Association Rule Mining

Project Overview

This project analyzes a “bankdata” dataset provided by our client. The dataset contains features on customer demographics, banking information, as well as whether or not the customer purchased a Personal Equity Plan (PEP) after the financial firm sent out a promotional direct mail piece advertising a call to action to purchase a PEP. The purpose of this analysis is to offer actionable insight based on customer profile models that better predict which customers are more likely or less likely to purchase Personal Equity Plans or similar type products. The analysis was conducted using Association Rule Mining using the Apriori algorithm conducted in a WEKA software suite. The following report provides a brief overview of Association Rule Mining, the process followed to conduct the analysis, and concludes with key findings, insights and recommendations base of the top 5 rules discovered through the analysis. The Appendices provide the detailed report outputs from WEKA.

Association Rule Mining

Association Rule Mining, often referred to as Market Basket Analysis, is a machine learning technique that identifies associations among features in a dataset to help identify useful and actionable patterns. One of the most widely used algorithms used for association rule mining is the Apriori algorithm. The Apriori algorithm uses a simple, but computationally intensive, approach to reduce the association rule space by requiring that the subset of a frequent feature (or combination of features) that imply some other feature’s outcome, must also be a frequent feature (or features) themselves.

Analyzing a dataset with Aprioir results in a collection of association rules that specify associations and relationships among features in the dataset. From a technical aspect, the rules are built as two-sided relationships that help show what group of features (left-hand side) often imply some other feature (right-hand side). Organizing the association in this way allows an analyst to identify a specific right-hand side feature, and then try to find what left-hand side feature (or features) lead to a specific nominal outcome of the right-hand side (e.g. “yes”, “no”).

The association rules that are generated are constrained, and later refined by three primary evaluation parameters: Support, Confidence and Lift. *Support* of a rule measures how frequently the features, or combination of features, occur in the dataset. It is calculated by taking the frequency of the feature (or combination of features e.g. X1 and X2) and dividing it by the total number of observations in the dataset (N). *Confidence* of a rule is a measure of its predictive power or accuracy. It is calculated dividing the support for a set of features that lead to some outcome (how feature X1 and X2 lead to -> Y), divided by the support for X1 and X2. *Lift* of a rule measures how

much more likely a feature (or combination of features, e.g. X1 and X2 –>Y) is likely to lead to some other feature's outcome (Z) relative to the rate of the other feature's outcome (Z) within the dataset absent the association. In other words, how often X leads to Z, compared to Z simply existing on its own, e.g. how many times did people buy pizza and wings, compared to how many times people just purchased wings.

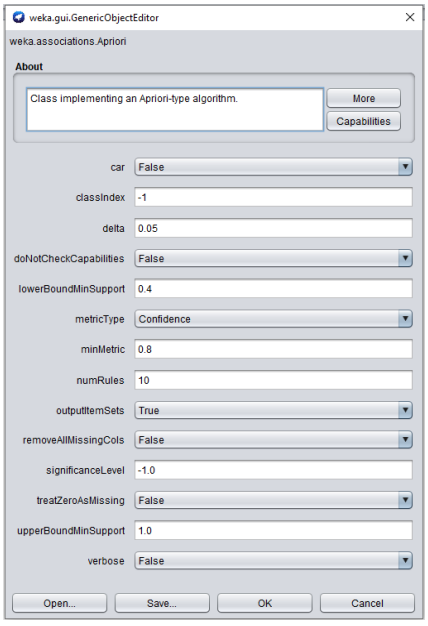
Some of the strengths of this algorithm include its applicability to large data sets, its usefulness in datamining and discovering unknow knowledge in datasets, and perhaps most importantly, the rules it generates are intuitive and easy to understand. Some of the drawbacks however are that the algorithm is not very useful on small datasets, and it is easy to draw the wrong conclusions from random patters. Additionally, association rules are not used for prediction but rather for unsupervised knowledge discovery. However, once the rules are established, organizations can use this new understanding of their data and apply classification and segmentation insights to their current marketing processes to improve outcomes.

Process Overview

Step 1: Prepare the dataset

This step was made easy by WEKA's intuitive system. Once the data was read in, it required only limited processing to prepare if for association rules mining. Out of the eleven features, one was removed, (ID Field) and only a few others required a simple discretization transformation to a nominal state from their original numeric state.

Step 2: Establish and tune parameters



This step begins with establishing the initialization of the Apriori function parameters and ends once the they have been tuned to meet a desired level of performance. The two metrics for evaluation are for Support and Confidence. For support there are three parameters to adjust and the initial settings are seen in the screen to the right. The Lower Bound minimum support was set to 0.4, and the Upper Bound minimum support to 1.0 with a Delta set to 0.05. This Delta setting allows the algorithm to step down from the upper limit to the lower limit in steps of 0.05. The number of rules to be generated was limited to 10. The Confidence value was set at 0.08 and the Output item sets was set to "True" to show the item sets the rules create. Below are the parameter definitions:

Apriori -I -N 10 -T 0 -C 0.8 -D 0.05 -U 1.0 -M 0.4 -S -1.0 -c -1

The results from these settings created 0 rules and the output can be seen in Appendix 1.A. On the second iteration, rule output was increased to 30 rules and the confidence was lowered to 0.6. This iteration produced 15 rules and produced one and two item itemsets. The output can be seen in Appendix 1.B.

