



Prof: Renato Rocha Souza

Aluno: Walter Dominguez

Trabalho sobre uso de ferramenta estatística ([WEKA v 3-9-0](#)) para identificar padrões ou relacionamentos comercialmente úteis em bancos de dados ou em outros repositórios de computador .

Resumo

Metodologia usada :

- roteiro para uso da ferramenta WEKA apresentado em https://github.com/rsouza/MMD/blob/master/activities/Exercise_Weka.md
- Os arquivos já estão preparados para serem lidos pelo WEKA (formato).
- O ciclo de mineração utilizado foi: Exploração (explorer), classificação (classifier), Agrupamento (cluster), Associação (associate), seleção de atributos e visualização.
- Algoritmo usado para classificação: **J48** é uma implementação de código aberto em [Java](#) do algoritmo C4.5 no aplicativo de mineração de dados [Weka](#)
- Algoritmo usado para agrupamento (cluster): simpleKMeans

Trab1

Arquivo fonte:

<https://github.com/rsouza/MMD/blob/master/datasets/bank.arff> :

Perfil de correntistas que aderiram ou não ao produto financeiro “PEP” (produto não descrito).

Objetivo

identificar grupos que compraram o produto ou não.

Exploração: São 600 correntista (instancias) com 11 atributos dos quais 274 compraram (Sim) e 326 não compraram (Não)

Classificação: A idade dos que tinham > 41 anos e 1 filho aderiram ao produto
Os que tinham faixa salarial mais baixa 3 mais de 41 anos aderiram ao produto.

Agrupamento(cluster): Campanha direcionada para os que tem > 41 anos e 1 filho
No topo da renda tem mais mulheres do que homens. A medida que a renda cresce fêmea vai diminuindo

Associação: Ao passar o nível de confiança de 0.9 para 0.7 aumentaram o número de regras

Atributos: Os melhores foram: renda (income), Casado(married) e com filhos (children)

Trab2

Arquivo fonte:

<https://github.com/rsouza/MMD/blob/master/datasets/credit-g.arff> :
Perfil de pessoas que já tiveram crédito ou não

Objetivo

O objetivo é identificar as características mais importantes para que o crédito no mercado seja bom ou ruim

Exploração: 1000 instâncias e 21 atributos

O status checking foi:

<0: 274, entre 0 e 200: 269, >=200: 63 e sem verificação: 394

Atributos: os melhores foram: checking_status, duration, credit_history

Trab3

Arquivo fonte:

<https://github.com/rsouza/MMD/blob/master/datasets/vote.arff> :
Perfil de votação dos congressistas

Objetivo

Identificar a posição política dos congressistas (Democrata ou Republicano) segundo atributos especificados na base de dados (áreas para aplicação de recursos) através da aplicação do algoritmo de árvore de decisão J48.

Exploração: São 435 instâncias (267 democratas, 168 republicanos)

A Base de Dados

Esta base de dados inclui votos para cada um dos Congressistas da

Câmara dos Representantes do EUA de 1984 em 16 categorias de votos identificados pelo CQA (Congressional Quarterly Almanac) para aplicação de recursos financeiros.

O CQA enumera nove tipos diferentes de votos

Atributos: Os melhores para caracterizar democratas e republicanos?

adoption-of-the-budget-resolution

physician-fee-freeze

immigration

synfuels-corporation-cutback

Trab4

Arquivo fonte:

<https://github.com/rsouza/MMD/blob/master/datasets/zoo.arff>

Perfil do zoológico

Objetivo

Identificar as características mais frequente dos animais

Exploração: Zoológico com 100 animais e 18 características de animais

Atributos: mais frequentes

animal

hair

feathers

milk

toothed

backbone

breathes

fins

legs

tail

Trab5

Arquivo fonte:

<https://github.com/rsouza/MMD/blob/master/datasets/weather.nominal.arff>

Perfil do tempo

Objetivo

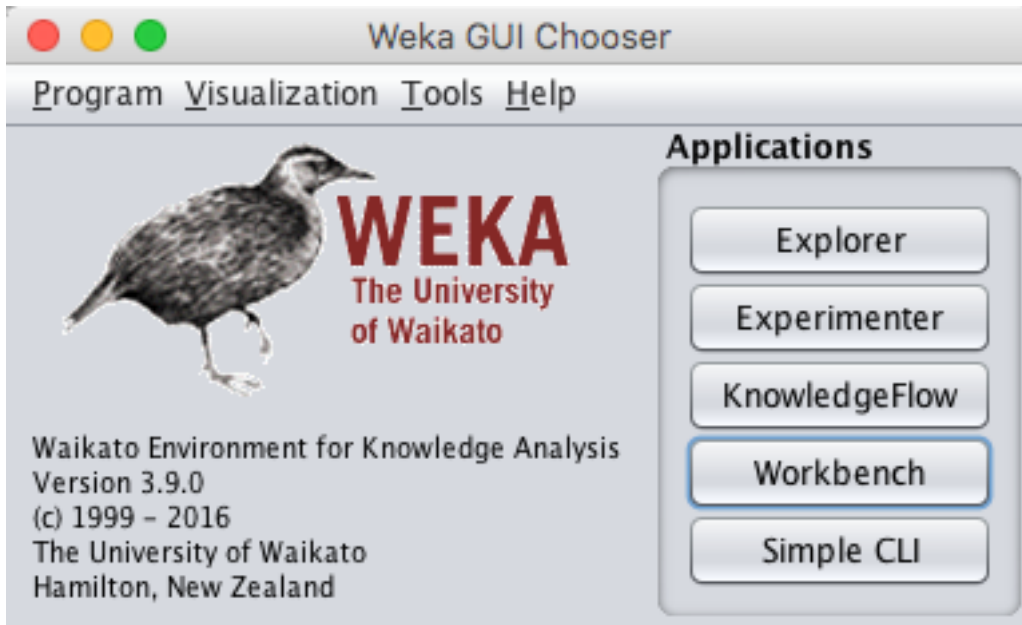
Identificar das características o que mais influencia o tempo

Exploração: 14 instancias e 3 atributos:

Atributos: Mais relevante Pluviosidade(Outlook)e humidade

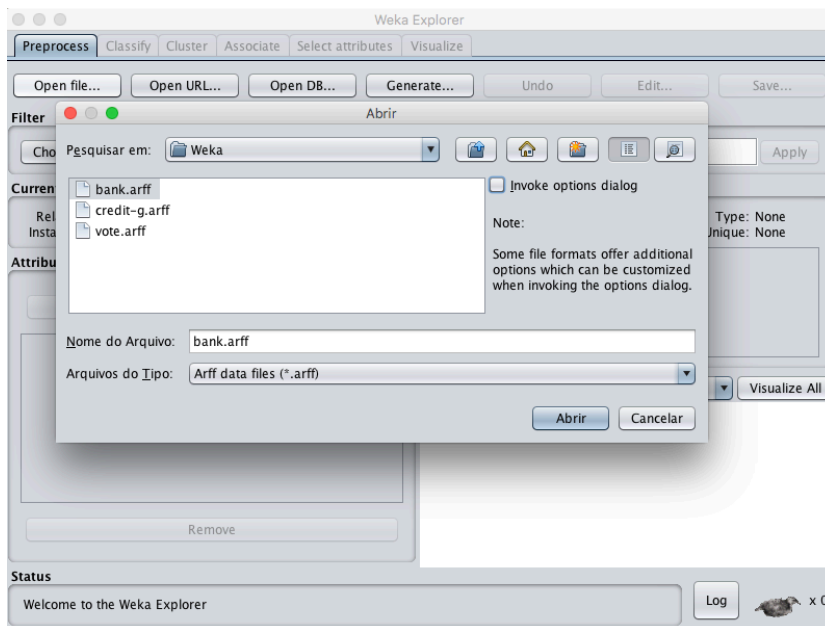
Exercícios com o Weka

1. Abra o Weka

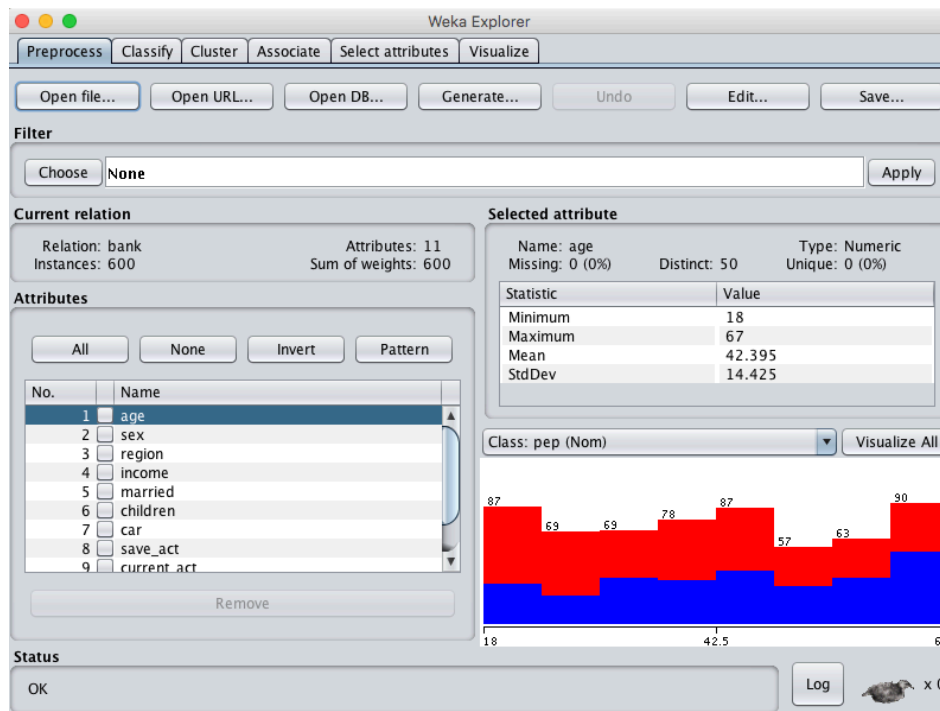


2. Escolha o módulo "Explorer"
 - 2.1. Carregue o arquivo "bank.arff" - este arquivo descreve o perfil de correntistas que aderiram ou não ao produto "PEP"

```
bank.arff
1 @relation bank
2
3 @attribute age numeric
4 @attribute sex {FEMALE,MALE}
5 @attribute region {INNER_CITY,TOWN,RURAL,SUBURBAN}
6 @attribute income numeric
7 @attribute married {NO,YES}
8 @attribute children {0,1,2,3}
9 @attribute car {NO,YES}
10 @attribute save_act {NO,YES}
11 @attribute current_act {NO,YES}
12 @attribute mortgage {NO,YES}
13 @attribute pep {YES,NO}
14
15 @data
16
17 48,FEMALE,INNER_CITY,17546,NO,1,NO,NO,NO,NO,YES
18 40,MALE,TOWN,30085.1,YES,3,YES,NO,YES,YES,NO
19 51,FEMALE,INNER_CITY,16575.4,YES,0,YES,YES,YES,NO,NO
20 23,FEMALE,TOWN,20375.4,YES,3,NO,NO,YES,NO,NO
21 57,FEMALE,RURAL,50576.3,YES,0,NO,YES,NO,NO,NO
22 57,FEMALE,TOWN,37869.6,YES,2,NO,YES,YES,NO,YES
23 22,MALE,RURAL,8877.07,NO,0,NO,NO,YES,NO,YES
24 58,MALE,TOWN,24946.6,YES,0,YES,YES,YES,NO,NO
25 37,FEMALE,SUBURBAN,25304.3,YES,2,YES,NO,NO,NO,NO
26 54,MALE,TOWN,24212.1,YES,2,YES,YES,YES,NO,NO
27 66,FEMALE,TOWN,59803.9,YES,0,NO,YES,YES,NO,NO
28 52,FEMALE,INNER_CITY,26658.8,NO,0,YES,YES,YES,YES,NO
29 44,FEMALE,TOWN,15735.8,YES,1,NO,YES,YES,YES,YES
30 66,FEMALE,TOWN,55204.7,YES,1,YES,YES,YES,YES,YES
31 36,MALE,RURAL,19474.6,YES,0,NO,YES,YES,YES,NO
32 38,FEMALE,INNER_CITY,22342.1,YES,0,YES,YES,YES,YES,NO
33 37,FEMALE,TOWN,17729.8,YES,2,NO,NO,NO,YES,NO
34 46,FEMALE,SUBURBAN,41016,YES,0,NO,YES,NO,YES,NO
35 62,FEMALE,INNER_CITY,26909.2,YES,0,NO,YES,NO,NO,YES
36 31,MALE,TOWN,22522.8,YES,0,YES,YES,YES,NO,NO
37 61,MALE,INNER_CITY,57880.7,YES,2,NO,YES,NO,NO,YES
38 50,MALE,TOWN,16497.3,YES,2,NO,YES,YES,NO,NO
39 54,MALE,INNER_CITY,38446.6,YES,0,NO,YES,YES,NO,NO
40 27,FEMALE,TOWN,15538.8,NO,0,YES,YES,YES,YES,NO
41 22,MALE,INNER_CITY,12640.3,NO,2,YES,YES,YES,NO,NO
42 56,MALE,INNER_CITY,41034,YES,0,YES,YES,YES,YES,NO
43 45,MALE,INNER_CITY,20809.7,YES,0,NO,YES,YES,YES,NO
44 39,FEMALE,TOWN,20114,YES,1,NO,NO,YES,NO,YES
45 39,FEMALE,INNER_CITY,29359.1,NO,3,YES,NO,YES,YES,NO
46 61,MALE,RURAL,24270.1,YES,1,NO,NO,YES,NO,YES
47 61,FEMALE,RURAL,22942.9,YES,2,NO,YES,YES,NO,NO
48 20,FEMALE,TOWN,16325.8,YES,2,NO,YES,NO,NO,NO
```



2.2. Explore os dados no "explorer"



Objetivo: Quantos correntistas escolheram ou não o produto por idade.

