

**Laboratório de Geoprocessamento da  
Escola de Arquitetura da UFMG**

**Dayan Magalhães de Castro**  
**Orientação: Profa Ana Clara M. Moura**

**APOSTILA DA AULA PRÁTICA DO SOFTWARE DE  
MINERAÇÃO DE DADOS WEKA**

**EXPORTAÇÃO DOS DADOS PELO ARCGIS  
E  
CRIANDO UMA IMAGEM BRUTA NO FORMATO PGM NO  
GIMP**

**Índice:**

1.	Gerando imagens no formato PGM e PNM no GIMP.....	2
1.1.	Imagens em escala de cinza – 256 tons .....	2
1.2.	Imagens em RGB – 256 tons .....	3
2.	Gerando os dados no ArcGis.....	4
2.1.	Extract Values to Point.....	4
2.1.1.	Vantagens .....	5
2.1.2.	Desvantagens .....	5
2.2.	Sample .....	5
2.2.1.	Vantagens .....	6
2.2.2.	Desvantagens .....	6
3.	Abrindo os dados no Weka .....	7
3.1.	Abrindo arquivo de banco de dados .....	7
3.2.	Selecionando o classificador a ser usado.....	8
3.3.	Executando o processo .....	11
3.4.	Visualizando os resultados .....	12

## 1. Gerando imagens no formato PGM e PNM no GIMP

### 1.1. Imagens em escala de cinza – 256 tons

**P2** – tipo de arquivo salvo

**#** – comentário salvo pelo Gimp

**2 2** – Altura e largura em pixels

**255** – Valor Máximo.

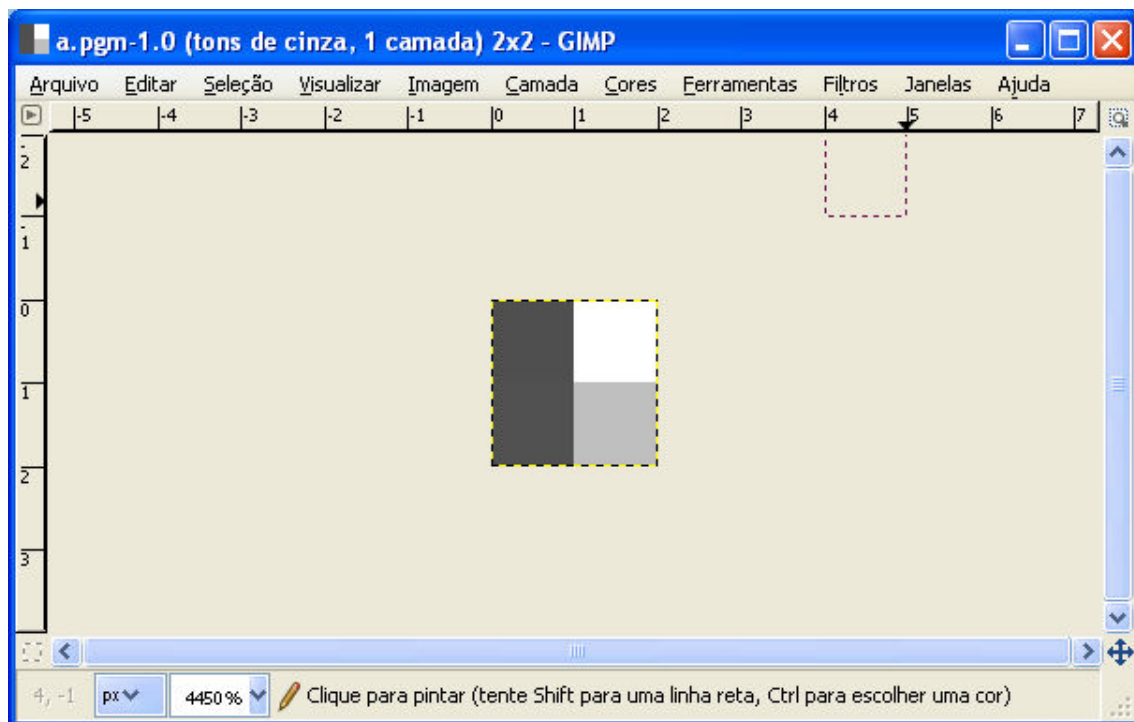
Matriz com os valores dos pixels



```
P2
# CREATOR: GIMP PNM Filter Version 1.1
2 2
255
79 255
81 191
```