Given that a rule's confidence is a measure of its predictive power or accuracy, its preferred to have the value be as high as possible but this of course comes with tradeoffs. As such the third iteration's confidence was increased to 0.75 and as tradeoff lowered the lower bound of support to 0.3. Interesting result on this run was that fewer rules (11 vs 15 from iteration 2) were generated, but the best rule has better parameters than iteration two as its confidence is 0.79 and lift is 1.04. This is an increase of 0.02 and 0.02 respectively. Additionally, these parameters offered the first three-item itemset (three features associated with an outcome) with four instances of this event. The results can be seen in Appendix 1.C.

In an effort to generate more rules, for the fourth interaction the confidence was lowered to .70 and support was held at 0.3. This tuning resulted in the execution of 14 cycles, and a total of 22 rules. The results can be seen in Appendix 1.D.

In the fifth iteration, minimum support was dropped to 0.25 to develop at least 30 rules and the iteration was successful. It generated 16 three-item itemsets and 30 rules. The results can be seen in Appendix 1.E.

Step 3: Classification

With ruleset tuning iterations complete, the focus turns to classification where the right hand side of our rules is set to the target feature of PEP (did the customer purchase a Personal Equity Plan, yes or no?) to determine what rules are generated and how we can best tune them to help determine which features, or combination of features, are most strongly associated with certain PEP outcomes.

With minimum support set to 0.2 and confidence set to 0.7, 16 cycles were performed that generated 3 three-item itemsets and a total of two rules. Each rule has a confidence above .75 which is good, but both rules only improve the understanding of which customers did not buy a PEP. To expand our ruleset to include instances where customers did buy PEPE, required making additional changes to the tuning parameters.

Lowering minimum support to 0.15 and confidence to .68 returned 10 rules. This iteration offered the same two high quality rules to help determine features leading to a "no" on PEP, and offer two valuable rules that leads to a "yes" on PEP. After several adjustments, the second iteration remained the best combination of parameters, and the top five rules from iteration two are covered below.

Key Findings, Insights, and Recommendations

Top Five Rules

What makes a rule valuable is that in addition to having high lift and confidence, the rule must also provide some non-trivial, actionable knowledge based on underlying business objectives. Generally, association rules can be divided into three categories: Actionable, trivial, inexplicable. The assumption made in this report is that the call to action listed in the direct mailer is successful if it drives the customer to perform the given action, which in this case is to buy a PEP. As such, actionable rules that offer insight into how to better target potential "yes", or to carve out potential "no" responses have been prioritized. Although the organization seeks to increase yes PEP responses, it is also valuable to understand the no PEP responses so the organization can limit future direct mailer costs by mailing only to the customers most likely to say yes. However, in order to prevent a potential false negative situation in which the analysis could potentially rule out certain customers that would be classified to select decline a PEP purchase when in fact they would have, the "no" rules must have high confidence to be considered a valuable rule. Given these constraints, the following rules and parameters that offer the best insight are listed below, and the details of the best performing prediction iteration can be found in Appendix 2.A

Rule #1

Rule: married=NO mortgage=NO ==> pep=YES

Parameters: Support: 0.217 / Confidence: 0.71 / Lift: 1.30

The first rule points to a "yes" PEP classification if a customer is "not married" and has "no mortgage." This rule's minimum support is 0.217 which is calculated by taking the frequency of the rule's instances (130) and dividing it by the total occurrences in the data set (600). The score of 0.217 means its relative occurrence in the dataset is low. This implies that most of the customers may not fit this "not married, no mortgage" profile if this dataset is representative of all their customers. The confidence of the rule is high with a score of 0.71 which derived by taking the rule's support of the instances where "not married" and "no mortgage" implied "Yes" to PEP (92/600 = 0.1533) and dividing it by the support for "not married" and "no mortgage" which is 0.217. Lastly, the lift, which is calculated by taking the confidence for "not married" and "no mortgage" leading to a "yes" on PEP (0.71) and divide it by the support of "Yes" on PEP in the dataset (326/600 = 0.5433), results in a lift score of 1.30. Since the score is greater than 1, this implies that the combination of these features and their values in this rule are found together more often than one would expect by chance, reflecting a true connection between these features.

Recommendation: If the client want to increase customers that elect to buy a PEP, this rule's high confidence and lift highlights a potential market opportunity to attract more customers that fit this profile.

Rule #2

children=1 ==> pep=YES

Parameters: Support: 0.225 / Confidence: 0.813 / Lift: 1.49

The second rule also points to “yes” on PEP classification if the customer is listed as having 1 child. Although this doesn’t provide a very rigorous profile of a customer, it does provide a feature to help segment the dataset and with its respectable support, confidence and lift scores, the rule remains valuable.

Recommendation: Prioritize effort on this segment to increase potential PEP sales. Additionally, the marketing team should segment the dataset to customers reporting just one child and conduct additional research to determine other attributes that might help explain similarities these customers have with other customers that are not included in this dataset.

Rule #3

married=YES children=0 save_act=YES ==> pep=NO

Parameters: Support: 0.198 / Confidence: 0.90 / Lift: 1.65

The third rule partitions the “no” on PEP classification. Married customers with no children, but have a savings account appear to be lagging on the uptake of the direct mailer call to action.

Recommendation: Families with no kids tend to be highly active and may avoid direct mail solicitation. I recommend the marketing team try a medium such as email to make the PEP offer.