Visualização:

Variáveis de dados:

- 3 categoria nominal (yes/no, masc/fem, região)
- 1 quantitativo escala (idade e escolheram o produto e não escolheram) .
- 1 quantitativa (frequência de correntista).

Variáveis visual:

cor (azul-sim vermelho-não).

Método: Ocorrências de correntista por idade separando os que optaram pelo produto e os que não optaram.

Iteração: atributos (idade, sexo,...), algoritmo. Incluir, alterar e excluir uma instância, ordenar as instancias.

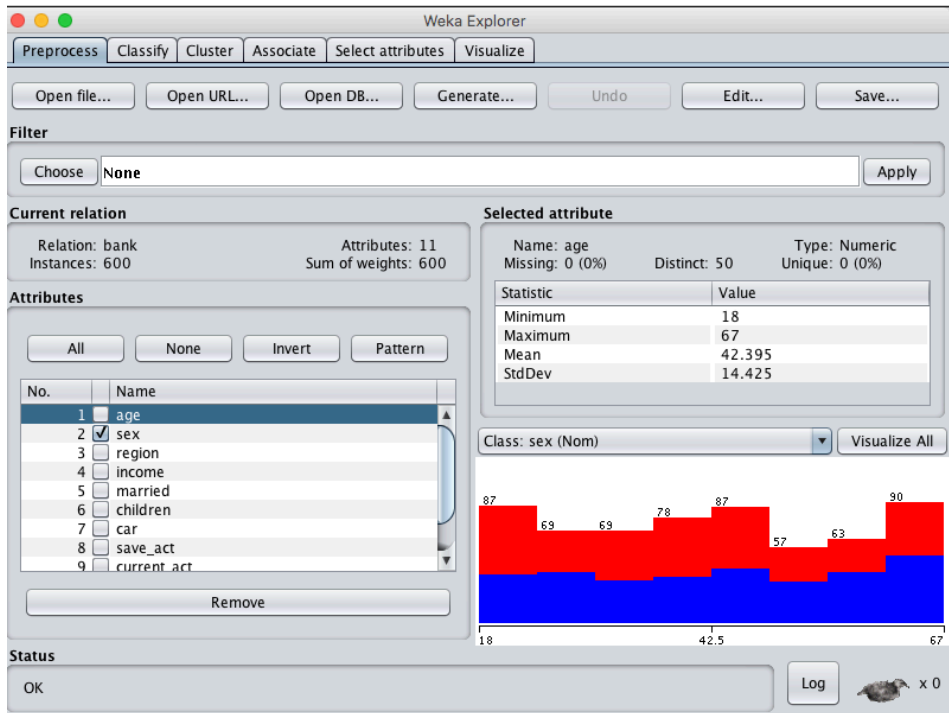
Descrição: Cada grupo de correntista por idade tem sua quantidade que escolheram e não escolheram. Contagem simples de correntista separando sim e não. Tendo o sim e não tamanhos proporcionais a quantidade encontrada. Saida em ordem de idade.

2.3. Visualize os dados com a opção "Edit" do Explorer

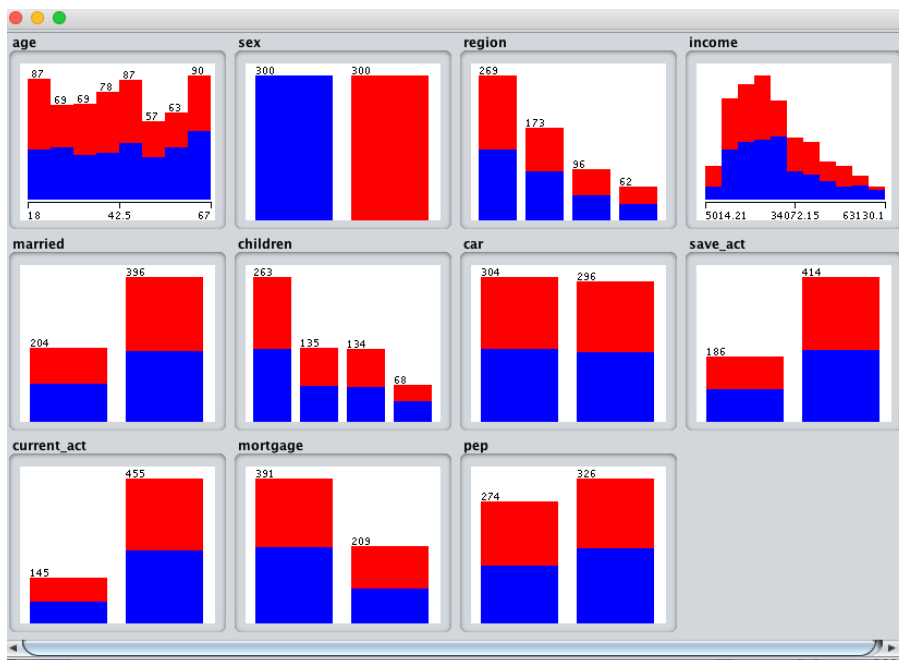
The screenshot shows a software interface with a 'Viewer' window open. The window title is 'Relation: bank'. The data is presented in a table with the following columns: No., 1: age, 2: sex, 3: region, 4: income, 5: married, 6: children, 7: car, 8: save_act. The data rows are sorted by age. A small bar chart is visible in the bottom right corner of the viewer window, showing two bars with values 63 and 90. The interface also includes a 'Filter' section on the left and a 'Status' section at the bottom.

No.	1: age	2: sex	3: region	4: income	5: married	6: children	7: car	8: save_act
1	48.0	FEM...	INNE...	17546.0	NO	1	NO	NO
2	40.0	MALE	TOWN	30085.1	YES	3	YES	NO
3	51.0	FEM...	INNE...	16575.4	YES	0	YES	YES
4	23.0	FEM...	TOWN	20375.4	YES	3	NO	NO
5	57.0	FEM...	RURAL	50576.3	YES	0	NO	YES
6	57.0	FEM...	TOWN	37869.6	YES	2	NO	YES
7	22.0	MALE	RURAL	8877.07	NO	0	NO	NO
8	58.0	MALE	TOWN	24946.6	YES	0	YES	YES
9	37.0	FEM...	SUBU...	25304.3	YES	2	YES	NO
...	54.0	MALE	TOWN	24212.1	YES	2	YES	YES
...	66.0	FEM...	TOWN	59803.9	YES	0	NO	YES
...	52.0	FEM...	INNE...	26658.8	NO	0	YES	YES
...	44.0	FEM...	TOWN	15735.8	YES	1	NO	YES
...	66.0	FEM...	TOWN	55204.7	YES	1	YES	YES
...	36.0	MALE	RURAL	19474.6	YES	0	NO	YES
...	38.0	FEM...	INNE...	22342.1	YES	0	YES	YES
...	37.0	FEM...	TOWN	17729.8	YES	2	NO	NO
...	46.0	FEM...	SUBU...	41016.0	YES	0	NO	YES
...	62.0	FEM...	INNE...	26909.2	YES	0	NO	YES
...	31.0	MALE	TOWN	22522.8	YES	0	YES	YES
...	61.0	MALE	INNE...	57880.7	YES	2	NO	YES
...	50.0	MALE	TOWN	16497.3	YES	2	NO	YES
...	54.0	MALE	INNE...	38446.6	YES	0	NO	YES

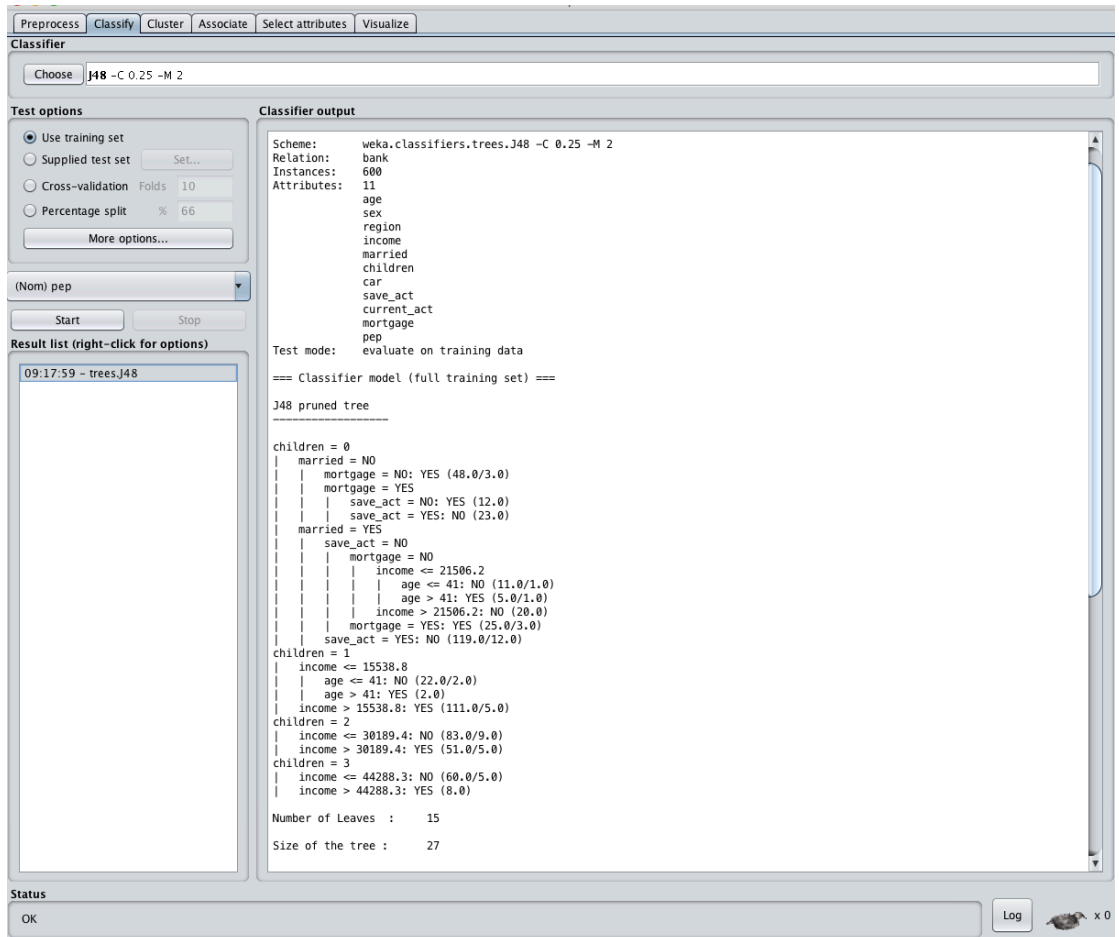
2.4. Varie as classes (quadrante inferior direito) e explore a opção "Visualize All"



