

Lab04 – weka

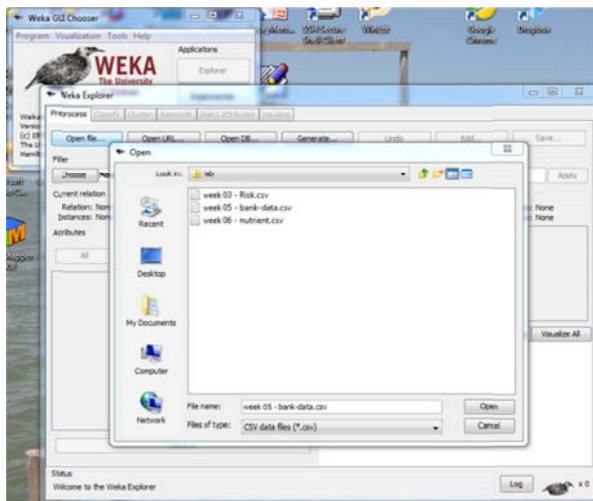
data file: http://paginas.fe.up.pt/~ec/files_1112/lab04-bank-data.csv

Attribute description

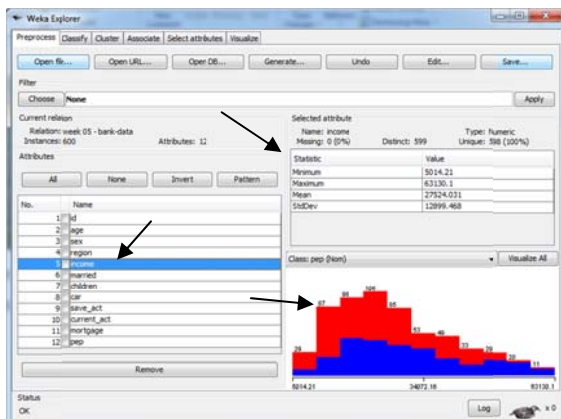
id	a unique identification number
age	age of customer in years (numeric)
sex	MALE / FEMALE
region	inner_city/rural/suburban/town
income	income of customer (numeric)
married	is the customer married (YES/NO)
children	number of children (numeric)
car	does the customer own a car (YES/NO)
save_acct	does the 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 PRE-PROCESSING

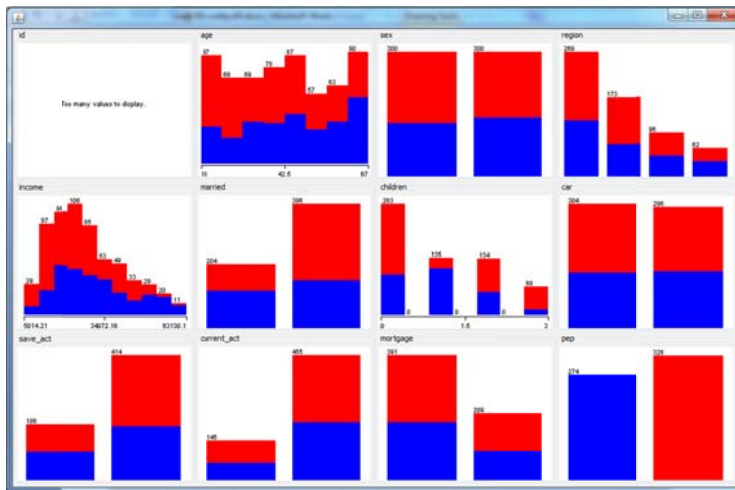
Open data file.