Para obter ajuda, pressione F1

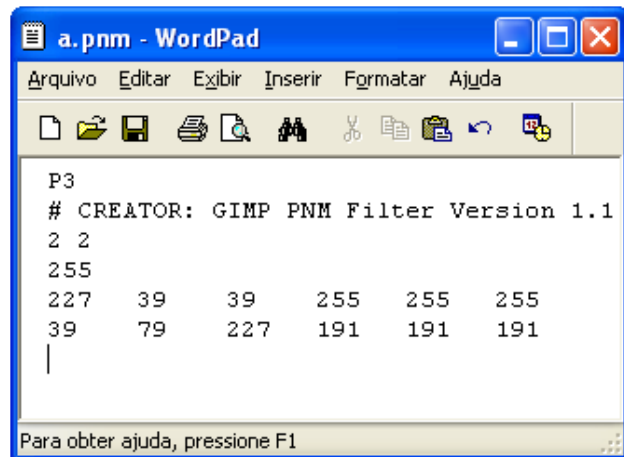
Imagem criada no GIMP



## 1.2. *Imagens em RGB – 256 tons*

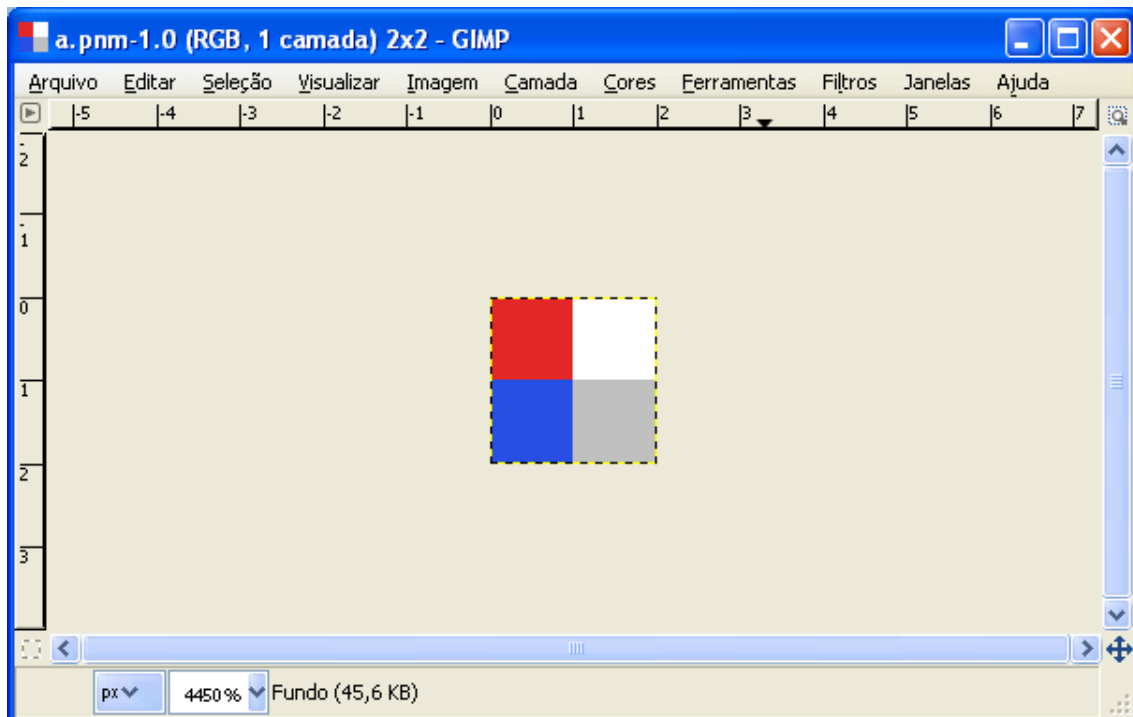
**P3** – tipo de arquivo salvo (RGB)  
**#** – comentário salvo pelo Gimp  
**2 2** – Altura e largura em pixels  
**255** – Valor Máximo.