Rule #4

married=YES children=0 mortgage=NO ==> pep=NO conf:(0.9)

Parameters: Support: 0.193 / Confidence: 0.90 / Lift: 1.65

The fourth rule partitions the “no” on PEP classification. Married customers with no children, and no mortgage appear to be lagging on the uptake of the direct mailer call to action.

Recommendation: This finding could be closely related to that of rule number 3 in that families with no kids tend to be highly active and may avoid direct mail solicitation. Additionally, if the family does not have a mortgage, they are most likely renters and move often potentially reducing the likelihood they received the direct mailer. I recommend the marketing team try a medium such as email to make the PEP offer.

Rule #5

children=0 save_act=YES current_act=YES ==> pep=NO

Parameters: Support: 0.222 / Confidence: 0.76 / Lift: 1.39

The fifth rule partitions the “no” on PEP classification for customers that have no children, have a savings account, and have a current account with the financial firm. This rule is interesting because it applies to nearly 25% of the dataset and this group of customers may represent a lucrative segment given these users show a propensity to save money given the have a savings account and also have a current account with the firm.

Recommendation: The financial firm should continue to foster their relationship with these consumers and attempt to identify what sort of features, rates, etc. would make purchasing a PEP appealing to them.

Appendix 1

A.

```
=== Run information ===

Scheme:       weka.associations.Apriori -I -N 10 -T 0 -C 0.8 -D 0.05 -U 1.0 -M 0.4 -S -1.0 -c -1
Relation:     mydata-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Discretize-B7-M-1.0-Rfirst-precision6-
weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R4-precision6-weka.filters.unsupervised.attribute.NumericToNominal-R6
Instances:    600
Attributes:   11
              age
              sex
              region
              income
              married
              children
              car
              save_act
              current_act
              mortgage
              pep
=== Associator model (full training set) ===

Apriori
=======

Minimum support: 0.4 (240 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Large Itemsets L(1):
sex=FEMALE 300
sex=MALE 300
region=INNER_CITY 269
income='(-inf-24386.173333]' 285
married=YES 396
children=0 263
car=NO 304
car=YES 296
save_act=YES 414
current_act=YES 455
mortgage=NO 391
pep=NO 326
pep=YES 274

Size of set of large itemsets L(2): 8

Large Itemsets L(2):
married=YES save_act=YES 277
married=YES current_act=YES 293
married=YES mortgage=NO 261
married=YES pep=NO 242
save_act=YES current_act=YES 319
save_act=YES mortgage=NO 270
current_act=YES mortgage=NO 301
current_act=YES pep=NO 244

Best rules found:
```

Appendix 1

B.

=== Run information ===

Scheme: weka.associations.Apriori -I -N 30 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.4 -S -1.0 -c -1
Relation: mydata-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Discretize-B7-M-1.0-Rfirst-precision6-weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R4-precision6-weka.filters.unsupervised.attribute.NumericToNominal-R6
Instances: 600
Attributes: 11
age
sex
region
income
married
children
car
save_act
current_act
mortgage
pep
=== Associator model (full training set) ===

Apriori
=======

Minimum support: 0.4 (240 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Large Itemsets L(1):
sex=FEMALE 300
sex=MALE 300
region=INNER_CITY 269
income='(-inf-24386.173333]' 285
married=YES 396
children=0 263
car=NO 304
car=YES 296
save_act=YES 414
current_act=YES 455
mortgage=NO 391
pep=NO 326
pep=YES 274

Size of set of large itemsets L(2): 8

Large Itemsets L(2):
married=YES save_act=YES 277
married=YES current_act=YES 293
married=YES mortgage=NO 261
married=YES pep=NO 242
save_act=YES current_act=YES 319
save_act=YES mortgage=NO 270
current_act=YES mortgage=NO 301
current_act=YES pep=NO 244

Best rules found:

1. save_act=YES 414 ==> current_act=YES 319 <conf:(0.77)> lift:(1.02) lev:(0.01) [5] conv:(1.04)
2. mortgage=NO 391 ==> current_act=YES 301 <conf:(0.77)> lift:(1.02) lev:(0.01) [4] conv:(1.04)
3. pep=NO 326 ==> current_act=YES 244 <conf:(0.75)> lift:(0.99) lev:(-0.01) [-3] conv:(0.95)
4. pep=NO 326 ==> married=YES 242 <conf:(0.74)> lift:(1.12) lev:(0.04) [26] conv:(1.3)
5. married=YES 396 ==> current_act=YES 293 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-7] conv:(0.92)
6. current_act=YES 455 ==> save_act=YES 319 <conf:(0.7)> lift:(1.02) lev:(0.01) [5] conv:(1.03)
7. married=YES 396 ==> save_act=YES 277 <conf:(0.7)> lift:(1.01) lev:(0.01) [3] conv:(1.02)
8. mortgage=NO 391 ==> save_act=YES 270 <conf:(0.69)> lift:(1) lev:(0) [0] conv:(0.99)
9. save_act=YES 414 ==> married=YES 277 <conf:(0.67)> lift:(1.01) lev:(0.01) [3] conv:(1.02)
10. mortgage=NO 391 ==> married=YES 261 <conf:(0.67)> lift:(1.01) lev:(0) [2] conv:(1.01)
11. current_act=YES 455 ==> mortgage=NO 301 <conf:(0.66)> lift:(1.02) lev:(0.01) [4] conv:(1.02)