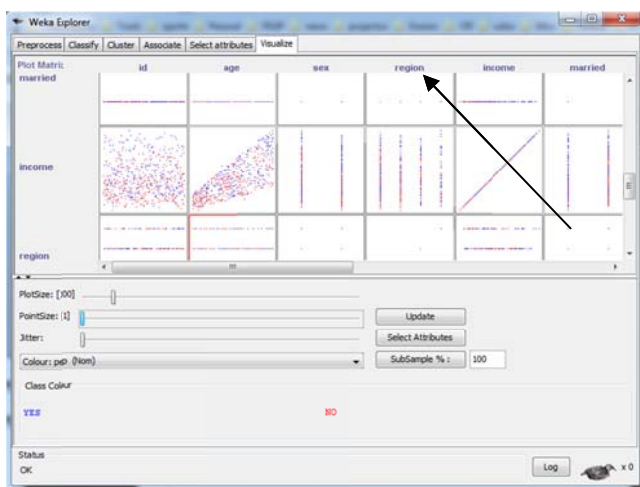
Analyze individual attributes



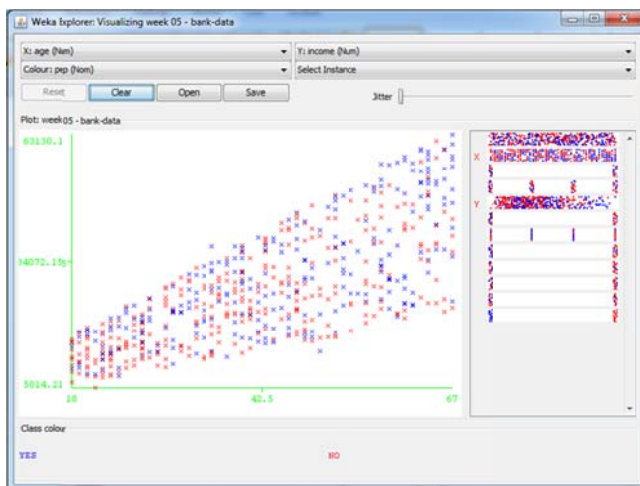
'Visualize all' button



Visualize tab

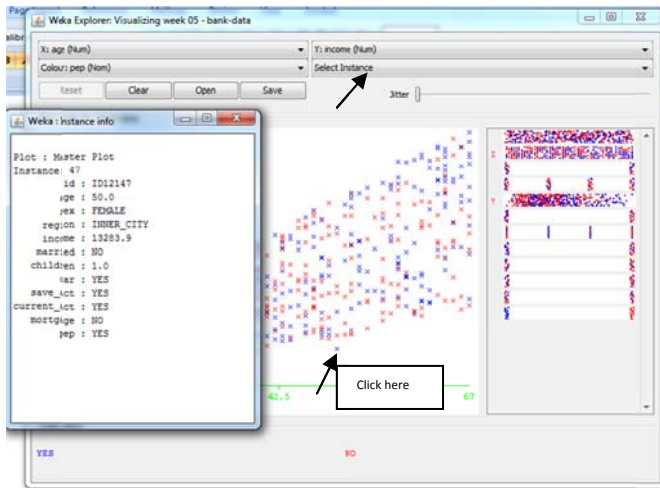


Double click on a scatter plot in the previous figure



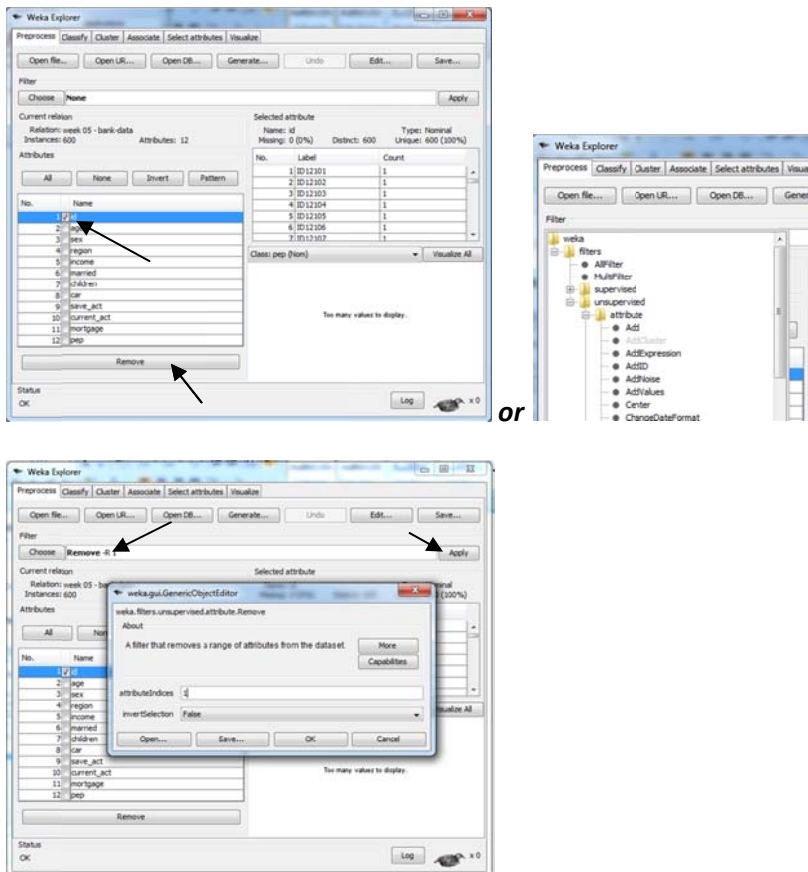
Double click on a scatter plot in the previous figure