Matriz com os valores dos pixels  
agora cada grupo de 3 valores  
formam 1 pixel  
Ex.: **227 R – 39 G – 39 B**



```
P3
# CREATOR: GIMP PNM Filter Version 1.1
2 2
255
227 39 39 255 255 255
39 79 227 191 191 191
|
```

## Imagem criada no GIMP

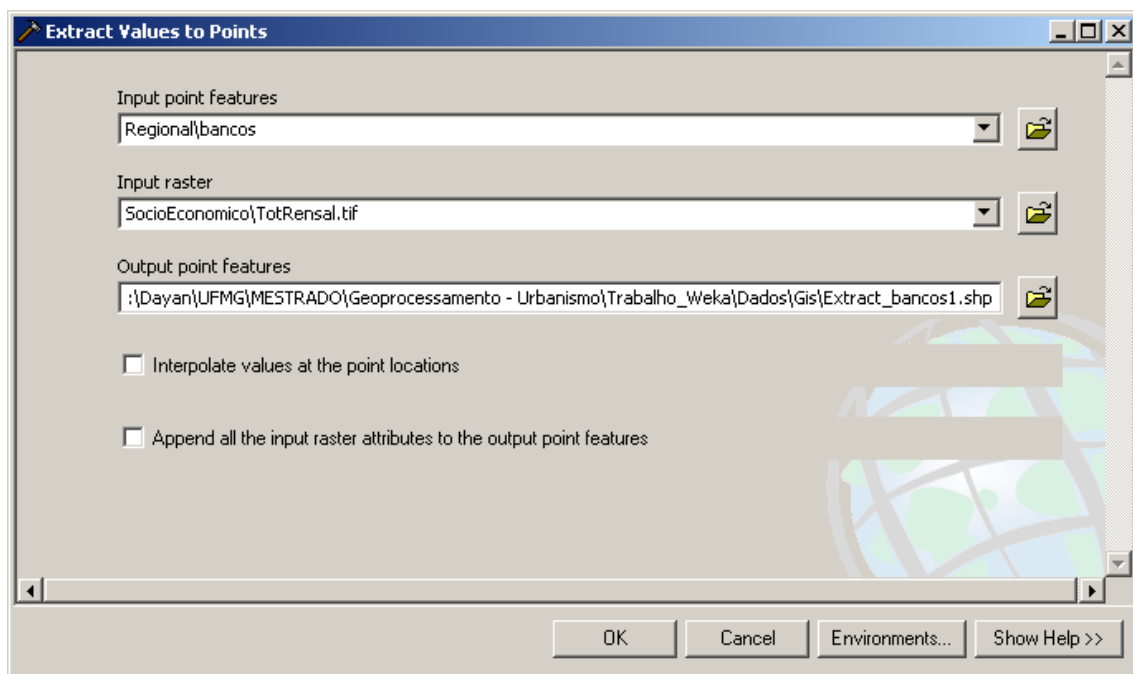