12. married=YES 396 ==> mortgage=NO 261 <conf:(0.66)> lift:(1.01) lev:(0) [2] conv:(1.01)
13. save_act=YES 414 ==> mortgage=NO 270 <conf:(0.65)> lift:(1) lev:(0) [0] conv:(0.99)
14. current_act=YES 455 ==> married=YES 293 <conf:(0.64)> lift:(0.98) lev:(-0.01) [-7] conv:(0.95)
15. married=YES 396 ==> pep=NO 242 <conf:(0.61)> lift:(1.12) lev:(0.04) [26] conv:(1.17)

Appendix 1

C.

```
=== Run information ===

Scheme:       weka.associations.Apriori -I -N 30 -T 0 -C 0.75 -D 0.05 -U 1.0 -M 0.3 -S -1.0 -c -1
Relation:     mydata-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Discretize-B7-M-1.0-Rfirst-precision6-
weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R4-precision6-weka.filters.unsupervised.attribute.NumericToNominal-R6
Instances:    600
Attributes:   11
              age
              sex
              region
              income
              married
              children
              car
              save_act
              current_act
              mortgage
              pep
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.3 (180 instances)
Minimum metric <confidence>: 0.75
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 17

Large Itemsets L(1):
sex=FEMALE 300
sex=MALE 300
region=INNER_CITY 269
income='(-inf-24386.173333]' 285
income='(24386.173333-43758.136667]' 235
married=NO 204
married=YES 396
children=0 263
car=NO 304
car=YES 296
save_act=NO 186
save_act=YES 414
current_act=YES 455
mortgage=NO 391
mortgage=YES 209
pep=NO 326
pep=YES 274

Size of set of large itemsets L(2): 34

Large Itemsets L(2):
sex=FEMALE married=YES 195
sex=FEMALE save_act=YES 206
sex=FEMALE current_act=YES 230
sex=FEMALE mortgage=NO 205
sex=MALE married=YES 201
sex=MALE save_act=YES 208
sex=MALE current_act=YES 225
sex=MALE mortgage=NO 186
region=INNER_CITY current_act=YES 205
income='(-inf-24386.173333]' married=YES 195
income='(-inf-24386.173333]' current_act=YES 215
income='(-inf-24386.173333]' mortgage=NO 188
married=YES children=0 180
married=YES car=NO 202
married=YES car=YES 194
```

married=YES save_act=YES 277
married=YES current_act=YES 293
married=YES mortgage=NO 261
married=YES pep=NO 242
children=0 current_act=YES 199
car=NO save_act=YES 205
car=NO current_act=YES 235
car=NO mortgage=NO 197
car=YES save_act=YES 209
car=YES current_act=YES 220
car=YES mortgage=NO 194
save_act=YES current_act=YES 319
save_act=YES mortgage=NO 270
save_act=YES pep=NO 235
current_act=YES mortgage=NO 301
current_act=YES pep=NO 244
current_act=YES pep=YES 211
mortgage=NO pep=NO 209
mortgage=NO pep=YES 182

Size of set of large itemsets L(3): 4

Large Itemsets L(3):
married=YES save_act=YES current_act=YES 206
married=YES save_act=YES mortgage=NO 184
married=YES current_act=YES mortgage=NO 199
save_act=YES current_act=YES mortgage=NO 212

Best rules found:

1. save_act=YES mortgage=NO 270 ==> current_act=YES 212 <conf:(0.79)> lift:(1.04) lev:(0.01) [7] conv:(1.11)
2. car=NO 304 ==> current_act=YES 235 <conf:(0.77)> lift:(1.02) lev:(0.01) [4] conv:(1.05)
3. save_act=YES 414 ==> current_act=YES 319 <conf:(0.77)> lift:(1.02) lev:(0.01) [5] conv:(1.04)
4. pep=YES 274 ==> current_act=YES 211 <conf:(0.77)> lift:(1.02) lev:(0.01) [3] conv:(1.03)
5. mortgage=NO 391 ==> current_act=YES 301 <conf:(0.77)> lift:(1.02) lev:(0.01) [4] conv:(1.04)
6. sex=FEMALE 300 ==> current_act=YES 230 <conf:(0.77)> lift:(1.01) lev:(0) [2] conv:(1.02)
7. married=YES mortgage=NO 261 ==> current_act=YES 199 <conf:(0.76)> lift:(1.01) lev:(0) [1] conv:(1)
8. region=INNER_CITY 269 ==> current_act=YES 205 <conf:(0.76)> lift:(1) lev:(0) [1] conv:(1)
9. children=0 263 ==> current_act=YES 199 <conf:(0.76)> lift:(1) lev:(-0) [0] conv:(0.98)
10. income='(-inf-24386.173333]' 285 ==> current_act=YES 215 <conf:(0.75)> lift:(0.99) lev:(-0) [-1] conv:(0.97)
11. sex=MALE 300 ==> current_act=YES 225 <conf:(0.75)> lift:(0.99) lev:(-0) [-2] conv:(0.95)