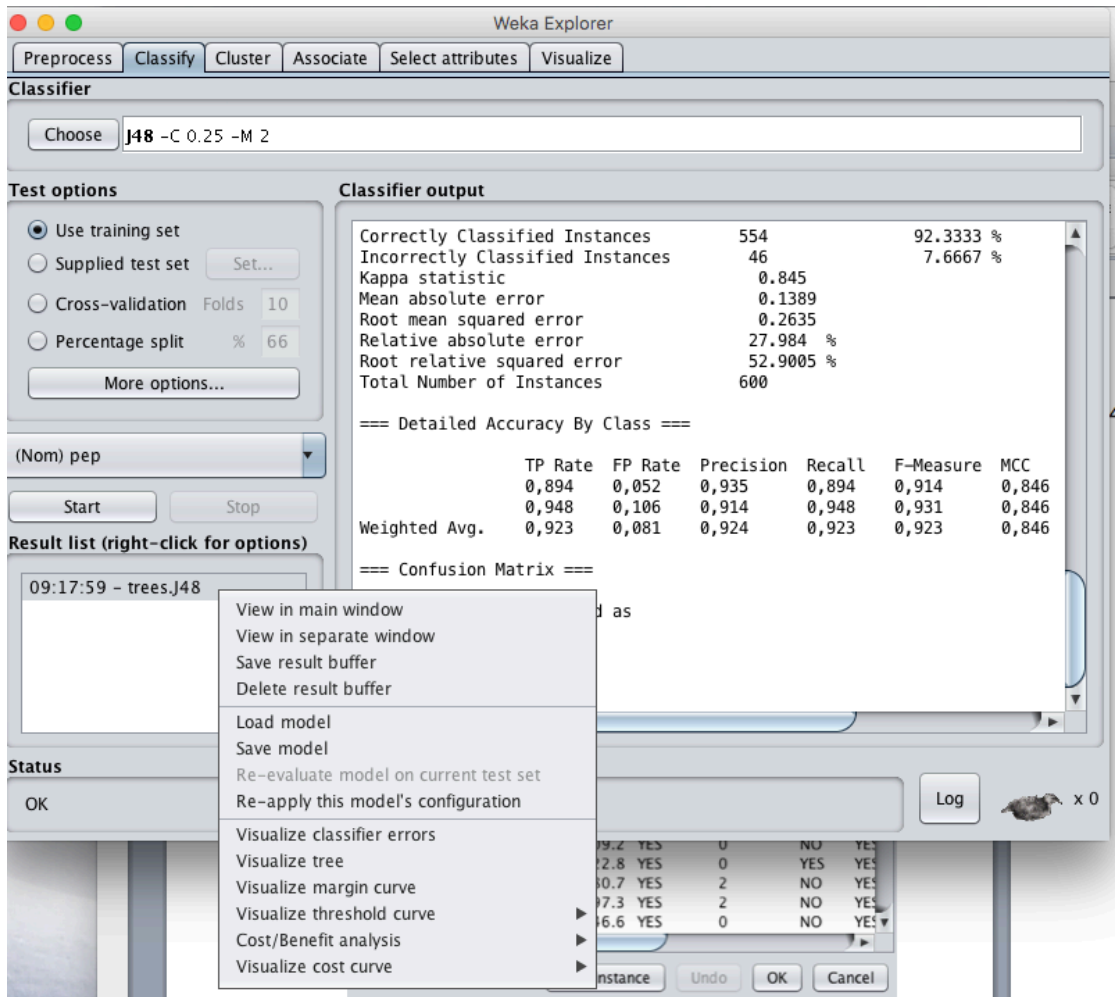
Classe Sexo:
Azul – feminino Vermelho – masculino



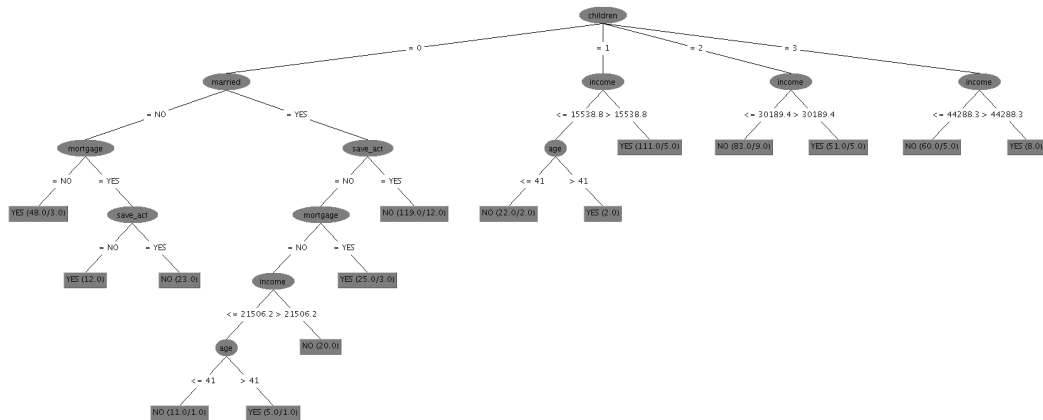
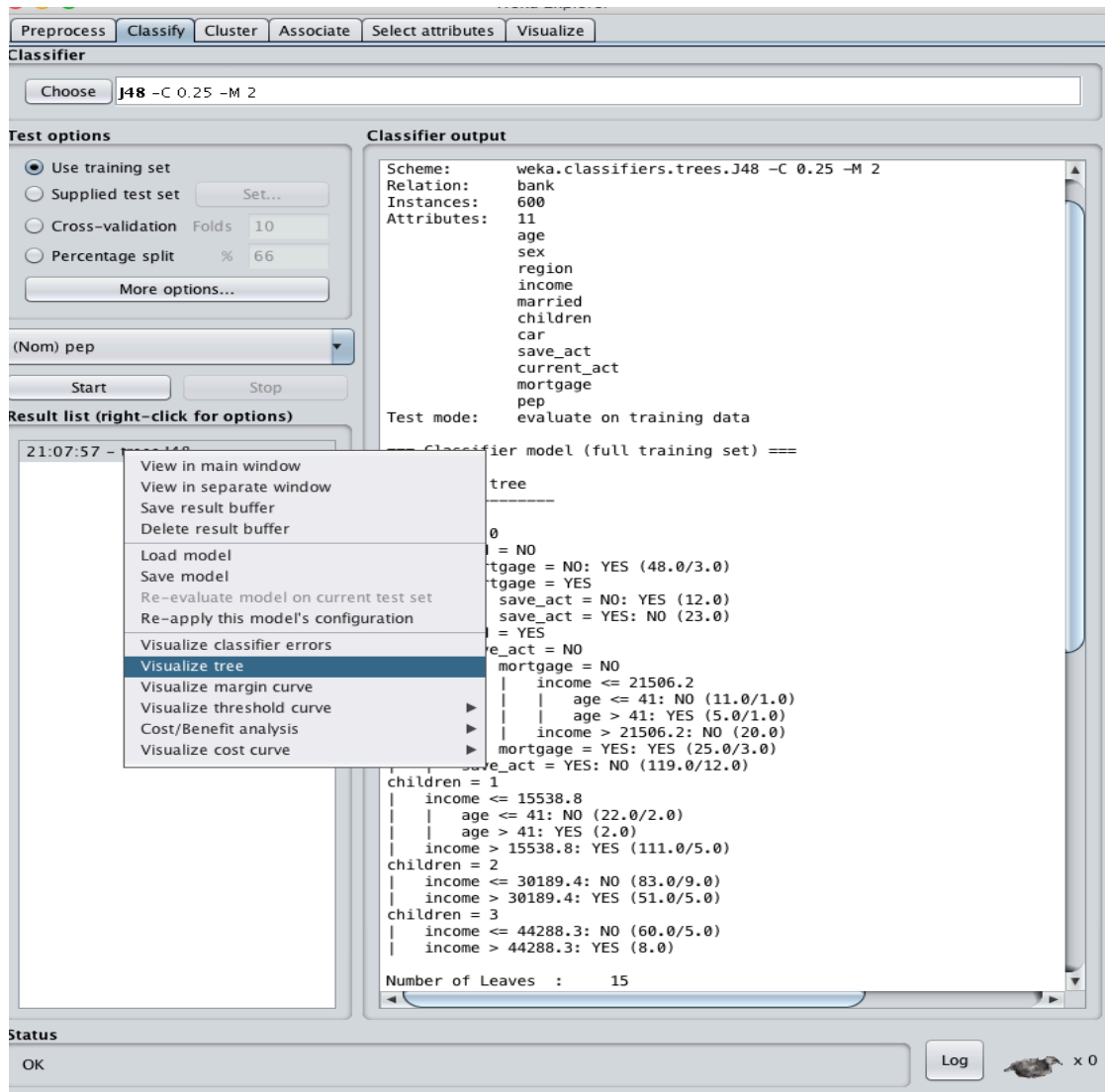
3. Vá para o módulo "Classify"
- 3.1. Monte uma árvore de decisão usando o algoritmo "J48"
- 3.2. Escolha "Use training set"



3.3. CLIque em "Start"



3.4. CLIque no resultado (no Result list) com o botão direito e escolha a opção "Visualize Tree"



3.5. Examine os perfis de correntistas que adotaram o

produto.

3.5.1. O que se pode dizer do produto em relação ao número de filhos?

A idade dos que tinham > 41 anos e 1 filho aderiram ao produto

3.5.2. E em relação ao salário (income)?

Os que tinham faixa salarial mais baixa e mais de 41 anos aderiram ao produto.

4. Vá para o módulo "Cluster"

4.1. Monte uma árvore de decisão usando o algoritmo "SimpleKMeans"

The screenshot shows the Weka Clusterer interface. The 'Cluster mode' section has 'Use training set' selected. The 'Clusterer output' section shows the following text:

```
save_act
current_act
mortgage
pep
Test mode: evaluate on training data

=== Clustering model (full training set) ===

KMeans
=====

Number of iterations: 4
Within cluster sum of squared errors: 2016.6752520938053

Initial starting points (random):
Cluster 0: 25, FEMALE, RURAL, 14505.3, NO, 3, NO, YES, YES, NO, NO
Cluster 1: 61, FEMALE, RURAL, 22942.9, YES, 2, NO, YES, YES, NO, NO

Missing values globally replaced with mean/mode

cluster centroids:
Attribute      Cluster#
              0          1
Income         (600.0)    (254.0)    (346.0)
=====
Age            42.395     38.1142    45.5376
Sex            FEMALE     FEMALE     MALE
MARRIED        YES         NO          YES
INNER_CITY    27524.0312 24264.972 29916.5198
children       0           0           0
car            NO          NO          YES
save_act       YES         YES         YES
current_act    YES         YES         YES
mortgage       NO          NO          NO
pep            NO          YES         NO

Time taken to build model (full training data) : 0 seconds

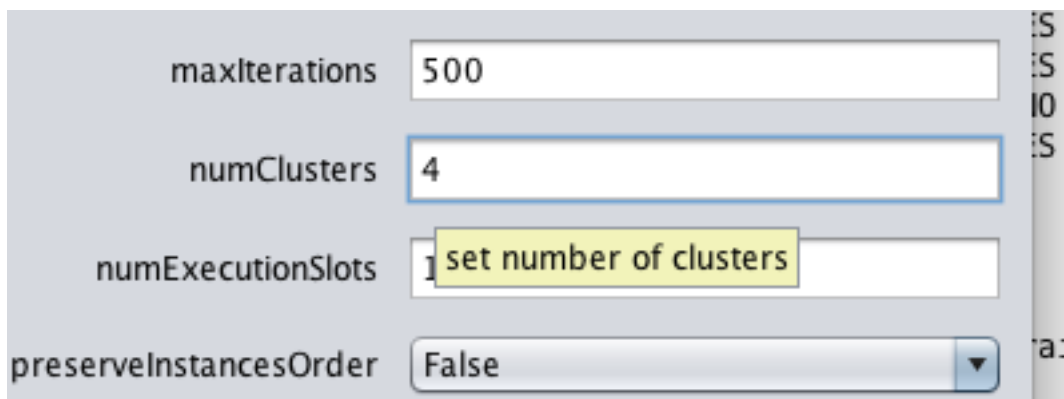
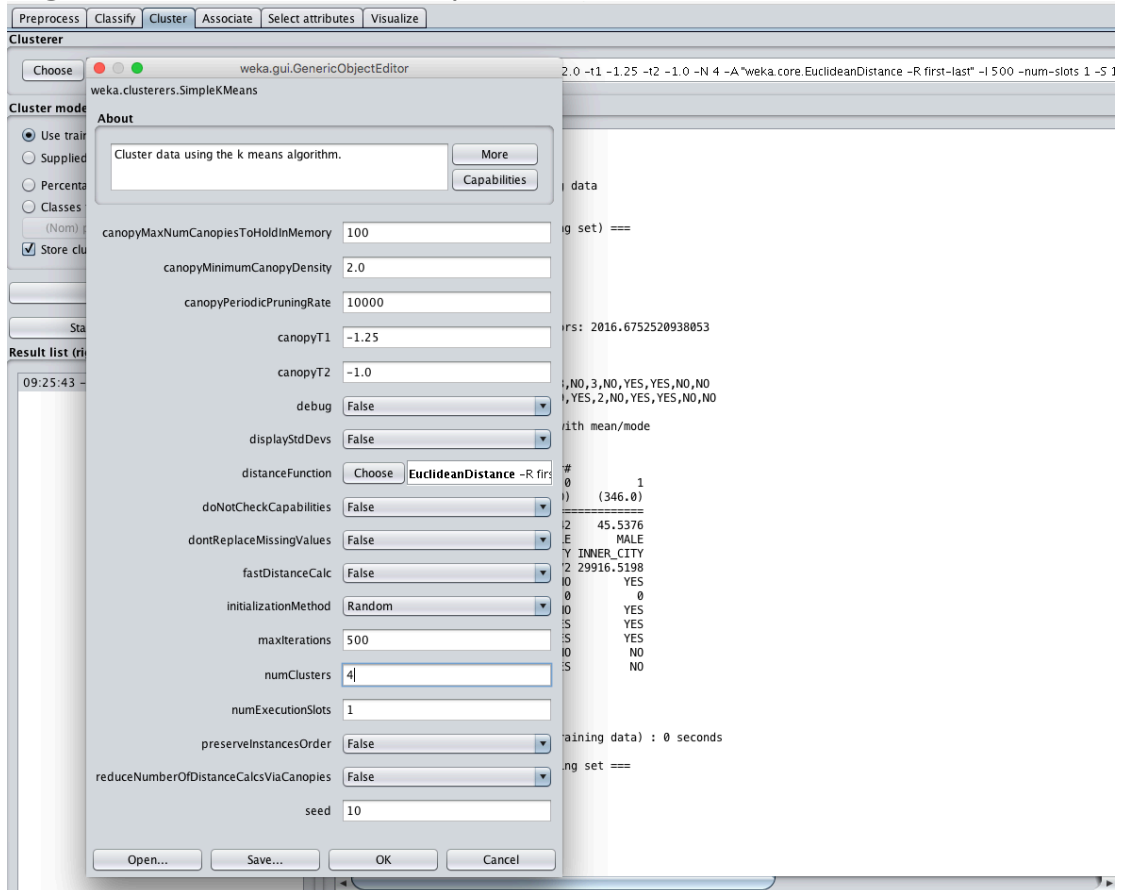
=== Model and evaluation on training set ===

Clustered Instances
0      254 ( 42%)
1      346 ( 58%)
```

A context menu is open over the 'Result list' area, showing options like 'View in main window', 'View in separate window', 'Save result buffer', 'Delete result buffer', 'Load model', 'Save model', 'Re-evaluate model on current test set', 'Re-apply this model's configuration', 'Visualize cluster assignments', 'Visualize cluster assignments', and 'Visualize tree'.

