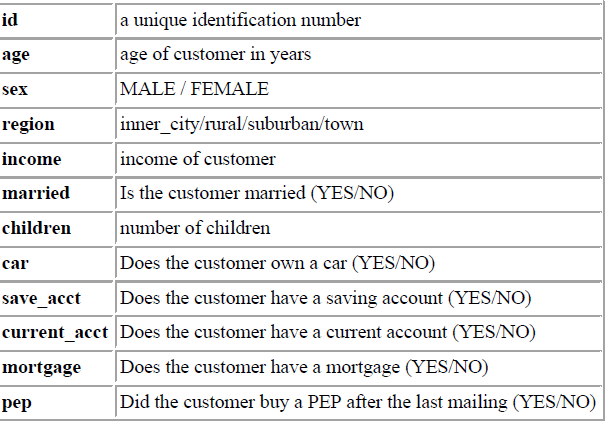
Jacob Dineen

IST 565 Data Mining

Homework 3

10/20/17

**Using Weka**



**Description of Preprocessing Steps:**

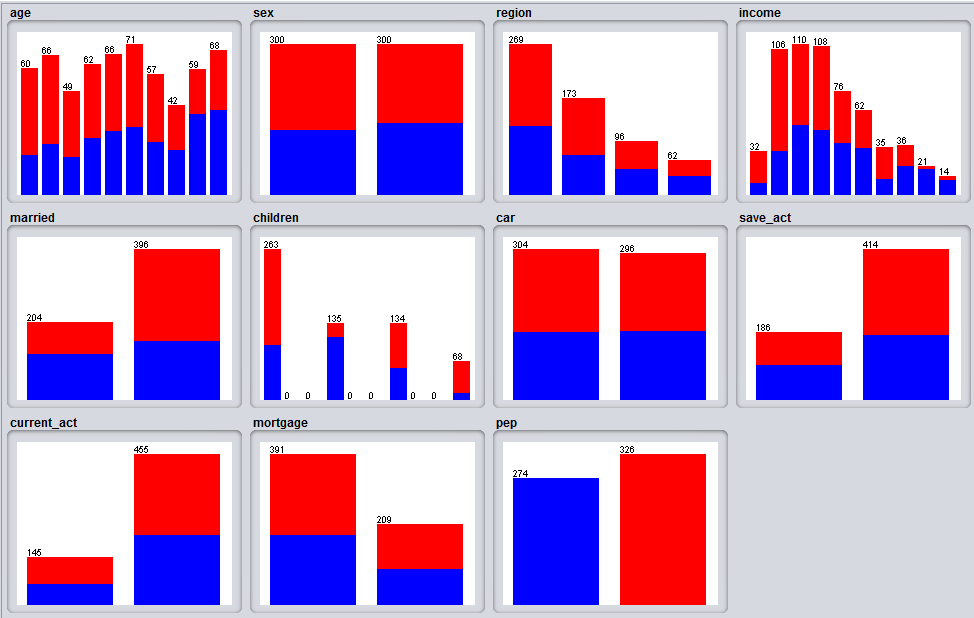
Remove ID field. Discretize Age and Income Variables (Currently Numeric)

Result: Age into 10 distinct bins. Income into 10 distinct bins.

Children is bucketed into ten bins, but is a rendering of four discrete values: 0, 1,2, 3+

No data is missing from any columns.

11 Nominal Attributes Remaining.



**Description of parameters and experiments in order to obtain strong rules:**

To start, I experimented with tapering confidence up and down. All rules generated have Car = TRUE, and the PEP field used as the right-hand side of the equation [x] < PEP == YES. I am interested in seeing what demographic and behavioral banking patterns are correlated with obtaining/owning a Personal Equity Plan, specifically in the instance defined by the definition: Did the customer buy a PEP after the last mailing? So what is known is that a campaign was previously launched to garner more traffic towards the offering. Most of the high confidence rules generated have high support for PEP == No, so while it would be interesting to see how we may change perception/persuasion techniques to alter thinking, we currently want to see what is working. Note: I switched up tactics because there weren’t enough interesting rules that contained PEP == Yes. Browsing down a list of 30 generated rules with confidence > .70 and min support of 0.10, we can see some patterns forming. Almost all rules contain a LHS item under: Married, Children, Savings Account, Current account, Mortgage and/or Sex. Rare instances of attribute occurrence are tied to Region and Car. Notably absent from the itemsets are the attributes: Age and Income.

