DATA MINING

Assignment 1

25-04-2014

Group Members:

Santhosh Alladi (2011A7PS086H)

Surya Nikhil R (2011A7PS201H)

Manideep Illendula (2011C6PS662H)

Association Rule Mining

Objective:

The objective of this assignment is to pre-process a data set, do associative rule mining and then figuring out interesting rules which could be useful. In our assignment we used **bank data** set with the following attributes:

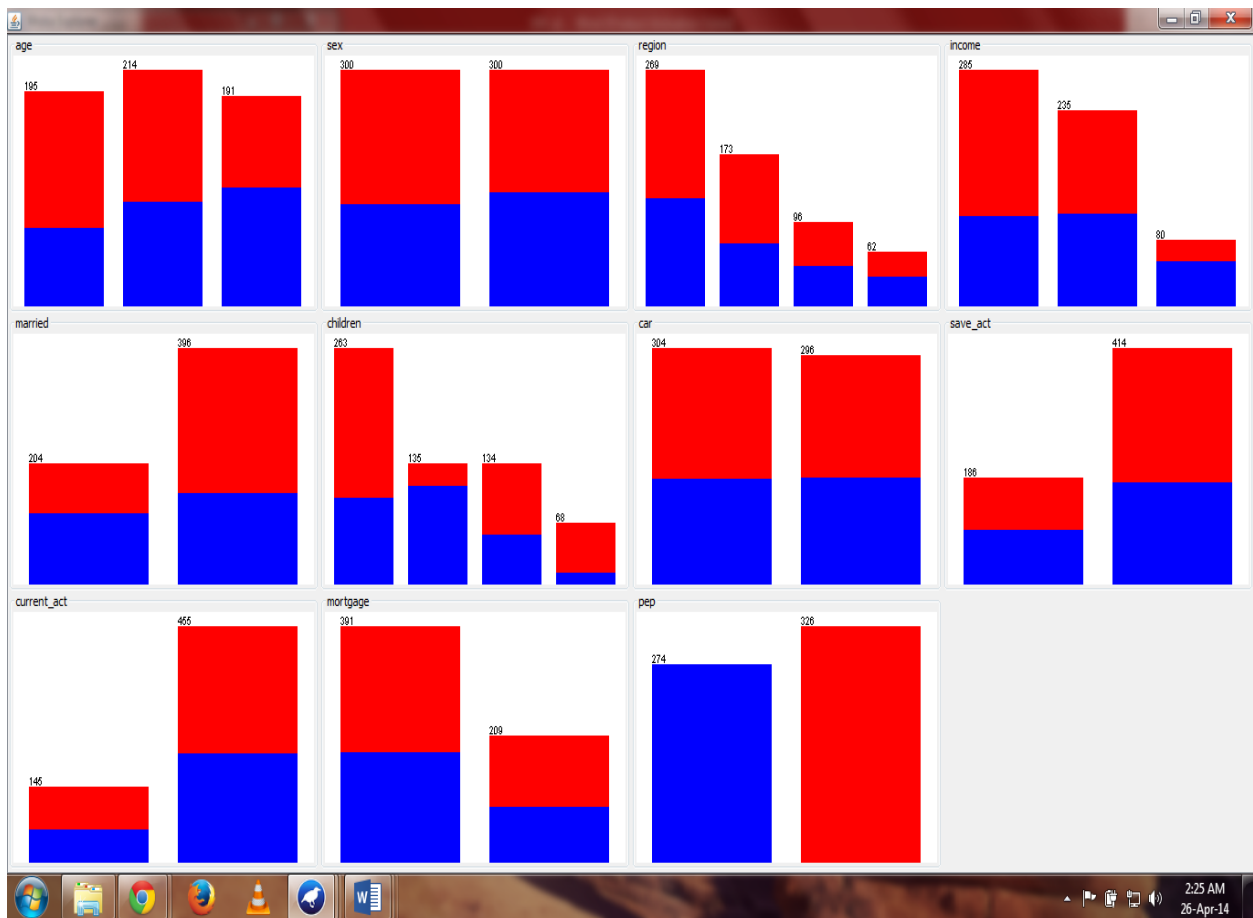| id	a unique identification number
age	age of customer in years (numeric)
sex	MALE / FEMALE
region	inner_city/rural/suburban/town
income	income of customer (numeric)
married	is the customer married (YES/NO)
children	number of children (numeric)
car	does the customer own a car (YES/NO)
save_acct	does the customer have a saving account (YES/NO)
current_acct	does the customer have a current account (YES/NO)
mortgage	does the customer have a mortgage (YES/NO)
pep	did the customer buy a PEP (Personal Equity Plan) after the last mailing (YES/NO)

Data set description:

As shown in the above table, the data set has 12 attributes. The unprocessed data set has 600 instances in total as shown in the attached file (bankData.arff). The following pre-processing steps are done for proper association rule mining-