Não permite

4.2. Configure o algoritmo para que sejam encontrados 4 clusters (Configure clicando em cima do nome do algoritmo em "Choose")



4.3. Escolha "Use training set"

4.4. Clique em "Start"

The screenshot shows the Weka Clusterer window with the SimpleKMeans algorithm selected. The interface includes a menu bar (Preprocess, Classify, Cluster, Associate, Select attributes, Visualize), a command line, and two main panels: Cluster mode and Clusterer output.

Cluster mode:

- Use training set (selected)
- Supplied test set: Set...
- Percentage split: % 66
- Classes to clusters evaluation: (Nom) pep
- Store clusters for visualization (checked)
- Ignore attributes: []
- Start / Stop buttons

Clusterer output:

```
==== Clustering model (full training set) ====  
kMeans  
=====  
Number of iterations: 7  
Within cluster sum of squared errors: 1710.6380875337327  
Initial starting points (random):  
Cluster 0: 25, FEMALE, RURAL, 14505.3, NO, 3, NO, YES, YES, NO, NO  
Cluster 1: 61, FEMALE, RURAL, 22942.9, YES, 2, NO, YES, YES, NO, NO  
Cluster 2: 54, FEMALE, INNER_CITY, 31095.6, YES, 2, NO, NO, YES, NO, YES  
Cluster 3: 36, FEMALE, TOWN, 26920.8, YES, 0, NO, NO, YES, NO, NO  
Missing values globally replaced with mean/mode  
Final cluster centroids:  
Attribute      Full Data      Cluster#  
                (600.0)      0          1          2          3  
-----  
age            42.395        34.4724    48.3636    43.8742    41.1093  
sex            FEMALE       MALE      FEMALE     MALE      FEMALE  
region        INNER_CITY   INNER_CITY INNER_CITY INNER_CITY TOWN  
income        27524.0312   21738.0819 32340.1644 29396.4921 25341.9421  
married       YES          NO        YES        YES        YES  
children      0            0         2          1          0  
car           NO          NO        YES        NO        NO  
save_act      YES         YES       YES        YES        YES  
current_act   YES         YES       YES        YES        YES  
mortgage      NO          NO        NO         NO        NO  
pep           NO          NO        NO         YES        NO
```

Time taken to build model (full training data) : 0 seconds

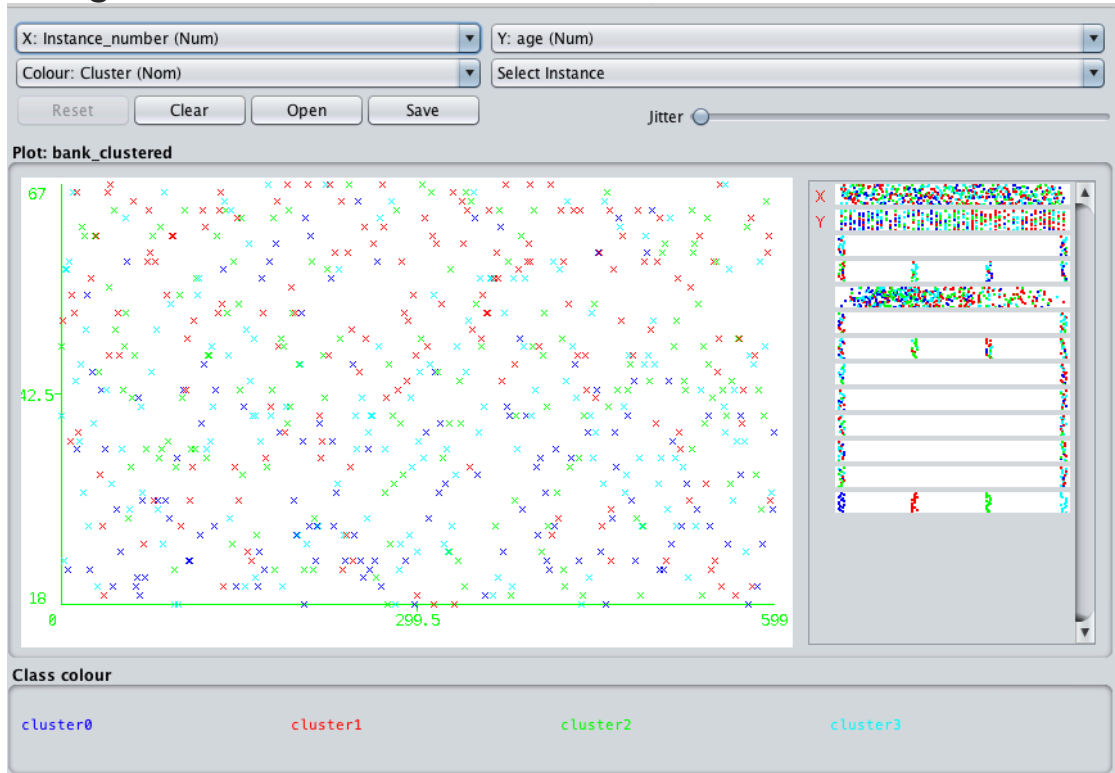
```
==== Model and evaluation on training set ====  
Clustered Instances  
0      127 ( 21%)  
1      165 ( 28%)  
2      151 ( 25%)  
3      157 ( 26%)
```

4.5. Examine na janela principal os centróides dos clusters formados.

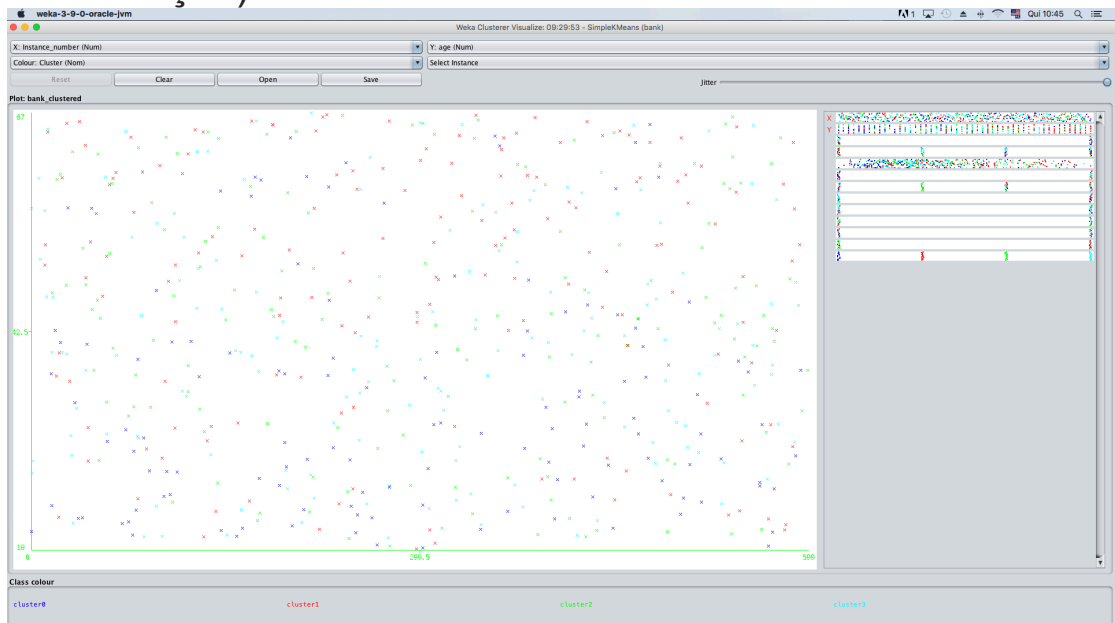
4.5.1. Há grupos específicos que devem ser alvos de campanhas futuras para contratação do produto PEP?

Sim. Os que tem > 41 anos e 1 filho

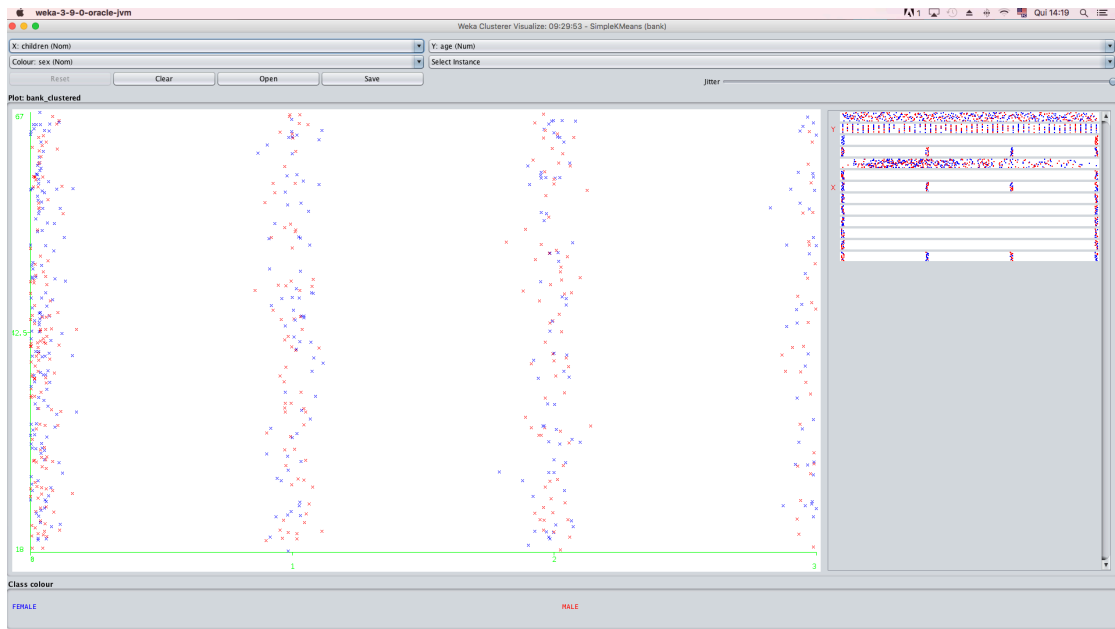
- 4.6. Clique no resultado (no Result list) com o botão direito e escolha a opção "Visualize Cluster Assignments"



- 4.7. Modifique o parâmetro "Jitter" para inserir uma separação artificial entre os pontos (melhora a visualização)



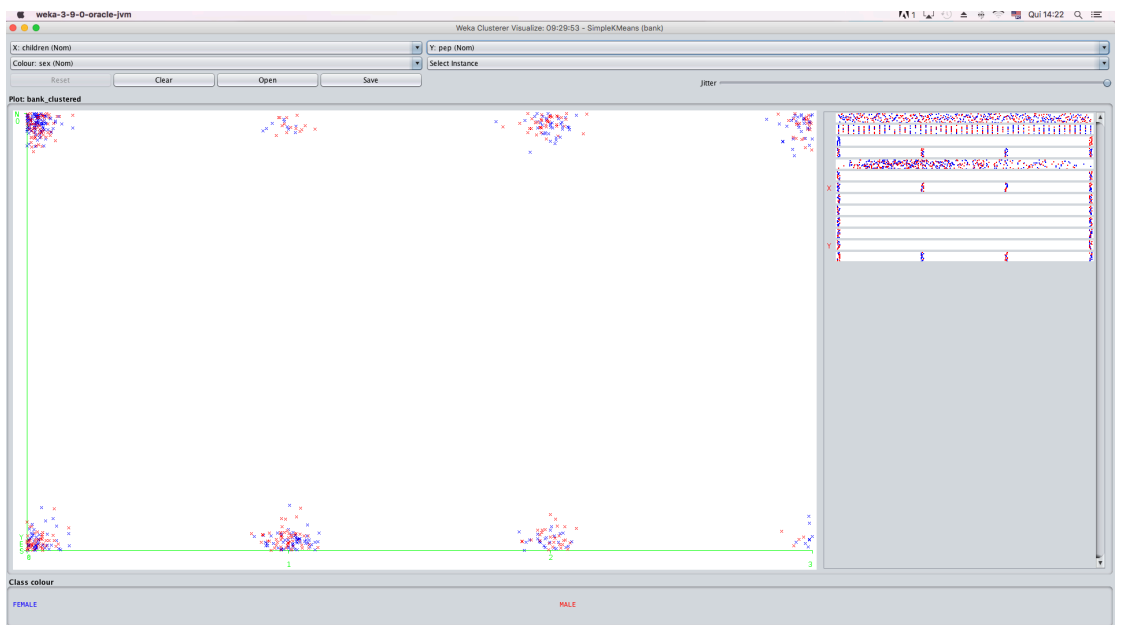
4.8. Alterne as variáveis que estão associadas à cor, ao eixo x e ao eixo y.