Appendix 1

D.

```
=== Run information ===

Scheme:       weka.associations.Apriori -I -N 30 -T 0 -C 0.7 -D 0.05 -U 1.0 -M 0.3 -S -1.0 -c -1
Relation:     mydata-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Discretize-B7-M-1.0-Rfirst-precision6-
weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R4-precision6-weka.filters.unsupervised.attribute.NumericToNominal-R6
Instances:    600
Attributes:   11
              age
              sex
              region
              income
              married
              children
              car
              save_act
              current_act
              mortgage
              pep
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.3 (180 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 17

Large Itemsets L(1):
sex=FEMALE 300
sex=MALE 300
region=INNER_CITY 269
income='(-inf-24386.173333]' 285
income='(24386.173333-43758.136667]' 235
married=NO 204
married=YES 396
children=0 263
car=NO 304
car=YES 296
save_act=NO 186
save_act=YES 414
current_act=YES 455
mortgage=NO 391
mortgage=YES 209
pep=NO 326
pep=YES 274

Size of set of large itemsets L(2): 34

Large Itemsets L(2):
sex=FEMALE married=YES 195
sex=FEMALE save_act=YES 206
sex=FEMALE current_act=YES 230
sex=FEMALE mortgage=NO 205
sex=MALE married=YES 201
sex=MALE save_act=YES 208
sex=MALE current_act=YES 225
sex=MALE mortgage=NO 186
region=INNER_CITY current_act=YES 205
income='(-inf-24386.173333]' married=YES 195
income='(-inf-24386.173333]' current_act=YES 215
income='(-inf-24386.173333]' mortgage=NO 188
married=YES children=0 180
married=YES car=NO 202
married=YES car=YES 194
```

married=YES save_act=YES 277
married=YES current_act=YES 293
married=YES mortgage=NO 261
married=YES pep=NO 242
children=0 current_act=YES 199
car=NO save_act=YES 205
car=NO current_act=YES 235
car=NO mortgage=NO 197
car=YES save_act=YES 209
car=YES current_act=YES 220
car=YES mortgage=NO 194
save_act=YES current_act=YES 319
save_act=YES mortgage=NO 270
save_act=YES pep=NO 235
current_act=YES mortgage=NO 301
current_act=YES pep=NO 244
current_act=YES pep=YES 211
mortgage=NO pep=NO 209
mortgage=NO pep=YES 182

Size of set of large itemsets L(3): 4

Large Itemsets L(3):
married=YES save_act=YES current_act=YES 206
married=YES save_act=YES mortgage=NO 184
married=YES current_act=YES mortgage=NO 199
save_act=YES current_act=YES mortgage=NO 212

Best rules found:

1. save_act=YES mortgage=NO 270 ==> current_act=YES 212 <conf:(0.79)> lift:(1.04) lev:(0.01) [7] conv:(1.11)
2. car=NO 304 ==> current_act=YES 235 <conf:(0.77)> lift:(1.02) lev:(0.01) [4] conv:(1.05)
3. save_act=YES 414 ==> current_act=YES 319 <conf:(0.77)> lift:(1.02) lev:(0.01) [5] conv:(1.04)
4. pep=YES 274 ==> current_act=YES 211 <conf:(0.77)> lift:(1.02) lev:(0.01) [3] conv:(1.03)
5. mortgage=NO 391 ==> current_act=YES 301 <conf:(0.77)> lift:(1.02) lev:(0.01) [4] conv:(1.04)
6. sex=FEMALE 300 ==> current_act=YES 230 <conf:(0.77)> lift:(1.01) lev:(0) [2] conv:(1.02)
7. married=YES mortgage=NO 261 ==> current_act=YES 199 <conf:(0.76)> lift:(1.01) lev:(0) [1] conv:(1)
8. region=INNER_CITY 269 ==> current_act=YES 205 <conf:(0.76)> lift:(1) lev:(0) [1] conv:(1)
9. children=0 263 ==> current_act=YES 199 <conf:(0.76)> lift:(1) lev:(-0) [0] conv:(0.98)
10. income='(-inf-24386.173333]' 285 ==> current_act=YES 215 <conf:(0.75)> lift:(0.99) lev:(-0) [-1] conv:(0.97)
11. sex=MALE 300 ==> current_act=YES 225 <conf:(0.75)> lift:(0.99) lev:(-0) [-2] conv:(0.95)
12. pep=NO 326 ==> current_act=YES 244 <conf:(0.75)> lift:(0.99) lev:(-0.01) [-3] conv:(0.95)
13. married=YES save_act=YES 277 ==> current_act=YES 206 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-4] conv:(0.93)
14. car=YES 296 ==> current_act=YES 220 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-4] conv:(0.93)
15. pep=NO 326 ==> married=YES 242 <conf:(0.74)> lift:(1.12) lev:(0.04) [26] conv:(1.3)
16. married=YES 396 ==> current_act=YES 293 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-7] conv:(0.92)
17. pep=NO 326 ==> save_act=YES 235 <conf:(0.72)> lift:(1.04) lev:(0.02) [10] conv:(1.1)
18. car=YES 296 ==> save_act=YES 209 <conf:(0.71)> lift:(1.02) lev:(0.01) [4] conv:(1.04)
19. married=YES mortgage=NO 261 ==> save_act=YES 184 <conf:(0.7)> lift:(1.02) lev:(0.01) [3] conv:(1.04)
20. current_act=YES mortgage=NO 301 ==> save_act=YES 212 <conf:(0.7)> lift:(1.02) lev:(0.01) [4] conv:(1.04)
21. married=YES current_act=YES 293 ==> save_act=YES 206 <conf:(0.7)> lift:(1.02) lev:(0.01) [3] conv:(1.03)
22. current_act=YES 455 ==> save_act=YES 319 <conf:(0.7)> lift:(1.02) lev:(0.01) [5] conv:(1.03)