When examining Lift, it is discovered that CAR cannot be set to specify a right-hand side. It is also discovered that setting the lower bound min support greater than .30 results in 0 rules generated, so we’ll use .2 as our LBMS, and 1.2 as our minimum lift, as that should separate mediocre rules from decent ones. This doesn’t return enough rules where PEP == Yes, so we will drop out LBMS to .15.

I will try to avoid iterations of rules containing the same items, although the instances containing child = 1 appear consistently in rules where the RHS is set to PEP.

Support, Confidence and Lift values

**Give the top 5 most interesting rules and the 3 items listed above for each rule:**

The below were the most interesting rules in terms of joint results of confidence and lift.

children='(0.9-1.2]' 135 ==> married=YES pep=YES 74 conf:(0.55) < lift:(2.14)> lev:(0.07) [39] conv:(1.62)

Support: 74/600 = .12

Confidence: 0.55

Lift: 2.14

This pattern suggests that, with a moderate level of confidence, individuals with one child are often married, and have a personal equity plan. If we are trying to be more targeted in who we persuade to sign up for a PEP, it may be beneficial to look at individuals that have a single child. Perhaps launch a follow up mailing campaign using this as a criteria.

children='(0.9-1.2]' current\_act=YES 101 ==> save\_act=YES pep=YES 63 conf:(0.62) < lift:(2.09)> lev:(0.05) [32] conv:(1.82)

Support: 63/600 = .11

Confidence: .62

Lift: 2.09

The above pattern recognizes an association between individuals who have 1 child and a current account at the bank as being both savings account and PEP members within the branch. Targeting existing customers may be a way to market this product, rather than launching a full-blown awareness campaign for nonexistent bank-ees, or conquesting from different branches. Maybe we can utilize our CRM to launch a loyalty program pertaining to the personal equity plan.

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

Support: 71/600 = .118

Confidence: .85

Lift: 1.85

This pattern follows along the lines of the first pattern, noting that an individual with a single child, but who does not have a current mortgage is likely to have a personal equity plan. Perhaps individuals without a mortgage payment are more likely to have spare money to place into a PEP, which is kind of along the same lines as a 401k for UK citizens, with the idiosyncrasy that it can be liquidated after the trust expires, which is normally one year. This makes a PEP an attractive plan for those who don’t have financial stresses.

married=YES children='(0.9-1.2]' 89 ==> pep=YES 74 conf:(0.83) < lift:(1.82)> lev:(0.06) [33] conv:(3.02)

Support: 74/600 = .123

Confidence: .83

Lift: 1.82

A slightly different iteration of a common rule involving children- Those with children and who are married are highly likely to contain a PEP. This helps us to narrow down our approach. Perhaps people with spouses and at least a single child are more concerned about the family’s financial future. Maybe it can be marketed as a surefire way to protect your family in case of job insecurity/death, etc.

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)

Support: 80/600 = .133

Confidence: 0.92

Lift: 1.69

I found this rule interesting because it contrasts an earlier rule about brand loyalty. Those who are married, but have no children and possess both savings and current accounts are not likely to have a PEP. The only real difference here is the fact that an individual has a child count < 1. Perhaps children, appearing in most rules, is a significant factor in one’s decision making process to protect their future financial position.

**For one rule, discuss the support, confidence and lift numbers and how they were**

**computed from the data:**

children='(0.9-1.2]' 135 ==> married=YES pep=YES 74 conf:(0.55) < lift:(2.14)> lev:(0.07) [39] conv:(1.62)

Support- The support for this rule would be computed by taking the total number of transactions that contain an itemset and dividing them by the total number of transactions.

74 instances of both transactions occurring

600 total transactions

74/600 = .12

74 individuals were recognized as having 1 child, being married, and have a PEP.

Confidence: A measure of how frequently items in Y appear in transactions containing X.

77/135 = 0.55

135 individuals contained a record of children = 1. Of those 135, 74 had a personal equity plan.

Lift: A measure to show correlation between X and Y.

(sup(X U Y)/ N) / (sup(X)/ N\*sup(Y)/ N )

This is found by taking the support count of occurrences of both X and Y, and fractionizing them against the total number of instances. You then divide that number by the product of the fractionized support of x and y.

**EG. ((74/600))/((135/600)\*(154/600)) = 2.135642**

So 74 is the number of times this occurred: (children=1, married=yes, pep=yes).

135 is the number of time that children = 1 occurred.

154 is the number of times that married= yes, pep = yes occurred.