- The id attribute has been removed, as it is of no importance.
- Then discretization is done:
  - Under this first the children attribute is discretized to values 0,1,2 and 3 as this attribute takes only these 4 values.
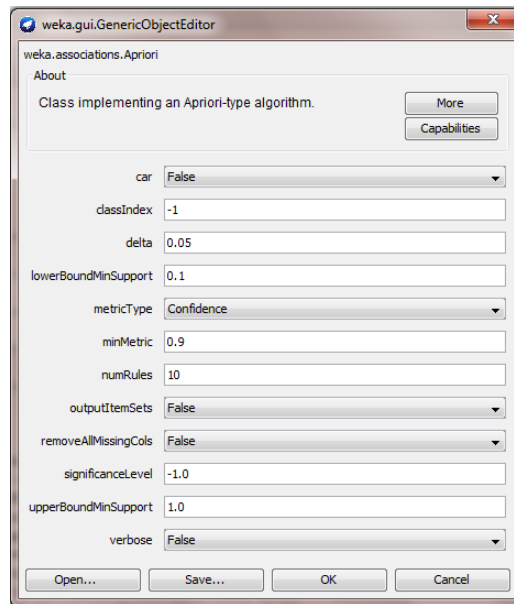 - The age and income attributes are also discretized into 3 bins each.

The data set after all the pre-processing steps is in the attached file (bankDataFinal.arff).

The screen shot gives the visualization of the attribute data after the pre-processing of the dataset:

Rule Mining:

We did the rule mining using Apriori principle. Also, we have done rule mining using two metrics- confidence and lift. The screen shots show the values of the different parameters.
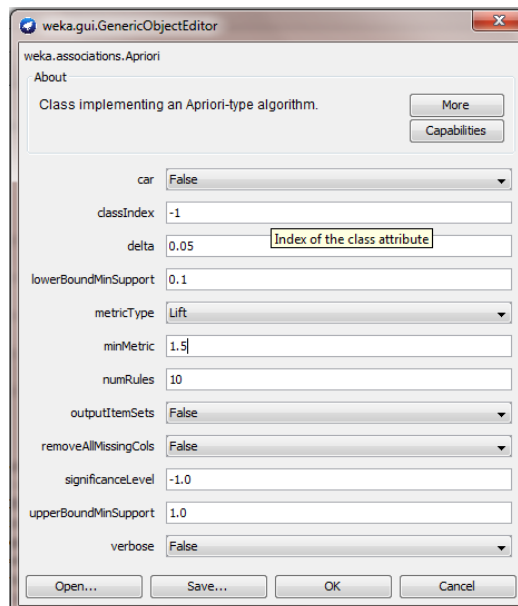

Parameters for confidence

Parameters for lift

Resulting Rules:

After association rule mining the following are the 10 interesting rules.

1) Rules based on **confidence**:

```
 1. income=43759_max 80 ==> save_act=YES 80     conf:(1)
 2. age=52_max income=43759_max 76 ==> save_act=YES 76     conf:(1)
 3. income=43759_max current_act=YES 63 ==> save_act=YES 63
conf:(1)
 4. age=52_max income=43759_max current_act=YES 61 ==> save_act=YES 61
conf:(1)
 5. children=0 save_act=YES mortgage=NO pep=NO 74 ==> married=YES 73
conf:(0.99)
 6. sex=FEMALE children=0 mortgage=NO pep=NO 64 ==> married=YES 63
conf:(0.98)
 7. children=0 current_act=YES mortgage=NO pep=NO 82 ==> married=YES
80     conf:(0.98)
 8. children=0 mortgage=NO pep=NO 107 ==> married=YES 104
conf:(0.97)
 9. income=43759_max current_act=YES 63 ==> age=52_max 61
conf:(0.97)
10. income=43759_max save_act=YES current_act=YES 63 ==> age=52_max 61
conf:(0.97)
```

2) Rules based on **lift**:

```
 1. age=0_34 195 ==> income=0_24386 current_act=YES 138     conf:(0.71)
< lift:(1.97)> lev:(0.11) [68] conv:(2.16)
 2. income=0_24386 current_act=YES 215 ==> age=0_34 138     conf:(0.64)
< lift:(1.97)> lev:(0.11) [68] conv:(1.86)
 3. income=0_24386 285 ==> age=0_34 car=NO 100     conf:(0.35) <
lift:(1.97)> lev:(0.08) [49] conv:(1.26)
 4. age=0_34 car=NO 107 ==> income=0_24386 100     conf:(0.93) <
lift:(1.97)> lev:(0.08) [49] conv:(7.02)
 5. age=0_34 195 ==> income=0_24386 pep=NO 111     conf:(0.57) <
lift:(1.94)> lev:(0.09) [53] conv:(1.62)
 6. income=0_24386 pep=NO 176 ==> age=0_34 111     conf:(0.63) <
lift:(1.94)> lev:(0.09) [53] conv:(1.8)
 7. age=0_34 195 ==> income=0_24386 save_act=YES 106     conf:(0.54) <
lift:(1.91)> lev:(0.08) [50] conv:(1.55)
 8. income=0_24386 save_act=YES 171 ==> age=0_34 106     conf:(0.62) <
lift:(1.91)> lev:(0.08) [50] conv:(1.75)
 9. income=0_24386 285 ==> age=0_34 mortgage=NO 113     conf:(0.4) <
lift:(1.9)> lev:(0.09) [53] conv:(1.3)
10. age=0_34 mortgage=NO 125 ==> income=0_24386 113     conf:(0.9) <
lift:(1.9)> lev:(0.09) [53] conv:(5.05)
```

The text files containing the detailed result of the rule mining are attached along with this report (bankData_a*.txt).