Select an instance from the plot

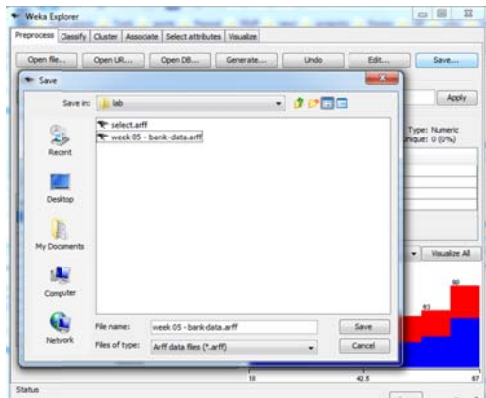


Try to select a rectangle and press the submit button

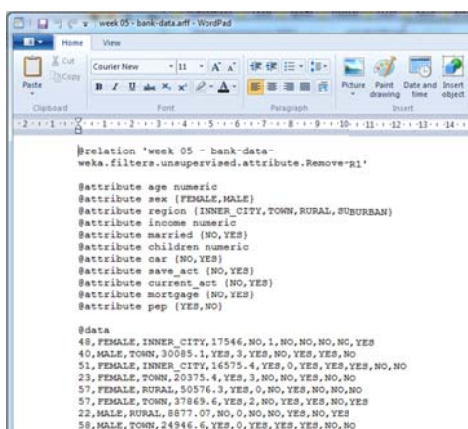
Remove ID attribute



Save file in weka format (.arff)



View file in WordPad



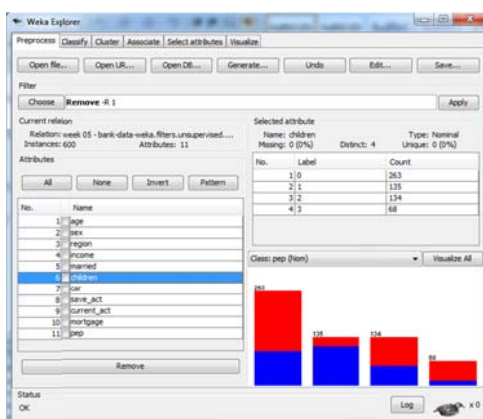
In order to apply association rules algorithms we cannot have continuous data. The next step will be to discretize the numeric attributes.

In WordPad edit the arff file and change

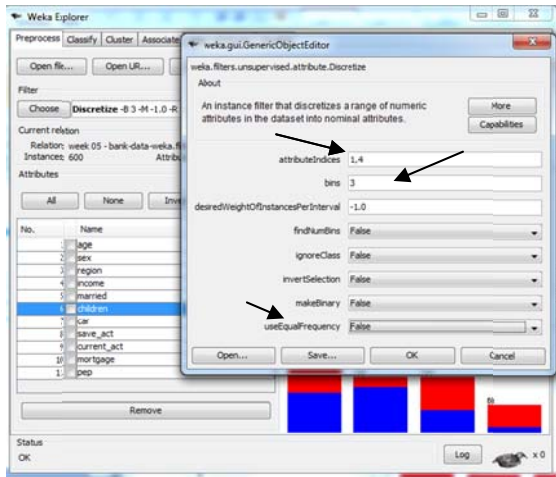
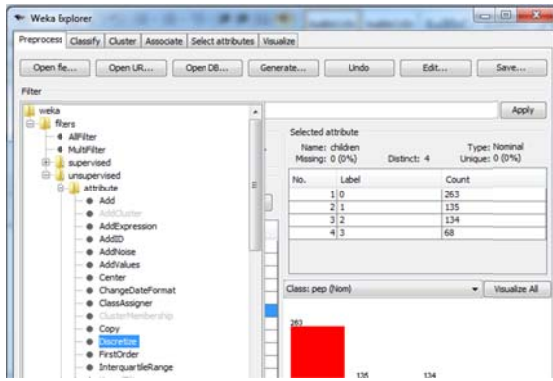
```
@attribute children numeric  
to
```

```
@attribute children {0,1,2,3}
```

Re-open the file.



