- |
|  |  |

**Results**

The general apriori association rule mining, generated a full set of 25 results:

Best rules found:

1. children='(-inf-0.3]' save\_act=YES mortgage=NO pep=NO 74 ==> married=YES 73 <conf:(0.99)> lift:(1.49) lev:(0.04) [24] conv:(12.58)

2. sex=FEMALE children='(-inf-0.3]' mortgage=NO pep=NO 64 ==> married=YES 63 <conf:(0.98)> lift:(1.49) lev:(0.03) [20] conv:(10.88)

3. children='(-inf-0.3]' current\_act=YES mortgage=NO pep=NO 82 ==> married=YES 80 <conf:(0.98)> lift:(1.48) lev:(0.04) [25] conv:(9.29)

4. children='(-inf-0.3]' mortgage=NO pep=NO 107 ==> married=YES 104 <conf:(0.97)> lift:(1.47) lev:(0.06) [33] conv:(9.1)

5. children='(-inf-0.3]' car=NO mortgage=NO pep=NO 62 ==> married=YES 60 <conf:(0.97)> lift:(1.47) lev:(0.03) [19] conv:(7.03)

6. married=YES children='(-inf-0.3]' save\_act=YES current\_act=YES 87 ==> pep=NO 80 <conf:(0.92)> lift:(1.69) lev:(0.05) [32] conv:(4.97)

7. married=YES children='(-inf-0.3]' save\_act=YES mortgage=NO 80 ==> pep=NO 73 <conf:(0.91)> lift:(1.68) lev:(0.05) [29] conv:(4.57)

8. married=YES children='(-inf-0.3]' current\_act=YES mortgage=NO 88 ==> pep=NO 80 <conf:(0.91)> lift:(1.67) lev:(0.05) [32] conv:(4.47)

9. sex=FEMALE married=YES children='(-inf-0.3]' mortgage=NO 70 ==> pep=NO 63 <conf:(0.9)> lift:(1.66) lev:(0.04) [24] conv:(4)

10. married=YES children='(-inf-0.3]' save\_act=YES 119 ==> pep=NO 107 <conf:(0.9)> lift:(1.65) lev:(0.07) [42] conv:(4.18)

11. age='(62.1-inf)' 68 ==> save\_act=YES 61 <conf:(0.9)> lift:(1.3) lev:(0.02) [14] conv:(2.63)

12. married=YES children='(-inf-0.3]' mortgage=NO 116 ==> pep=NO 104 <conf:(0.9)> lift:(1.65) lev:(0.07) [40] conv:(4.07)

13. married=YES children='(-inf-0.3]' car=NO mortgage=NO 67 ==> pep=NO 60 <conf:(0.9)> lift:(1.65) lev:(0.04) [23] conv:(3.82)

14. children='(-inf-0.3]' car=NO pep=NO 91 ==> married=YES 80 <conf:(0.88)> lift:(1.33) lev:(0.03) [19] conv:(2.58)

15. region=INNER\_CITY children='(-inf-0.3]' pep=NO 73 ==> married=YES 64 <conf:(0.88)> lift:(1.33) lev:(0.03) [15] conv:(2.48)

16. children='(-inf-0.3]' car=NO current\_act=YES pep=NO 69 ==> married=YES 60 <conf:(0.87)> lift:(1.32) lev:(0.02) [14] conv:(2.35)

17. sex=FEMALE children='(-inf-0.3]' pep=NO 90 ==> married=YES 78 <conf:(0.87)> lift:(1.31) lev:(0.03) [18] conv:(2.35)

18. car=NO mortgage=NO pep=YES 89 ==> current\_act=YES 77 <conf:(0.87)> lift:(1.14) lev:(0.02) [9] conv:(1.65)

19. children='(0.9-1.2]' save\_act=YES current\_act=YES 73 ==> pep=YES 63 <conf:(0.86)> lift:(1.89) lev:(0.05) [29] conv:(3.61)

20. sex=FEMALE children='(-inf-0.3]' current\_act=YES pep=NO 70 ==> married=YES 60 <conf:(0.86)> lift:(1.3) lev:(0.02) [13] conv:(2.16)

21. car=YES save\_act=YES mortgage=NO pep=NO 74 ==> married=YES 63 <conf:(0.85)> lift:(1.29) lev:(0.02) [14] conv:(2.1)

22. region=INNER\_CITY current\_act=YES mortgage=NO pep=NO 78 ==> married=YES 66 <conf:(0.85)> lift:(1.28) lev:(0.02) [14] conv:(2.04)

23. children='(0.9-1.2]' mortgage=NO 84 ==> pep=YES 71 <conf:(0.85)> lift:(1.85) lev:(0.05) [32] conv:(3.26)

24. save\_act=YES mortgage=NO pep=NO 142 ==> married=YES 120 <conf:(0.85)> lift:(1.28) lev:(0.04) [26] conv:(2.1)