X:Children Colour:sex Y:age

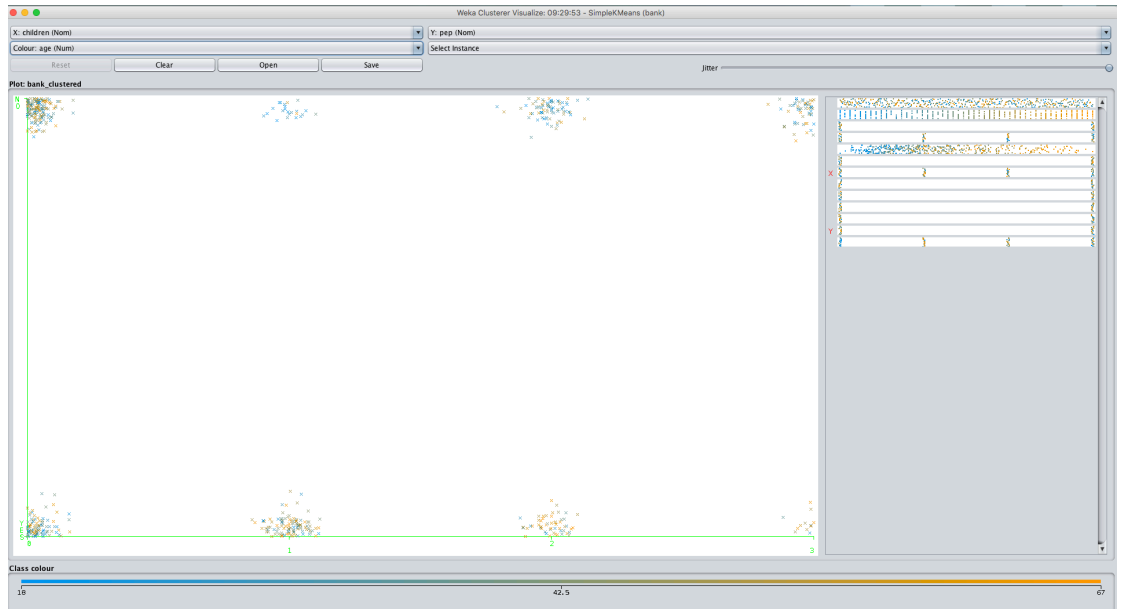
4.8.1. Experimente X:Cluster, Colour:Sex e Y:PEP. ---> Gênero foi um critério para montagem dos clusters?

Sim

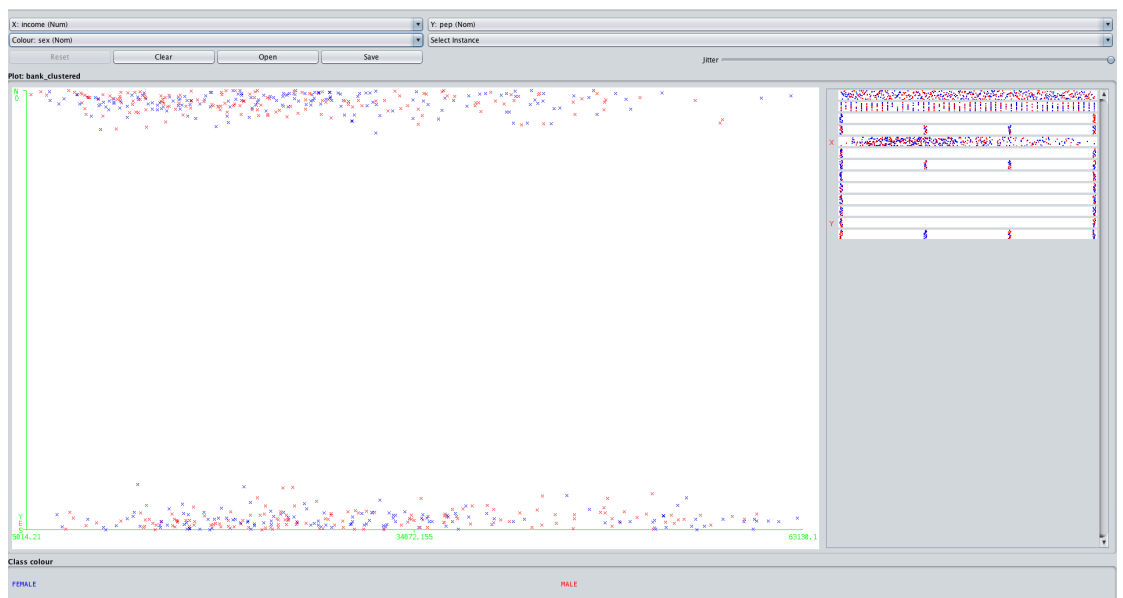


4.8.2. Varie o parâmetro em "Colour" e verifique outros critérios de formação de clusters.

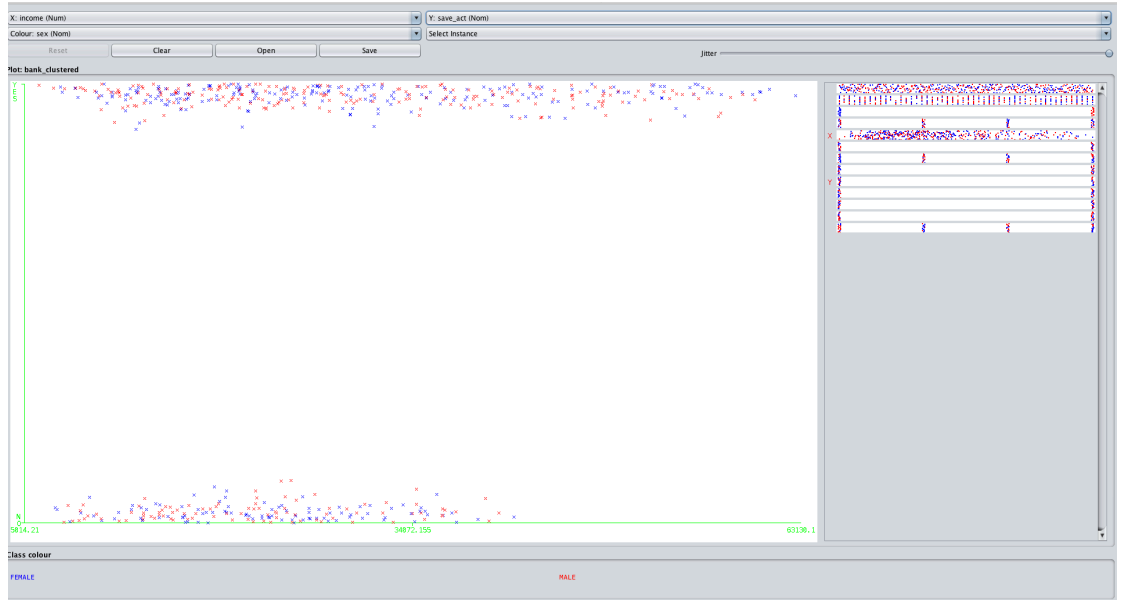
Colour: age



4.8.3. Experimente X:Income, Colour:Sex e Y:PEP. ---> Qual gênero com alta renda é mais propenso a contratar o PEP? E com baixa renda? alta renda → Femele baixa renda → nada a dizer



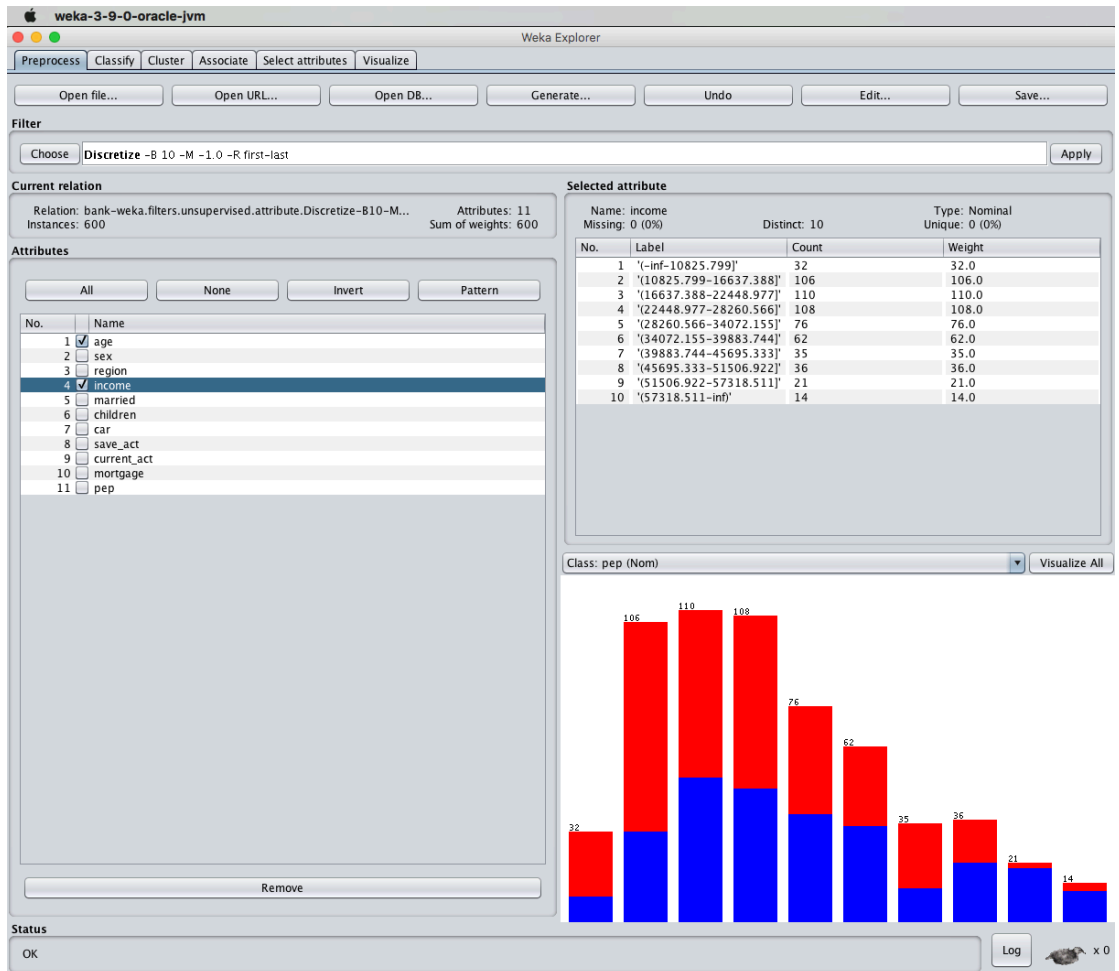
- 4.8.4. Experimente X:Income, Colour:Sex e Y:Save-Act.
---> o que se pode dizer das pessoas que poupam mais?



5. Volte ao módulo "Preprocess"
- 5.1. Escolha e aplique um filtro de discretização aos dados:
(Filter/Choose/filters/unsupervised/attribute/Discretize)
- 5.2. Marque os dados numéricos (age e income)
- 5.3. Examine os parâmetros de configuração do filtro. 10 bins é apropriado?