Select the Discretize filter



To make variable readable open arff in wordpad and replace

'\(-inf-34.333333]\'' with 0_34

'\'(34.333333-50.666667]\'' with 35_51

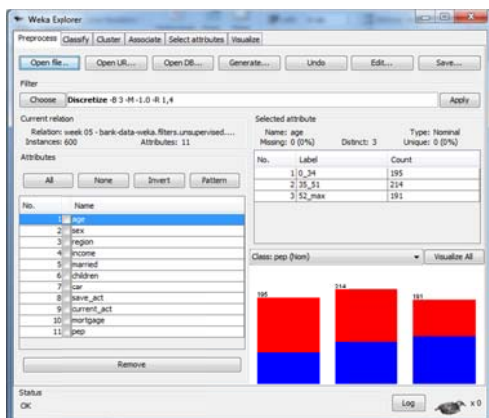
'\'(50.666667-inf)\'' with 52_max

'\(-inf-24386.173333]\'' with 0_24386

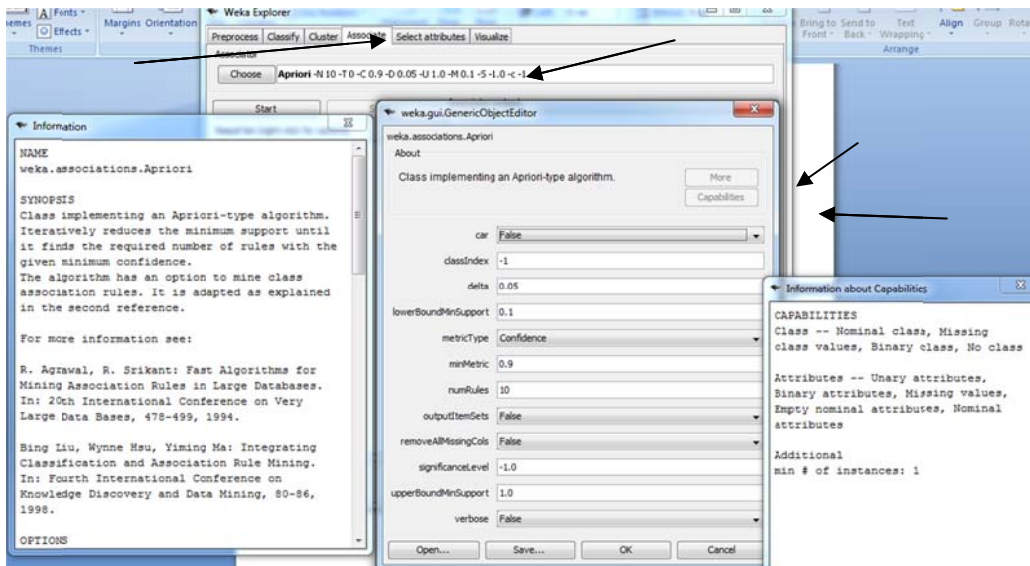
'\'(24386.173333-43758.136667]\'' with 24387_43758

, '\'(43758.136667-inf)\'' with 43759_max

Save the file and re-open it in weka.