## 2. Gerando os dados no ArcGis

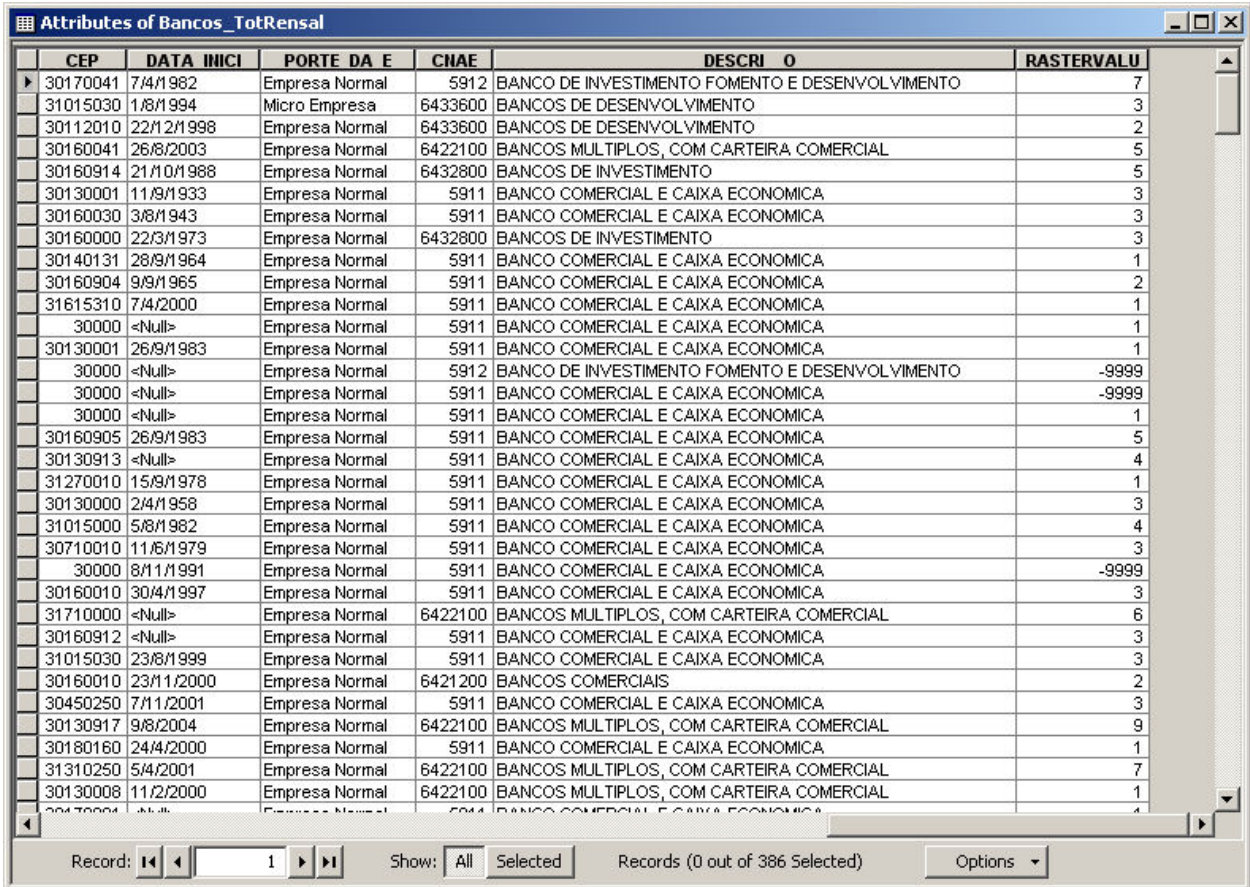
Existe duas maneiras de se executar a retirada dos valores dos pontos em cada imagem.

### 2.1. *Extract Values to Point*

É encontrada no Arctoolbox na parte Spatial Analyst Tools → Extraction → Extract Values to Points. Veja a imagem abaixo.