5.4. Aplique o filtro aos dados escolhidos (age e income) - Clique em "Apply".

No topo da renda tem mais mulheres do que homens. A medida que a renda cresce fêmea vai diminuindo

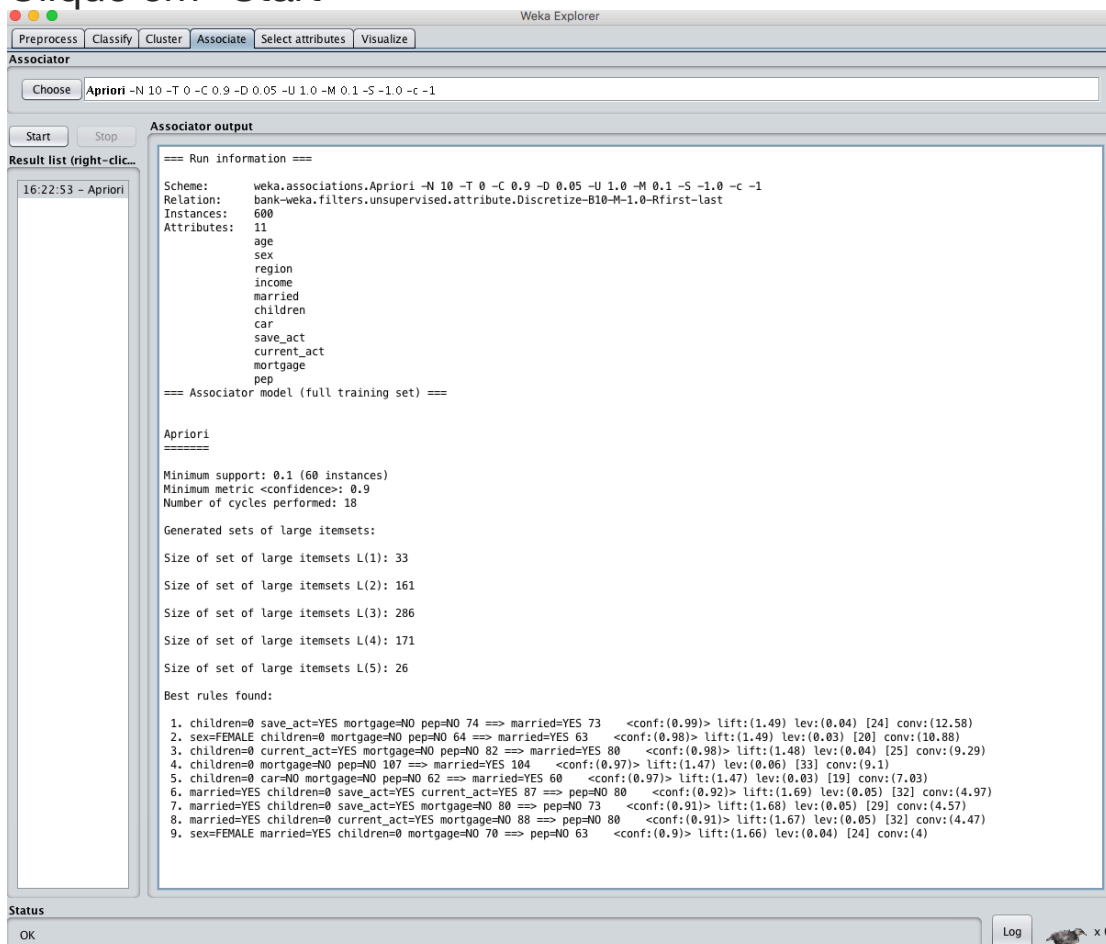


5.5. Examine os dados outrora contínuos discretizados.

6. Vá para o módulo "Associate"

6.1. Escolha o algoritmo de associação "APriori"

6.2. Clique em "Start"



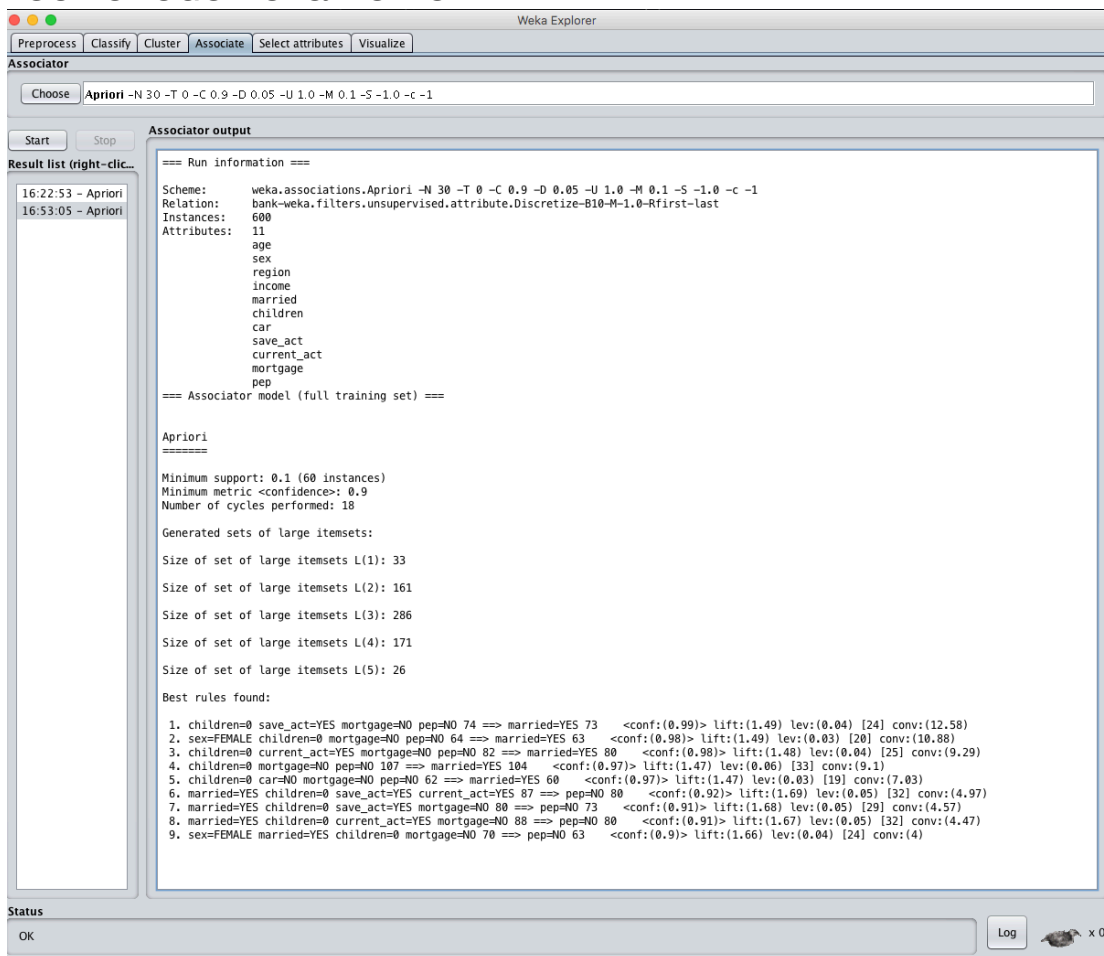
6.3. Examine as melhores regras encontradas em "Associator Output"

Best rules found:

1. children=0 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=0 mortgage=NO pep=NO 64 ==> married=YES 63 <conf:(0.98)> lift:(1.49) lev:(0.03) [20] conv:(10.88)
3. children=0 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=0 mortgage=NO pep=NO 107 ==> married=YES 104 <conf:(0.97)> lift:(1.47) lev:(0.06) [33] conv:(9.1)
5. children=0 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=0 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=0 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=0 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=0 mortgage=NO 70 ==> pep=NO 63 <conf:(0.9)> lift:(1.66) lev:(0.04) [24] conv:(4)

1. children=0 save_act=YES mortgage=NO
pep=NO 74 ==> married=YES 73 <conf:(0.99)>
lift:(1.49) lev:(0.04) [24] conv:(12.58)

6.4. Configure o filtro: aumente o número de regras para "30" e rode novamente.



Best rules found:

1. children=0 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=0 mortgage=NO pep=NO 64 ==> married=YES 63 <conf:(0.98)> lift:(1.49) lev:(0.03) [20] conv:(10.88)
3. children=0 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=0 mortgage=NO pep=NO 107 ==> married=YES 104 <conf:(0.97)> lift:(1.47) lev:(0.06) [33] conv:(9.1)
5. children=0 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=0 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=0 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=0 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=0 mortgage=NO 70 ==> pep=NO 63 <conf:(0.9)> lift:(1.66) lev:(0.04) [24] conv:(4)

6.4.1. Houve mudança significativa?

Não

6.5.

Configure o filtro: diminua a confiança para 0.7 e rode novamente.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -N 30 -T 0 -C 0.7 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start Stop

Result list (right-click...)

16:22:53 - Apriori
16:53:05 - Apriori
17:01:45 - Apriori

Associator output

==== 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): 16
Size of set of large itemsets L(2): 47
Size of set of large itemsets L(3): 16

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:(1.1)
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. sex=MALE 300 ==> current_act=YES 225 <conf:(0.75)> lift:(0.99) lev:(-0) [-2] conv:(0.95)
21. pep=NO 326 ==> current_act=YES 244 <conf:(0.75)> lift:(0.99) lev:(-0.01) [-3] conv:(0.95)
22. married=YES car=NO 202 ==> current_act=YES 151 <conf:(0.75)> lift:(0.99) lev:(-0) [-2] conv:(0.94)
23. save_act=YES pep=NO 235 ==> married=YES 175 <conf:(0.74)> lift:(1.13) lev:(0.03) [19] conv:(1.31)
24. married=YES save_act=YES 277 ==> current_act=YES 206 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-4] conv:(0.93)
25. car=YES 296 ==> current_act=YES 220 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-4] conv:(0.93)
26. pep=NO 326 ==> married=YES 242 <conf:(0.74)> lift:(1.12) lev:(0.04) [26] conv:(1.3)
27. married=YES 396 ==> current_act=YES 293 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-7] conv:(0.92)
28. mortgage=YES 209 ==> current_act=YES 154 <conf:(0.74)> lift:(0.97) lev:(-0.01) [-4] conv:(0.9)
29. current_act=YES pep=NO 244 ==> save_act=YES 179 <conf:(0.73)> lift:(1.06) lev:(0.02) [10] conv:(1.15)
30. married=YES pep=NO 242 ==> current_act=YES 177 <conf:(0.73)> lift:(0.96) lev:(-0.01) [-6] conv:(0.89)

Status

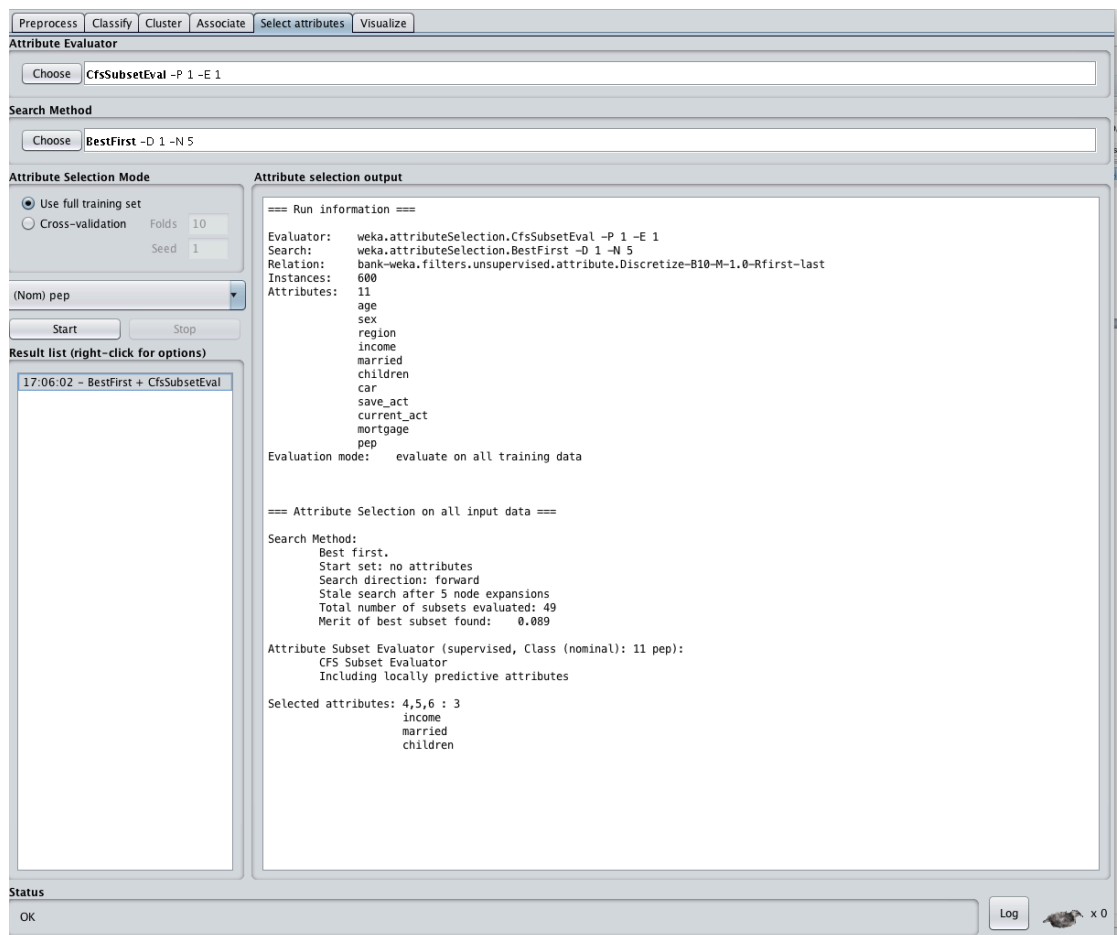
OK Log x 0

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:(1.1)
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. sex=MALE 300 ==> current_act=YES 225 <conf:(0.75)> lift:(0.99) lev:(-0) [-2] conv:(0.95)
21. pep=NO 326 ==> current_act=YES 244 <conf:(0.75)> lift:(0.99) lev:(-0.01) [-3] conv:(0.95)
22. married=YES car=NO 202 ==> current_act=YES 151 <conf:(0.75)> lift:(0.99) lev:(-0) [-2] conv:(0.94)
23. save_act=YES pep=NO 235 ==> married=YES 175 <conf:(0.74)> lift:(1.13) lev:(0.03) [19] conv:(1.31)
24. married=YES save_act=YES 277 ==> current_act=YES 206 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-4] conv:(0.93)
25. car=YES 296 ==> current_act=YES 220 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-4] conv:(0.93)
26. pep=NO 326 ==> married=YES 242 <conf:(0.74)> lift:(1.12) lev:(0.04) [26] conv:(1.3)
27. married=YES 396 ==> current_act=YES 293 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-7] conv:(0.92)
28. mortgage=YES 209 ==> current_act=YES 154 <conf:(0.74)> lift:(0.97) lev:(-0.01) [-4] conv:(0.9)
29. current_act=YES pep=NO 244 ==> save_act=YES 179 <conf:(0.73)> lift:(1.06) lev:(0.02) [10] conv:(1.15)
30. married=YES pep=NO 242 ==> current_act=YES 177 <conf:(0.73)> lift:(0.96) lev:(-0.01) [-6] conv:(0.89)