Appendix 1

E.

```
=== Run information ===

Scheme:       weka.associations.Apriori -I -N 30 -T 0 -C 0.7 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1
Relation:     mydata-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Discretize-B7-M-1.0-Rfirst-precision6-
weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R4-precision6-weka.filters.unsupervised.attribute.NumericToNominal-R6
Instances:    600
Attributes:   11
              age
              sex
              region
              income
              married
              children
              car
              save_act
              current_act
              mortgage
              pep
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.25 (150 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 18

Large Itemsets L(1):
sex=FEMALE 300
sex=MALE 300
region=INNER_CITY 269
region=TOWN 173
income='(-inf-24386.173333]' 285
income='(24386.173333-43758.136667]' 235
married=NO 204
married=YES 396
children=0 263
car=NO 304
car=YES 296
save_act=NO 186
save_act=YES 414
current_act=YES 455
mortgage=NO 391
mortgage=YES 209
pep=NO 326
pep=YES 274

Size of set of large itemsets L(2): 55

Large Itemsets L(2):
sex=FEMALE married=YES 195
sex=FEMALE car=NO 153
sex=FEMALE save_act=YES 206
sex=FEMALE current_act=YES 230
sex=FEMALE mortgage=NO 205
sex=FEMALE pep=NO 170
sex=MALE married=YES 201
sex=MALE car=NO 151
sex=MALE save_act=YES 208
sex=MALE current_act=YES 225
sex=MALE mortgage=NO 186
sex=MALE pep=NO 156
region=INNER_CITY married=YES 178
region=INNER_CITY save_act=YES 173
```

region=INNER_CITY current_act=YES 205
region=INNER_CITY mortgage=NO 175
income='(-inf-24386.173333]' married=YES 195
income='(-inf-24386.173333]' car=NO 162
income='(-inf-24386.173333]' save_act=YES 171
income='(-inf-24386.173333]' current_act=YES 215
income='(-inf-24386.173333]' mortgage=NO 188
income='(-inf-24386.173333]' pep=NO 176
income='(24386.173333-43758.136667]' save_act=YES 163
income='(24386.173333-43758.136667]' current_act=YES 177
married=NO current_act=YES 162
married=YES children=0 180
married=YES car=NO 202
married=YES car=YES 194
married=YES save_act=YES 277
married=YES current_act=YES 293
married=YES mortgage=NO 261
married=YES pep=NO 242
married=YES pep=YES 154
children=0 save_act=YES 174
children=0 current_act=YES 199
children=0 mortgage=NO 164
children=0 pep=NO 167
car=NO save_act=YES 205
car=NO current_act=YES 235
car=NO mortgage=NO 197
car=NO pep=NO 168
car=YES save_act=YES 209
car=YES current_act=YES 220
car=YES mortgage=NO 194
car=YES pep=NO 158
save_act=YES current_act=YES 319
save_act=YES mortgage=NO 270
save_act=YES pep=NO 235
save_act=YES pep=YES 179
current_act=YES mortgage=NO 301
current_act=YES mortgage=YES 154
current_act=YES pep=NO 244
current_act=YES pep=YES 211
mortgage=NO pep=NO 209
mortgage=NO pep=YES 182

Size of set of large itemsets L(3): 16

Large Itemsets L(3):
sex=FEMALE save_act=YES current_act=YES 160
sex=FEMALE current_act=YES mortgage=NO 159
sex=MALE save_act=YES current_act=YES 159
married=YES car=NO current_act=YES 151
married=YES save_act=YES current_act=YES 206
married=YES save_act=YES mortgage=NO 184
married=YES save_act=YES pep=NO 175
married=YES current_act=YES mortgage=NO 199
married=YES current_act=YES pep=NO 177
married=YES mortgage=NO pep=NO 171
car=NO save_act=YES current_act=YES 159
car=NO current_act=YES mortgage=NO 158
car=YES save_act=YES current_act=YES 160
save_act=YES current_act=YES mortgage=NO 212
save_act=YES current_act=YES pep=NO 179
current_act=YES mortgage=NO pep=NO 158

Best rules found:

1. mortgage=NO pep=NO 209 ==> married=YES 171 <conf:(0.82)> lift:(1.24) lev:(0.06) [33] conv:(1.82)
2. car=NO mortgage=NO 197 ==> current_act=YES 158 <conf:(0.8)> lift:(1.06) lev:(0.01) [8] conv:(1.19)
3. married=NO 204 ==> current_act=YES 162 <conf:(0.79)> lift:(1.05) lev:(0.01) [7] conv:(1.15)
4. save_act=YES mortgage=NO 270 ==> current_act=YES 212 <conf:(0.79)> lift:(1.04) lev:(0.01) [7] conv:(1.11)
5. sex=FEMALE save_act=YES 206 ==> current_act=YES 160 <conf:(0.78)> lift:(1.02) lev:(0.01) [3] conv:(1.06)
6. sex=FEMALE mortgage=NO 205 ==> current_act=YES 159 <conf:(0.78)> lift:(1.02) lev:(0.01) [3] conv:(1.05)
7. car=NO save_act=YES 205 ==> current_act=YES 159 <conf:(0.78)> lift:(1.02) lev:(0.01) [3] conv:(1.05)
8. car=NO 304 ==> current_act=YES 235 <conf:(0.77)> lift:(1.02) lev:(0.01) [4] conv:(1.05)
9. save_act=YES 414 ==> current_act=YES 319 <conf:(0.77)> lift:(1.02) lev:(0.01) [5] conv:(1.04)
10. pep=YES 274 ==> current_act=YES 211 <conf:(0.77)> lift:(1.02) lev:(0.01) [3] conv:(1.03)