**Figura 1 - Tela de execução da extração**



CEP	DATA INICI	PORTE DA E	CNAE	DESCR O	RASTERVALU
30170041	7/4/1982	Empresa Normal	5912	BANCO DE INVESTIMENTO FOMENTO E DESENVOLVIMENTO	7
31015030	1/8/1994	Micro Empresa	6433600	BANCOS DE DESENVOLVIMENTO	3
30112010	22/12/1998	Empresa Normal	6433600	BANCOS DE DESENVOLVIMENTO	2
30160041	26/8/2003	Empresa Normal	6422100	BANCOS MULTIPLOS, COM CARTEIRA COMERCIAL	5
30160914	21/10/1988	Empresa Normal	6432800	BANCOS DE INVESTIMENTO	5
30130001	11/9/1933	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	3
30160030	3/8/1943	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	3
30160000	22/3/1973	Empresa Normal	6432800	BANCOS DE INVESTIMENTO	3
30140131	28/9/1964	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	1
30160904	9/9/1965	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	2
31615310	7/4/2000	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	1
30000	<Null>	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	1
30130001	26/9/1983	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	1
30000	<Null>	Empresa Normal	5912	BANCO DE INVESTIMENTO FOMENTO E DESENVOLVIMENTO	-9999
30000	<Null>	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	-9999
30000	<Null>	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	1
30160905	26/9/1983	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	5
30130913	<Null>	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	4
31270010	15/9/1978	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	1
30130000	2/4/1958	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	3
31015000	5/8/1982	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	4
30710010	11/6/1979	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	3
30000	8/11/1991	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	-9999
30160010	30/4/1997	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	3
31710000	<Null>	Empresa Normal	6422100	BANCOS MULTIPLOS, COM CARTEIRA COMERCIAL	6
30160912	<Null>	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	3
31015030	23/8/1999	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	3
30160010	23/11/2000	Empresa Normal	6421200	BANCOS COMERCIAIS	2
30450250	7/11/2001	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	3
30130917	9/8/2004	Empresa Normal	6422100	BANCOS MULTIPLOS, COM CARTEIRA COMERCIAL	9
30180160	24/4/2000	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	1
31310250	5/4/2001	Empresa Normal	6422100	BANCOS MULTIPLOS, COM CARTEIRA COMERCIAL	7
30130008	11/2/2000	Empresa Normal	6422100	BANCOS MULTIPLOS, COM CARTEIRA COMERCIAL	1
30170004	<Null>	Empresa Normal	5911	BANCO COMERCIAL E CAIXA ECONOMICA	4

**Figura 2 - Tabela de atributos do shape de saída da execução**

### 2.1.1. Vantagens

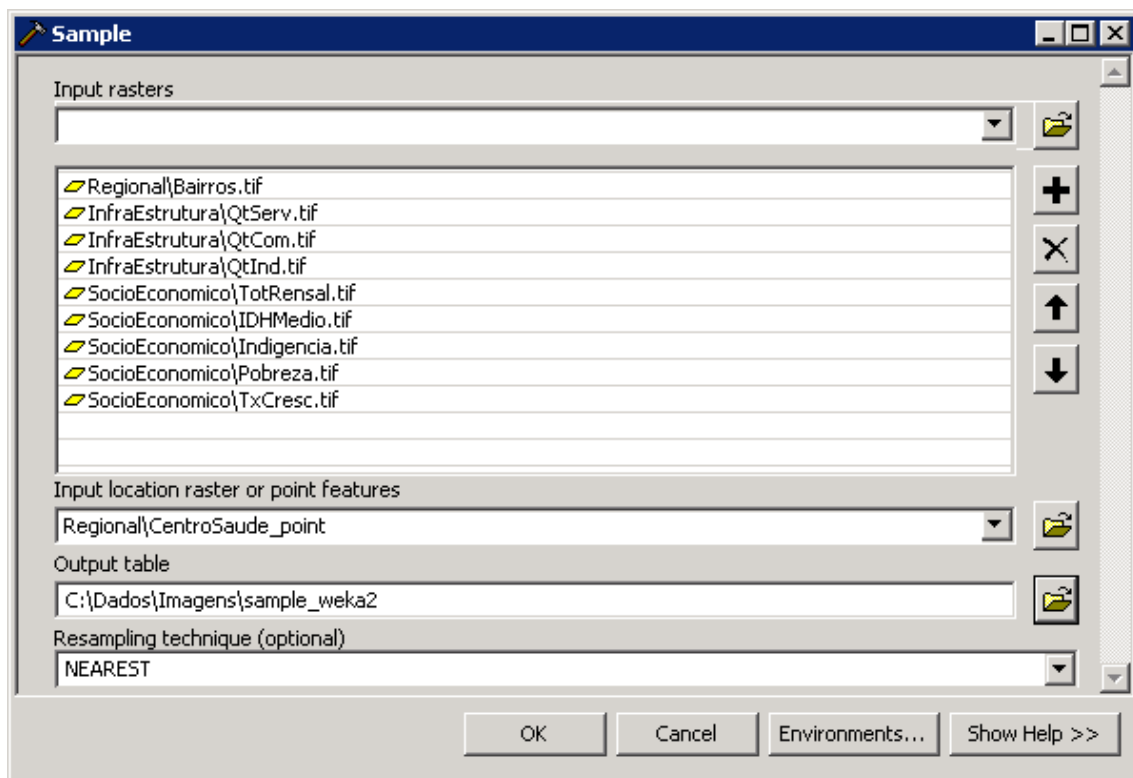
Como saída para esta ferramenta é criado um Shape com uma nova coluna onde encontramos os valores encontrados na imagem para os pontos dados, além de todas as informações presentes no shape de pontos informados inicialmente.