6.5.1. Houve mudança significativa?
Sim. Aumentaram o número de regras (30)

- 7. Vá para o módulo "Attributes"
- 7.1. Escolha o algoritmo de avaliação de atributos "CfsSubsetEval"
- 7.2. Escolha o algoritmo de busca de atributos "BestFirst"
- 7.3. Clique em "Start"



7.4. Examine as melhores regras encontradas em "Attribute Selection Output"

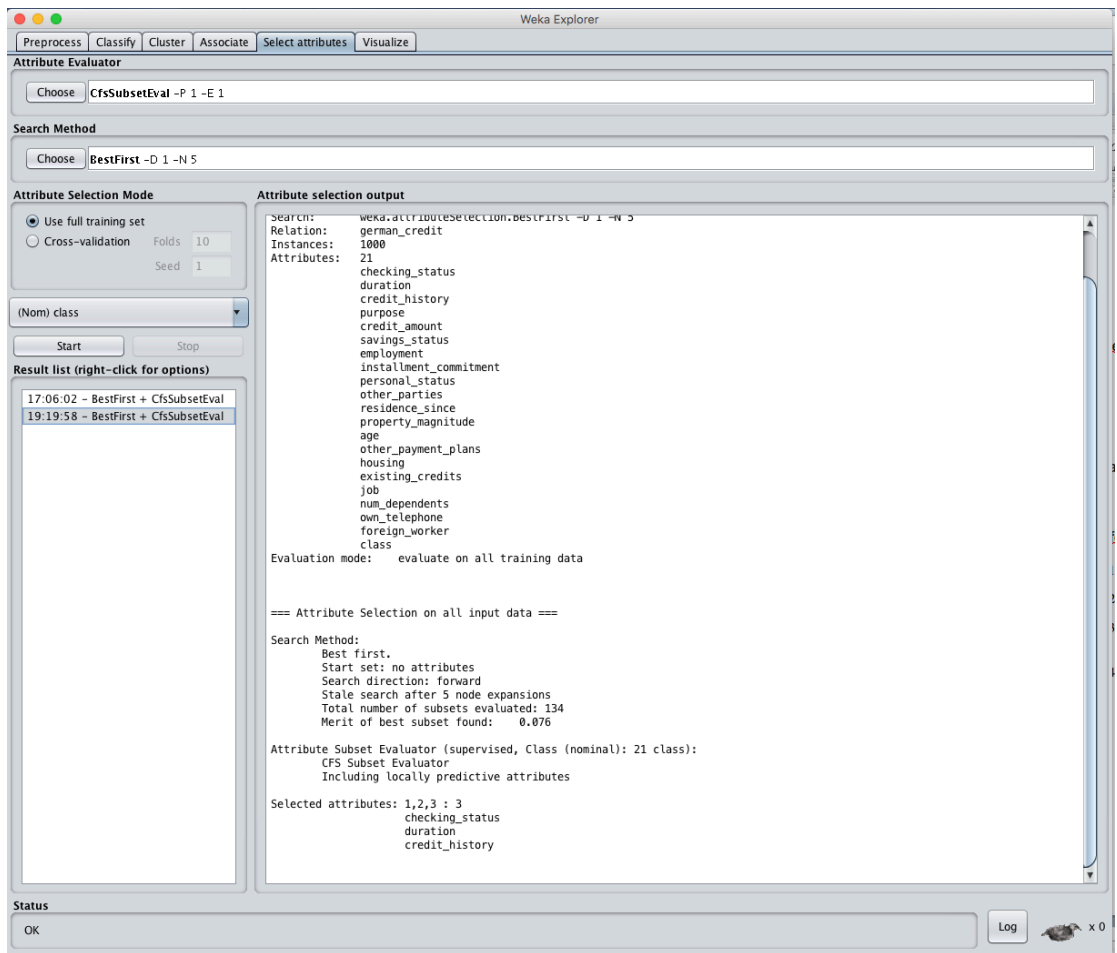
7.4.1. Quais são os atributos mais importantes para este dataset? Porquê?
4,5,6 : 3

income
married
children

8. Repita os passos para os datasets:

8.1. credit-g.arff (Quais características são importantes para que o crédito no mercado seja bom ou ruim? - parâmetro "class")

```
bank.arff x credit-g.arff x
279 @relation german_credit
280 @attribute checking_status { '<0', '0<=X<200', '>=200', 'no checking'}
281 @attribute duration real
282 @attribute credit_history { 'no credits/all paid', 'all paid', 'existing paid', 'delayed
previously', 'critical/other existing credit'}
283 @attribute purpose { 'new car', 'used car', 'furniture/equipment', 'radio/tv', 'domestic appliance',
repairs, education, vacation, retraining, business, other}
284 @attribute credit_amount real
285 @attribute savings_status { '<100', '100<=X<500', '500<=X<1000', '>=1000', 'no known savings'}
286 @attribute employment { unemployed, '<1', '1<=X<4', '4<=X<7', '>=7'}
287 @attribute installment_commitment real
288 @attribute personal_status { 'male div/sep', 'female div/dep/mar', 'male single', 'male mar/wid',
'female single'}
289 @attribute other_parties { none, 'co applicant', guarantor}
290 @attribute residence_since real
291 @attribute property_magnitude { 'real estate', 'life insurance', 'car', 'no known property'}
292 @attribute age real
293 @attribute other_payment_plans { bank, stores, none}
294 @attribute housing { rent, own, 'for free'}
295 @attribute existing_credits real
296 @attribute job { 'unemp/unskilled non res', 'unskilled resident', skilled, 'high qualif/self
emp/mgmt'}
297 @attribute num_dependents real
298 @attribute own_telephone { none, yes}
299 @attribute foreign_worker { yes, no}
300 @attribute class { good, bad}
301 @data
302 '<0',6,'critical/other existing credit',radio/tv,1169,'no known savings','>=7',4,'male
single',none,4,'real estate',67,none,own,2,skilled,1,yes,yes,good
303 '0<=X<200',48,'existing paid',radio/tv,5951,'<100',1<=X<4',2,'female div/dep/mar',none,2,'real
estate',22,none,own,1,skilled,1,none,yes,bad
304 'no checking',12,'critical/other existing credit',education,2096,'<100',4<=X<7',2,'male
single',none,3,'real estate',49,none,own,1,'unskilled resident',2,none,yes,good
305 '<0',42,'existing paid',furniture/equipment,7882,'<100',4<=X<7',2,'male single',guarantor,4,'life
insurance',45,none,'for free',1,skilled,2,none,yes,good
306 '<0',24,'delayed previously', 'new car',4870,'<100',1<=X<4',3,'male single',none,4,'no known
property',53,none,'for free',2,skilled,2,none,yes,bad
307 'no checking',36,'existing paid',education,9055,'no known savings',1<=X<4',2,'male
single',none,4,'no known property',35,none,'for free',1,'unskilled resident',2,yes,yes,good
308 'no checking',24,'existing paid',furniture/equipment,2835,'500<=X<1000', '>=7',3,'male
single',none,4,'life insurance',53,none,own,1,skilled,1,none,yes,good
309 '0<=X<200',36,'existing paid', 'used car',6948,'<100',1<=X<4',2,'male
single',none,2,'car',35,none,rent,1,'high qualif/self emp/mgmt',1,yes,yes,good
310 'no checking',12,'existing paid',radio/tv,3059,'>=1000',4<=X<7',2,'male div/sep',none,4,'real
estate',61,none,own,1,'unskilled resident',1,none,yes,good
311 '0<=X<200',30,'critical/other existing credit', 'new car',5234,'<100',unemployed,4,'male
mar/wid',none,2,'car',28,none,own,2,'high qualif/self emp/mgmt',1,none,yes,bad
312 '0<=X<200',12,'existing paid', 'new car',1295,'<100', '<1',3,'female
div/dep/mar',none,1,'car',25,none,rent,1,skilled,1,none,yes,bad
```



Quais características são importantes para que o crédito no mercado seja bom ou ruim? - parâmetro "class

checking_status
duration
credit_history

8.2.

vote.arff (Como podemos caracterizar demócratas e republicanos?)

```
bank.arff x credit-g.arff x vote.arff x
195 @relation vote
196 @attribute 'handicapped-infants' { 'n', 'y' }
197 @attribute 'water-project-cost-sharing' { 'n', 'y' }
198 @attribute 'adoption-of-the-budget-resolution' { 'n', 'y' }
199 @attribute 'physician-fee-freeze' { 'n', 'y' }
200 @attribute 'el-salvador-aid' { 'n', 'y' }
201 @attribute 'religious-groups-in-schools' { 'n', 'y' }
202 @attribute 'anti-satellite-test-ban' { 'n', 'y' }
203 @attribute 'aid-to-nicaraguan-contras' { 'n', 'y' }
204 @attribute 'mx-missile' { 'n', 'y' }
205 @attribute 'immigration' { 'n', 'y' }
206 @attribute 'synfuels-corporation-cutback' { 'n', 'y' }
207 @attribute 'education-spending' { 'n', 'y' }
208 @attribute 'superfund-right-to-sue' { 'n', 'y' }
209 @attribute 'crime' { 'n', 'y' }
210 @attribute 'duty-free-exports' { 'n', 'y' }
211 @attribute 'export-administration-act-south-africa' { 'n', 'y' }
212 @attribute 'Class' { 'democrat', 'republican' }
213 @data
214 'n', 'y', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'y', '?', 'y', 'y', 'y', 'n', 'y', 'republican'
215 'n', 'y', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'y', 'y', 'y', 'n', '?', 'republican'
216 '?', 'y', 'y', '?', 'y', 'n', 'n', 'n', 'n', 'y', 'n', 'y', 'y', 'n', 'n', 'democrat'
217 'n', 'y', 'y', 'n', '?', 'y', 'n', 'n', 'n', 'n', 'y', 'n', 'y', 'n', 'n', 'y', 'democrat'
218 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'y', '?', 'y', 'y', 'y', 'y', 'democrat'
219 'n', 'y', 'y', 'n', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'y', 'y', 'y', 'y', 'democrat'
220 'n', 'y', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'n', '?', 'y', 'y', 'y', 'democrat'
221 'n', 'y', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'y', 'y', 'y', '?', 'y', 'republican'
222 'n', 'y', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'y', 'y', 'n', 'y', 'republican'
223 'y', 'y', 'y', 'n', 'n', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'democrat'
224 'n', 'y', 'n', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'republican'
225 'n', 'y', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'y', '?', 'y', 'y', '?', 'republican'
226 'n', 'y', 'y', 'n', 'n', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'y', 'n', 'n', 'n', 'n', 'democrat'
227 'y', 'y', 'y', 'n', 'n', 'y', 'y', 'y', 'y', '?', 'y', 'y', '?', 'n', 'n', 'y', '?', 'democrat'
228 'n', 'y', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'y', '?', 'n', '?', 'republican'
229 'n', 'y', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'y', 'n', 'y', 'y', 'y', 'n', '?', 'republican'
230 'y', 'n', 'y', 'n', 'n', 'y', 'y', 'y', '?', 'y', 'y', 'y', '?', 'n', 'n', 'y', 'democrat'
231 'y', '?', 'y', 'n', 'n', 'n', 'y', 'y', 'y', 'y', 'n', 'n', 'n', 'y', 'n', 'y', 'y', 'democrat'
232 'n', 'y', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'y', 'n', 'n', 'republican'
233 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'y', 'y', 'y', 'n', 'y', 'n', 'n', 'y', 'y', 'democrat'
234 'y', 'y', 'y', 'n', 'n', 'n', '?', 'y', 'y', 'n', 'n', 'y', 'n', 'n', 'n', 'y', 'y', 'democrat'
235 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'democrat'
236 'y', '?', 'y', 'n', 'n', 'n', 'y', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'y', 'y', 'democrat'
237 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'y', 'democrat'
238 'y', 'n', 'y', 'n', 'n', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'democrat'
239 'y', 'n', 'y', 'n', 'n', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'n', 'n', 'n', 'y', 'democrat'
240 'y', 'n', 'y', 'n', 'n', 'n', 'y', 'y', 'y', 'n', 'y', 'n', 'n', 'n', 'n', 'y', 'y', 'democrat'
241 'y', 'y', 'y', 'n', 'n', 'n', 'y', 'y', 'y', 'y', 'n', 'y', 'n', 'n', 'y', 'y', 'democrat'
242 'y', 'n', 'n', 'y', 'y', 'n', 'y', 'y', 'y', 'n', 'n', 'y', 'y', 'y', 'n', 'y', 'republican'
243 'y', 'y', 'y', 'n', 'n', 'n', 'y', 'y', 'y', 'y', 'n', 'n', 'n', 'n', 'y', 'y', 'democrat'
```

The screenshot shows the Weka Explorer interface. The 'Attribute Evaluator' tab is active, displaying the results of a search for the best subset of attributes. The search method used is 'BestFirst -D 1 -N 5'. The results show that 17 attributes were selected from the original set of 17 attributes. The selected attributes are: adoption-of-the-budget-resolution, physician-fee-freeze, immigration, and synfuels-corporation-cutback. The evaluation mode is set to 'evaluate on all training data'. The search method details include: Best First, Start set: no attributes, Search direction: forward, Stale search after 5 node expansions, Total number of subsets evaluated: 85, Merit of best subset found: 0.729. The Attribute Subset Evaluator (supervised, Class (nominal): 17 Class): CFS Subset Evaluator, Including locally predictive attributes. The selected attributes are: 3,4,10,11 : 4, adoption-of-the-budget-resolution, physician-fee-freeze, immigration, synfuels-corporation-cutback.