25. sex=FEMALE married=YES children='(-inf-0.3]' current\_act=YES 71 ==> pep=NO 60 <conf:(0.85)> lift:(1.56) lev:(0.04) [21] conv:(2.7)

However, the class association rule mining, only returned 17 results:

Best rules found:

1. married=YES children='(-inf-0.3]' save\_act=YES current\_act=YES 87 ==> pep=NO 80 conf:(0.92)

2. married=YES children='(-inf-0.3]' save\_act=YES mortgage=NO 80 ==> pep=NO 73 conf:(0.91)

3. married=YES children='(-inf-0.3]' current\_act=YES mortgage=NO 88 ==> pep=NO 80 conf:(0.91)

4. sex=FEMALE married=YES children='(-inf-0.3]' mortgage=NO 70 ==> pep=NO 63 conf:(0.9)

5. married=YES children='(-inf-0.3]' save\_act=YES 119 ==> pep=NO 107 conf:(0.9)

6. married=YES children='(-inf-0.3]' mortgage=NO 116 ==> pep=NO 104 conf:(0.9)

7. married=YES children='(-inf-0.3]' car=NO mortgage=NO 67 ==> pep=NO 60 conf:(0.9)

8. children='(0.9-1.2]' save\_act=YES current\_act=YES 73 ==> pep=YES 63 conf:(0.86)

9. children='(0.9-1.2]' mortgage=NO 84 ==> pep=YES 71 conf:(0.85)

10. sex=FEMALE married=YES children='(-inf-0.3]' current\_act=YES 71 ==> pep=NO 60 conf:(0.85)

11. children='(0.9-1.2]' save\_act=YES 95 ==> pep=YES 80 conf:(0.84)

12. children='(0.9-1.2]' current\_act=YES 101 ==> pep=YES 84 conf:(0.83)

13. married=YES children='(0.9-1.2]' 89 ==> pep=YES 74 conf:(0.83)

14. sex=FEMALE married=YES children='(-inf-0.3]' 94 ==> pep=NO 78 conf:(0.83)

15. children='(0.9-1.2]' 135 ==> pep=YES 110 conf:(0.81)

16. married=YES children='(-inf-0.3]' car=NO current\_act=YES 74 ==> pep=NO 60 conf:(0.81)

17. married=YES children='(-inf-0.3]' car=NO 100 ==> pep=NO 80 conf:(0.8)

Both analysis was set to return a maximum of 25 records, with identical attribute configurations. The top 5 interesting rules were somewhat related:

married=YES children='(-inf-0.3]' save\_act=YES mortgage=NO 80 ==> pep=NO 73 <conf:(0.91)> lift:(1.68) lev:(0.05) [29] conv:(4.57)

married=YES children='(-inf-0.3]' current\_act=YES mortgage=NO 88 ==> pep=NO 80 conf:(0.91)

sex=FEMALE married=YES children='(-inf-0.3]' mortgage=NO 70 ==> pep=NO 63 <conf:(0.9)>

married=YES children='(-inf-0.3]' mortgage=NO 116 ==> pep=NO 104 conf:(0.9)

married=YES children='(-inf-0.3]' car=NO mortgage=NO 67 ==> pep=NO 60 <conf:(0.9)>

Specifically, when a customer is married, with newly born children(s), and not owning a home, it seems highly unlikely they would purchase PEP. More specifically, when interpreting the first of the above interesting rules, the *confidence* was found to be 0.91. This indicates a strong confidence that the rules are accurate. The 1.68 lift indicates the antecedent and consequence are significantly related. The support count for the left side items (i.e. antecedent) is 80, while the support count for the right-side items (i.e. consequent) is 73.

**Conclusions**

From this investigation, it has been found that investments in PEP are more likely for mature households than young families. Wealth and income are presupposed to be higher for the former based on the ability to support a mortgage. While specifically for young families with infants, financial resources are less likely to be prioritized towards a mortgage much less into PEP. Therefore, for PEP marketing purposes, it may beneficial to target families who can support a mortgage and those without young children.

However, in the future it would be interesting to expand this study by assessing whether location plays a role in PEP investments. Specifically, it would be a meaningful factor for the already PEP investing group of mature households. Those within this group who live in the suburbia and town could be assumed to hold more wealth than those living in the city. Since this is a group that is already assumed to be higher in wealth and income, it would be interesting to see whether living in the city affects their PEP investment ability.

Another factor that affects the wealth of a household is the ownership of a car. It would be interesting to investigate whether families would prioritize investing into PEP if they don’t own a car. This would help illustrate the importance of investing into PEP, and whether marketing firms should allocate resources advertising to non-car owners.

Overall, this has been a successful study that elucidated to target mature households. But, location and car-ownership are still important factors to investigate in the future.