### 2.1.2. Desvantagens

É necessário executar esta tarefa para cada imagem.

## 2.2. Sample

É encontrada no Arctoolbox na parte Spatial Analyst Tools → Extraction → Sample. Veja a imagem abaixo.



**Figura 3 - Tela de execução do Sample**

Attributes of sample_weka_banco													
Rowid	MASK	X	Y	Z Z Z2C1	Z Z Z2C2	Z Z Z2C3	Z Z Z2C4	Z Z Z2C5	Z Z Z2C6	Z Z Z2C7	Z Z Z2C8	Z Z Z2C9	
1	0	609710,739253	7796167,192747	88	2554	771	81	6	93	100	79	-1	
2	1	613663,326401	7798237,327922	116	815	589	69	3	89	64	44	0	
3	2	611696,536646	7794983,9388	256	7924	4049	336	2	94	100	80	0	
4	3	611029,222625	7796779,796623	110	18913	9390	531	5	93	100	79	-1	
5	4	611029,222625	7796779,796623	110	18913	9390	531	5	93	100	79	-1	
6	5	611085,187468	7797241,529295	110	18913	9390	531	3	93	100	79	-1	
7	6	611254,34512	7797086,853694	110	18913	9390	531	3	93	100	79	-1	
8	7	611107,010705	7797073,212287	110	18913	9390	531	3	93	100	79	-1	
9	8	611406,898558	7795205,696022	256	7924	4049	336	9	94	100	80	0	
10	9	611405,700407	7797110,661008	110	18913	9390	531	2	93	100	79	-1	
11	10	605804,899662	7810006,751281	205	169	173	14	3	78	51	38	0	
12	11	610084,847325	7807389,10529	196	106	91	14	1	87	41	30	0	
13	12	610930,120117	7797508,191784	110	18913	9390	531	1	93	100	79	-1	

**Figura 4 - Resultado da execução do Sample**

### 2.2.1. Vantagens

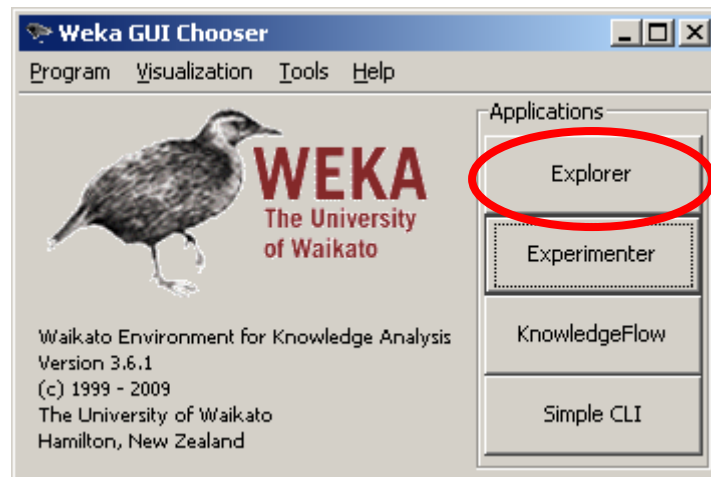
A análise é feita em mais de um raster por vez. Agilizando o processamento das informações.