Como podemos caracterizar democratas e republicanos?

adoption-of-the-budget-resolution

physician-fee-freeze

immigration

synfuels-corporation-cutback

- resolução da adoção do orçamento
- congelamento taxa médico
- Imigração
- Corte no combustível sintético

9. Examine os datasets zoo.arff, weather.nominal.arff zoo.arff

```
bank.arff x credit-g.arff x vote.arff x zoo.arff x weather.nominal.arff x UNREGISTERED
/ 3
74 @RELATION zoo
75
76 @ATTRIBUTE animal {aardvark,antelope,bass,bear,boar,buffalo,calf,carp,catfish,cavy,cheetah,chicken,ch
ub,clam,crab,crayfish,crow,deer,dogfish,dolphin,dove,duck,elephant,flamingo,flea,frog,fruitbat,giraff
e,girl,gnat,goat,gorilla,gull,haddock,hamster,hare,hawk,herring,honeybee,housefly,kiwi,ladybird,lark,
leopard,lion,lobster,lynx,mink,mole,mongoose,moth,newt,octopus,opossum,oryx,ostrich,parakeet,penguin,
pheasant,pike,piranha,pitviper,platypus,polecat,pony,porpoise,puma,pussycat,raccoon,reindeer,rhea,scor
pion,seahorse,seal,sealion,seasnake,seawasp,skimmer,skua,slowworm,slug,sole,sparrow,squirrel,starfis
h,stingray,swan,termite,toad,tortoise,tuatara,tuna,vampire,vole,vulture,wallaby,wasp,wolf,worm,wren}
77 @ATTRIBUTE hair {false, true}
78 @ATTRIBUTE feathers {false, true}
79 @ATTRIBUTE eggs {false, true}
80 @ATTRIBUTE milk {false, true}
81 @ATTRIBUTE airborne {false, true}
82 @ATTRIBUTE aquatic {false, true}
83 @ATTRIBUTE predator {false, true}
84 @ATTRIBUTE toothed {false, true}
85 @ATTRIBUTE backbone {false, true}
86 @ATTRIBUTE breathes {false, true}
87 @ATTRIBUTE venomous {false, true}
88 @ATTRIBUTE fins {false, true}
89 @ATTRIBUTE legs INTEGER [0,9]
90 @ATTRIBUTE tail {false, true}
91 @ATTRIBUTE domestic {false, true}
92 @ATTRIBUTE catsize {false, true}
93 @ATTRIBUTE type {mammal, bird, reptile, fish, amphibian, insect, invertebrate }
94
95 @DATA
96 %
97 % Instances (101):
98 %
99 aardvark,true,false,true,false,false,true,true,true,true,false,false,4,false,false,true,mammal
100 antelope,true,false,false,true,false,false,false,true,true,true,false,false,4,true,false,true,mammal
101 bass,false,false,true,false,false,true,true,true,true,false,false,true,0,true,false,false,fish
102 bear,true,false,false,true,false,false,true,true,true,true,false,false,4,false,false,true,mammal
103 boar,true,false,false,true,false,false,true,true,true,true,false,false,4,true,false,true,mammal
104 buffalo,true,false,false,true,false,false,false,true,true,true,false,false,4,true,false,true,mammal
105 calf,true,false,false,true,false,false,false,true,true,true,false,false,4,true,true,true,mammal
106 carp,false,false,true,false,false,true,true,true,true,false,false,true,0,true,false,fish
107 catfish,false,false,true,false,false,true,true,true,true,false,false,true,0,true,false,fish
108 cavy,true,false,true,false,false,true,true,true,true,false,false,4,false,true,false,mammal
109 cheetah,true,false,false,true,false,false,true,true,true,true,false,false,4,true,false,true,mammal
110 chicken,false,true,true,false,true,false,false,true,true,true,false,false,2,true,true,false,bird
111 chub,false,false,true,false,true,true,true,true,true,false,false,true,0,true,false,fish
112 clam,false,false,true,false,false,false,true,false,false,false,false,0,false,false,false,invertebrate
113 crab,false,false,true,false,false,true,true,false,false,false,false,4,false,false,false,invertebrate
```

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose Discretize -B 10 -M -1.0 -R first-last Apply

Current relation: Relation: zoo Instances: 101 Attributes: 18 Sum of weights: 101

Selected attribute: Name: animal Missing: 0 (0%) Distinct: 100 Type: Nominal Unique: 99 (98%)

Attributes: All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> animal
2	<input type="checkbox"/> hair
3	<input type="checkbox"/> feathers
4	<input type="checkbox"/> eggs
5	<input type="checkbox"/> milk
6	<input type="checkbox"/> airborne
7	<input type="checkbox"/> aquatic
8	<input type="checkbox"/> predator
9	<input type="checkbox"/> toothed
10	<input type="checkbox"/> backbone
11	<input type="checkbox"/> breathes
12	<input type="checkbox"/> venomous
13	<input type="checkbox"/> fins
14	<input type="checkbox"/> legs
15	<input type="checkbox"/> tail
16	<input type="checkbox"/> domestic
17	<input type="checkbox"/> catsize
18	<input type="checkbox"/> type

No.	Label	Count	Weight
1	aardvark	1	1.0
2	antelope	1	1.0
3	bass	1	1.0
4	bear	1	1.0
5	boar	1	1.0
6	buffalo	1	1.0
7	calf	1	1.0
8	carp	1	1.0
9	catfish	1	1.0
10	cavy	1	1.0
11	cheetah	1	1.0
12	chicken	1	1.0
13	chub	1	1.0
14	clam	1	1.0
15	crab	1	1.0
16	crayfish	1	1.0
17	crow	1	1.0
18	deer	1	1.0
19	dogfish	1	1.0

Class: type (Nom) Visualize All

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Attribute Evaluator: Choose CfsSubsetEval -P 1 -E 1

Search Method: Choose BestFirst -D 1 -N 5

Attribute Selection Mode:

 Use full training set

 Cross-validation Folds 10 Seed 1

(Nom) type

Start Stop

Result list (right-click for options):

17:06:02 - BestFirst + CfsSubsetEval

19:19:58 - BestFirst + CfsSubsetEval

19:31:22 - BestFirst + CfsSubsetEval

19:50:23 - BestFirst + CfsSubsetEval

Attribute selection output:

```

animal
hair
feathers
eggs
milk
airborne
aquatic
predator
toothed
backbone
breathes
venomous
fins
legs
tail
domestic
catsize
type

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 149
Merit of best subset found: 0.864

Attribute Subset Evaluator (supervised, Class (nominal): 18 type):
CFS Subset Evaluator
Including locally predictive attributes

Selected attributes: 1,2,3,5,9,10,11,13,14,15 : 10
animal
hair
feathers
milk
toothed
backbone
breathes
fins
legs
tail
  
```


Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose Discretize -B 10 -M -1.0 -R first-last Apply

Current relation
 Relation: weather.symbolic
 Instances: 14
 Attributes: 5
 Sum of weights: 14

Attributes
 All | None | Invert | Pattern

No.	Name
<input checked="" type="checkbox"/>	outlook
<input type="checkbox"/>	temperature
<input type="checkbox"/>	humidity
<input type="checkbox"/>	windy
<input type="checkbox"/>	play

Selected attribute
 Name: outlook
 Missing: 0 (0%)
 Distinct: 3
 Type: Nominal
 Unique: 0 (0%)

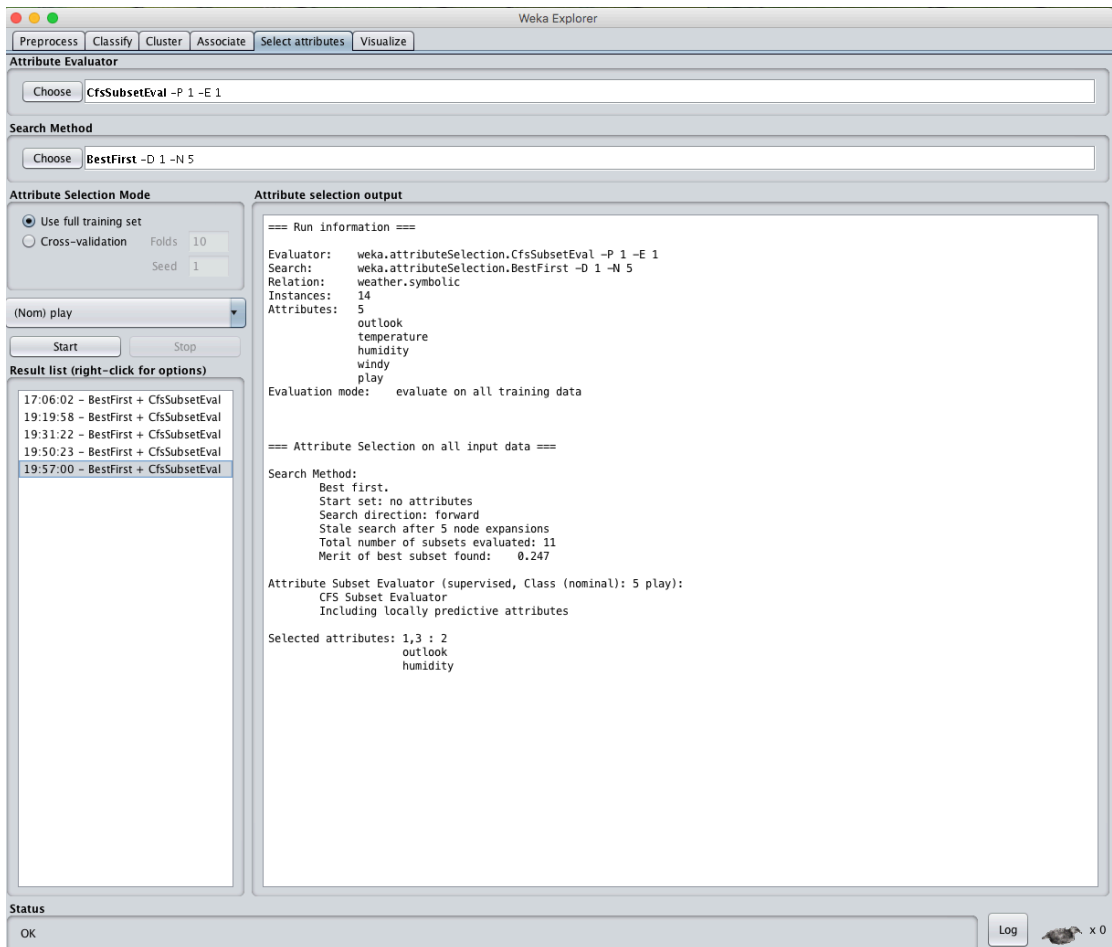
No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

Class: play (Nom) Visualize All

outlook	play (blue)	no play (red)
sunny	3	2
overcast	4	0
rainy	3	2

Remove

outlook
humidity



11. Referencias

1. wdz.eng.br/PropVisualizacaoFinal.docx – Trabalho apresentado na disciplina [Visualização de informação](#) . Foi utilizado para verificar quais eram os atributos para fazer visualização.
2. <https://pt.wikipedia.org/wiki/Weka> - o que é o Weka
3. https://sourceforge.net/projects/weka/files/weka-3-9/3.9.0/weka-3-9-0-oracle-jvm.dmg/download?use_mirror=freefr - download do weka
4. <http://weka.wikispaces.com/ARFF> - cria arquivo .arff

5. <https://prezi.com/rx9kqu3rm6ye/arvore-de-decisao-j48/> - algoritmo para arvore de decisão
6. <http://pt.slideshare.net/iaudesc/algoritmoid3ec45gilcimar> algoritmo id3 e c45.
7. https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwj6tK_Xn9PNAhXR15AKHfxyzB9AQFggeMAA&url=https%3A%2F%2Fcoruja.di.fc.ul.pt%2Fmod%2Fresource%2Fview.php%3Fid%3D15995&usg=AFQjCNG5fFx7nqiyec3AXzHIkOSS72u9Rg&sig2=65lMYgWN06qyqlBFkKI5NA - algoritmo simplKMeans
8. https://pt.wikipedia.org/wiki/Regras_de_associação
9. <http://www.ibm.com/developerworks/br/opensource/library/os-weka2/>
10. <https://mineracaodedados.wordpress.com/tag/distancia-euclidiana/>