11. mortgage=NO 391 ==> current_act=YES 301 <conf:(0.77)> lift:(1.02) lev:(0.01) [4] conv:(1.04)
12. sex=FEMALE 300 ==> current_act=YES 230 <conf:(0.77)> lift:(1.01) lev:(0) [2] conv:(1.02)
13. car=YES save_act=YES 209 ==> current_act=YES 160 <conf:(0.77)> lift:(1.01) lev:(0) [1] conv:(1.01)
14. sex=MALE save_act=YES 208 ==> current_act=YES 159 <conf:(0.76)> lift:(1.01) lev:(0) [1] conv:(1.01)
15. married=YES mortgage=NO 261 ==> current_act=YES 199 <conf:(0.76)> lift:(1.01) lev:(0) [1] conv:(1)
16. region=INNER_CITY 269 ==> current_act=YES 205 <conf:(0.76)> lift:(1) lev:(0) [1] conv:(1)
17. save_act=YES pep=NO 235 ==> current_act=YES 179 <conf:(0.76)> lift:(1) lev:(0) [0] conv:(1)
18. children=0 263 ==> current_act=YES 199 <conf:(0.76)> lift:(1) lev:(-0) [0] conv:(0.98)
19. mortgage=NO pep=NO 209 ==> current_act=YES 158 <conf:(0.76)> lift:(1) lev:(-0) [0] conv:(0.97)
20. income='(-inf-24386.173333]' 285 ==> current_act=YES 215 <conf:(0.75)> lift:(0.99) lev:(-0) [-1] conv:(0.97)
21. income='(24386.173333-43758.136667]' 235 ==> current_act=YES 177 <conf:(0.75)> lift:(0.99) lev:(-0) [-1] conv:(0.96)
22. sex=MALE 300 ==> current_act=YES 225 <conf:(0.75)> lift:(0.99) lev:(-0) [-2] conv:(0.95)
23. pep=NO 326 ==> current_act=YES 244 <conf:(0.75)> lift:(0.99) lev:(-0.01) [-3] conv:(0.95)
24. married=YES car=NO 202 ==> current_act=YES 151 <conf:(0.75)> lift:(0.99) lev:(-0) [-2] conv:(0.94)
25. save_act=YES pep=NO 235 ==> married=YES 175 <conf:(0.74)> lift:(1.13) lev:(0.03) [19] conv:(1.31)
26. married=YES save_act=YES 277 ==> current_act=YES 206 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-4] conv:(0.93)
27. car=YES 296 ==> current_act=YES 220 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-4] conv:(0.93)
28. pep=NO 326 ==> married=YES 242 <conf:(0.74)> lift:(1.12) lev:(0.04) [26] conv:(1.3)
29. married=YES 396 ==> current_act=YES 293 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-7] conv:(0.92)
30. mortgage=YES 209 ==> current_act=YES 154 <conf:(0.74)> lift:(0.97) lev:(-0.01) [-4] conv:(0.9)

Appendix 2

A.

```
=== Run information ===

Scheme:       weka.associations.Apriori -I -N 15 -T 0 -C 0.65 -D 0.02 -U 1.0 -M 0.15 -S -1.0 -A -c -1
Relation:     mydata-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Discretize-B7-M-1.0-Rfirst-precision6-
weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R4-precision6-weka.filters.unsupervised.attribute.NumericToNominal-R6
Instances:    600
Attributes:   11
              age
              sex
              region
              income
              married
              children
              car
              save_act
              current_act
              mortgage
              pep
=== Associator model (full training set) ===

Apriori
=======

Minimum support: 0.15 (90 instances)
Minimum metric <confidence>: 0.65
Number of cycles performed: 43

Generated sets of large itemsets:

Size of set of large itemsets L(1): 31

Large Itemsets L(1):
sex=FEMALE 300
0  170
sex=FEMALE 300
1  130
sex=MALE 300
0  156
sex=MALE 300
1  144
region=INNER_CITY 269
0  146
region=INNER_CITY 269
1  123
region=TOWN 173
0  102
income='(-inf-24386.173333]' 285
0  176
income='(-inf-24386.173333]' 285
1  109
income='(24386.173333-43758.136667]' 235
0  124
income='(24386.173333-43758.136667]' 235
1  111
married=NO 204
1  120
married=YES 396
0  242
married=YES 396
1  154
children=0 263
0  167
children=0 263
1  96
children=1 135
1  110
car=NO 304
0  168
car=NO 304
```

1 136
car=YES 296
0 158
car=YES 296
1 138
save_act=NO 186
0 91
save_act=NO 186
1 95
save_act=YES 414
0 235
save_act=YES 414
1 179
current_act=YES 455
0 244
current_act=YES 455
1 211
mortgage=NO 391
0 209
mortgage=NO 391
1 182
mortgage=YES 209
0 117
mortgage=YES 209
1 92