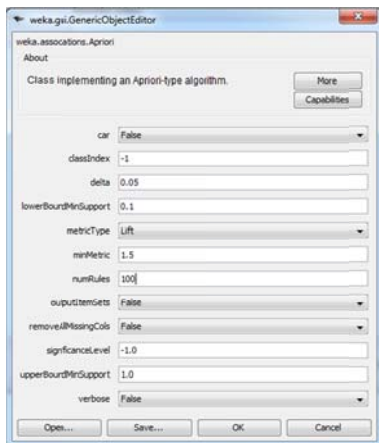


ASSOCIATION RULES MINING

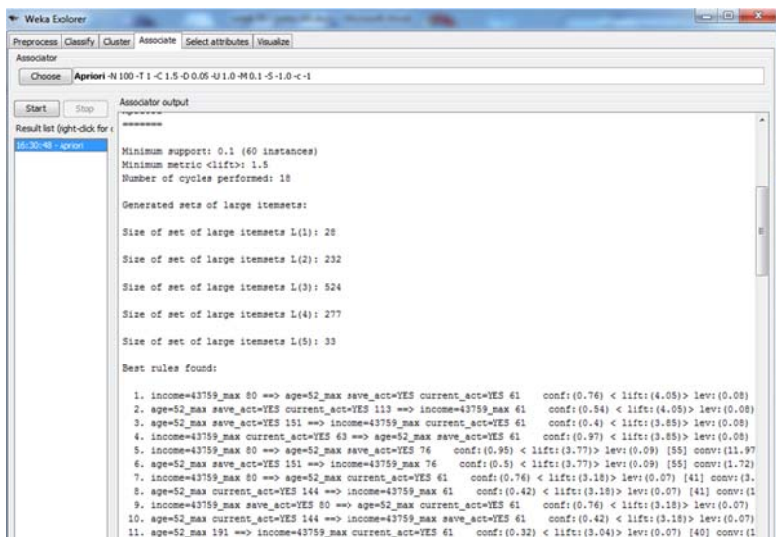


Read the information provided when the more button is clicked

Setting the apriori algorithm parameters

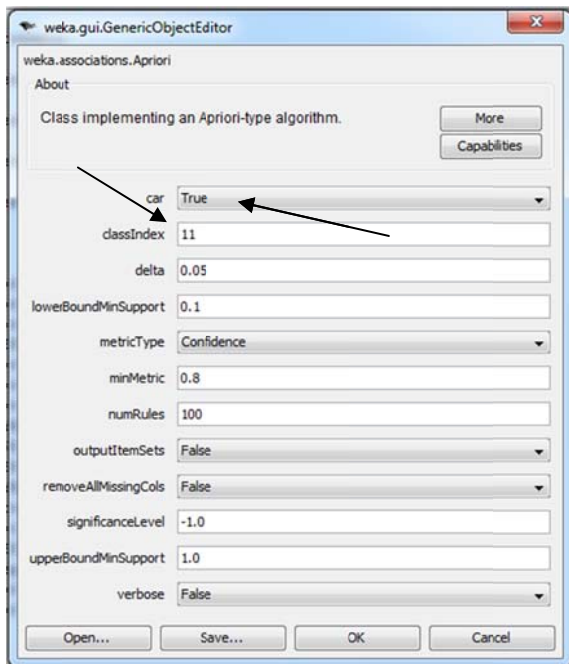


Press start button and check the results



You can (should) run the algorithm several times varying the parameters. You can save the results in a text file by right-clicking on the corresponding run item in the result list panel. Note that the results show the values for confidence, lift, leverage and conviction.

Try different settings



```
Associator output
Minimum support: 0.1 (60 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 42

Size of set of large itemsets L(2): 156

Size of set of large itemsets L(3): 125

Size of set of large itemsets L(4): 24

Best rules found:

1. married=YES children=0 save_act=YES current_act=YES 87 ==> pep=NO 80   conf:(0.92)
2. married=YES children=0 save_act=YES mortgage=NO 80 ==> pep=NO 73   conf:(0.91)
3. married=YES children=0 current_act=YES mortgage=NO 88 ==> pep=NO 80   conf:(0.91)
4. sex=FEMALE married=YES children=0 mortgage=NO 70 ==> pep=NO 63   conf:(0.9)
5. married=YES children=0 save_act=YES 119 ==> pep=NO 107   conf:(0.9)
6. married=YES children=0 mortgage=NO 116 ==> pep=NO 104   conf:(0.9)
7. married=YES children=0 car=NO mortgage=NO 67 ==> pep=NO 60   conf:(0.9)
8. children=1 save_act=YES current_act=YES 73 ==> pep=YES 63   conf:(0.86)
9. children=1 mortgage=NO 84 ==> pep=YES 71   conf:(0.85)
10. sex=FEMALE married=YES children=0 current_act=YES 71 ==> pep=NO 60   conf:(0.85)
11. children=1 save_act=YES 95 ==> pep=YES 80   conf:(0.84)
12. children=1 current_act=YES 101 ==> pep=YES 84   conf:(0.83)
13. married=YES children=1 89 ==> pep=YES 74   conf:(0.83)
14. sex=FEMALE married=YES children=0 94 ==> pep=NO 78   conf:(0.83)
15. children=1 135 ==> pep=YES 110   conf:(0.81)
16. married=YES children=0 car=NO current_act=YES 74 ==> pep=NO 60   conf:(0.81)
17. married=YES children=0 car=NO 100 ==> pep=NO 80   conf:(0.8)
```

The results now only give rules with pep as consequent and the values of confidence.