Size of set of large itemsets L(2): 59

Large Itemsets L(2):
sex=FEMALE married=YES 195
0 127
sex=FEMALE children=0 132
0 90
sex=FEMALE save_act=YES 206
0 122
sex=FEMALE current_act=YES 230
0 128
sex=FEMALE mortgage=NO 205
0 115
sex=FEMALE current_act=YES 230
1 102
sex=FEMALE mortgage=NO 205
1 90
sex=MALE income='(-inf-24386.173333]' 149
0 91
sex=MALE married=YES 201
0 115
sex=MALE save_act=YES 208
0 113
sex=MALE current_act=YES 225
0 116
sex=MALE mortgage=NO 186
0 94
sex=MALE save_act=YES 208
1 95
sex=MALE current_act=YES 225
1 109
sex=MALE mortgage=NO 186
1 92
region=INNER_CITY married=YES 178
0 112
region=INNER_CITY save_act=YES 173
0 100
region=INNER_CITY current_act=YES 205
0 115
region=INNER_CITY mortgage=NO 175
0 96
region=INNER_CITY current_act=YES 205
1 90
income='(-inf-24386.173333]' married=YES 195
0 128
income='(-inf-24386.173333]' car=NO 162
0 100
income='(-inf-24386.173333]' save_act=YES 171

0 119
income='(-inf-24386.173333]' current_act=YES 215
0 132
income='(-inf-24386.173333]' mortgage=NO 188
0 118
income='(24386.173333-43758.136667]' married=YES 148
0 92
income='(24386.173333-43758.136667]' save_act=YES 163
0 90
married=NO current_act=YES 162
1 95
married=NO mortgage=NO 130
1 92
married=YES children=0 180
0 141
married=YES car=NO 202
0 126
married=YES car=YES 194
0 116
married=YES save_act=YES 277
0 175
married=YES current_act=YES 293
0 177
married=YES mortgage=NO 261
0 171
married=YES save_act=YES 277
1 102
married=YES current_act=YES 293
1 116
married=YES mortgage=NO 261
1 90
children=0 car=NO 139
0 91
children=0 save_act=YES 174
0 131
children=0 current_act=YES 199
0 127
children=0 mortgage=NO 164
0 107
car=NO save_act=YES 205
0 117
car=NO current_act=YES 235
0 125
car=NO mortgage=NO 197
0 108
car=NO current_act=YES 235
1 110
car=YES save_act=YES 209
0 118
car=YES current_act=YES 220
0 119
car=YES mortgage=NO 194
0 101
car=YES save_act=YES 209
1 91
car=YES current_act=YES 220
1 101
car=YES mortgage=NO 194
1 93
save_act=YES current_act=YES 319
0 179
save_act=YES mortgage=NO 270
0 142
save_act=YES mortgage=YES 144
0 93
save_act=YES current_act=YES 319
1 140
save_act=YES mortgage=NO 270
1 128
current_act=YES mortgage=NO 301
0 158
current_act=YES mortgage=NO 301
1 143

Size of set of large itemsets L(3): 19

Large Itemsets L(3):
sex=FEMALE married=YES save_act=YES 137
0 91
sex=FEMALE married=YES current_act=YES 146
0 93
sex=FEMALE married=YES mortgage=NO 138
0 93
sex=FEMALE save_act=YES current_act=YES 160
0 92
income='(-inf-24386.173333]' married=YES current_act=YES 144
0 94
income='(-inf-24386.173333]' married=YES mortgage=NO 133
0 93
income='(-inf-24386.173333]' save_act=YES current_act=YES 132
0 90
income='(-inf-24386.173333]' current_act=YES mortgage=NO 143
0 90
married=YES children=0 save_act=YES 119
0 107
married=YES children=0 current_act=YES 133
0 105
married=YES children=0 mortgage=NO 116
0 104
married=YES car=NO current_act=YES 151
0 92
married=YES save_act=YES current_act=YES 206
0 130
married=YES save_act=YES mortgage=NO 184
0 120
married=YES current_act=YES mortgage=NO 199
0 129
children=0 save_act=YES current_act=YES 133
0 101
car=YES save_act=YES current_act=YES 160
0 92
save_act=YES current_act=YES mortgage=NO 212
0 108
save_act=YES current_act=YES mortgage=NO 212
1 104

Size of set of large itemsets L(4): 1
Large Itemsets L(4):
married=YES save_act=YES current_act=YES mortgage=NO 142
0 91

Best rules found:

 1. married=YES children=0 save_act=YES 119 ==> pep=NO 107   conf:(0.9)
 2. married=YES children=0 mortgage=NO 116 ==> pep=NO 104   conf:(0.9)
 3. children=1 135 ==> pep=YES 110   conf:(0.81)
 4. married=YES children=0 current_act=YES 133 ==> pep=NO 105   conf:(0.79)
 5. married=YES children=0 180 ==> pep=NO 141   conf:(0.78)
 6. children=0 save_act=YES current_act=YES 133 ==> pep=NO 101   conf:(0.76)
 7. children=0 save_act=YES 174 ==> pep=NO 131   conf:(0.75)
 8. married=NO mortgage=NO 130 ==> pep=YES 92   conf:(0.71)
 9. income='(-inf-24386.173333]' married=YES mortgage=NO 133 ==> pep=NO 93   conf:(0.7)
10. income='(-inf-24386.173333]' save_act=YES 171 ==> pep=NO 119   conf:(0.7)
11. sex=FEMALE children=0 132 ==> pep=NO 90   conf:(0.68)
12. income='(-inf-24386.173333]' save_act=YES current_act=YES 132 ==> pep=NO 90   conf:(0.68)
13. sex=FEMALE married=YES mortgage=NO 138 ==> pep=NO 93   conf:(0.67)
14. sex=FEMALE married=YES save_act=YES 137 ==> pep=NO 91   conf:(0.66)
15. income='(-inf-24386.173333]' married=YES 195 ==> pep=NO 128   conf:(0.66)