### 2.2.2. Desvantagens

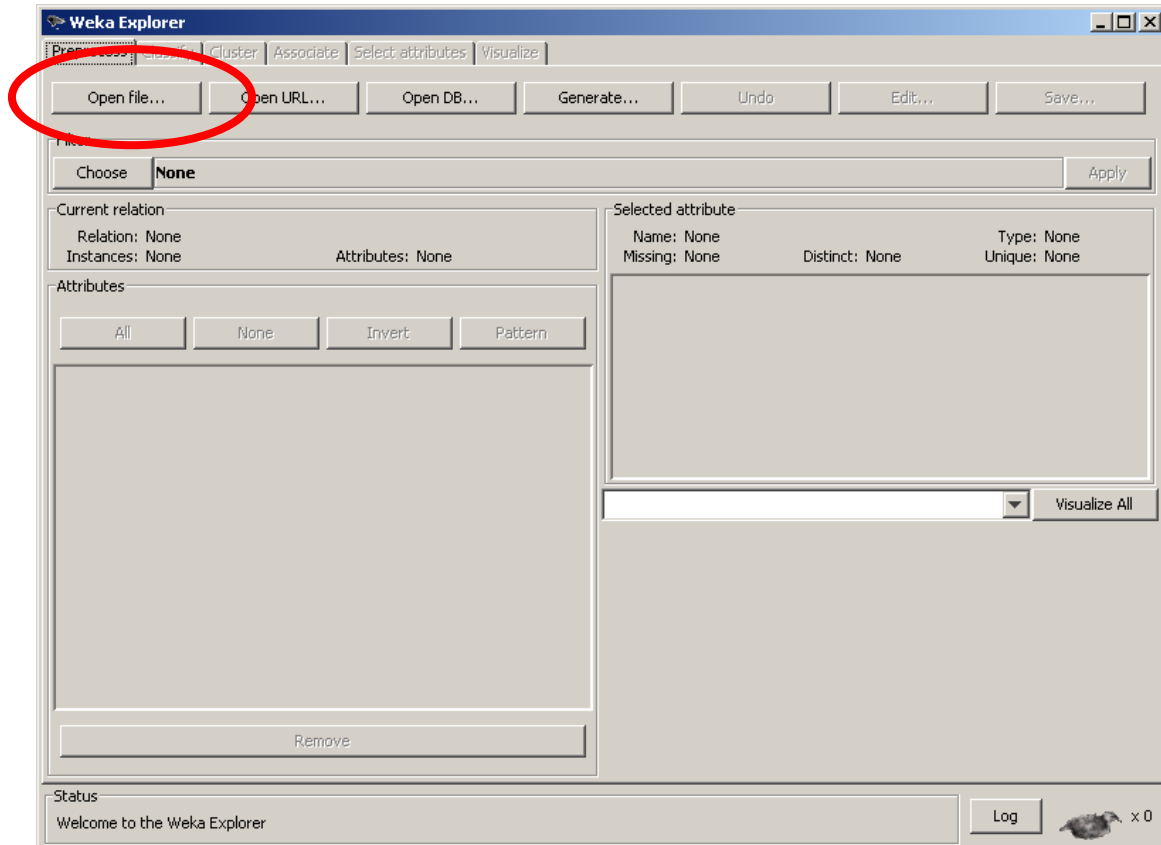
Perdem-se as informações contidas no shape de pontos, pois a saída é uma tabela contendo o X e o Y de cada ponto e os respectivos valores obtidos nas imagens.

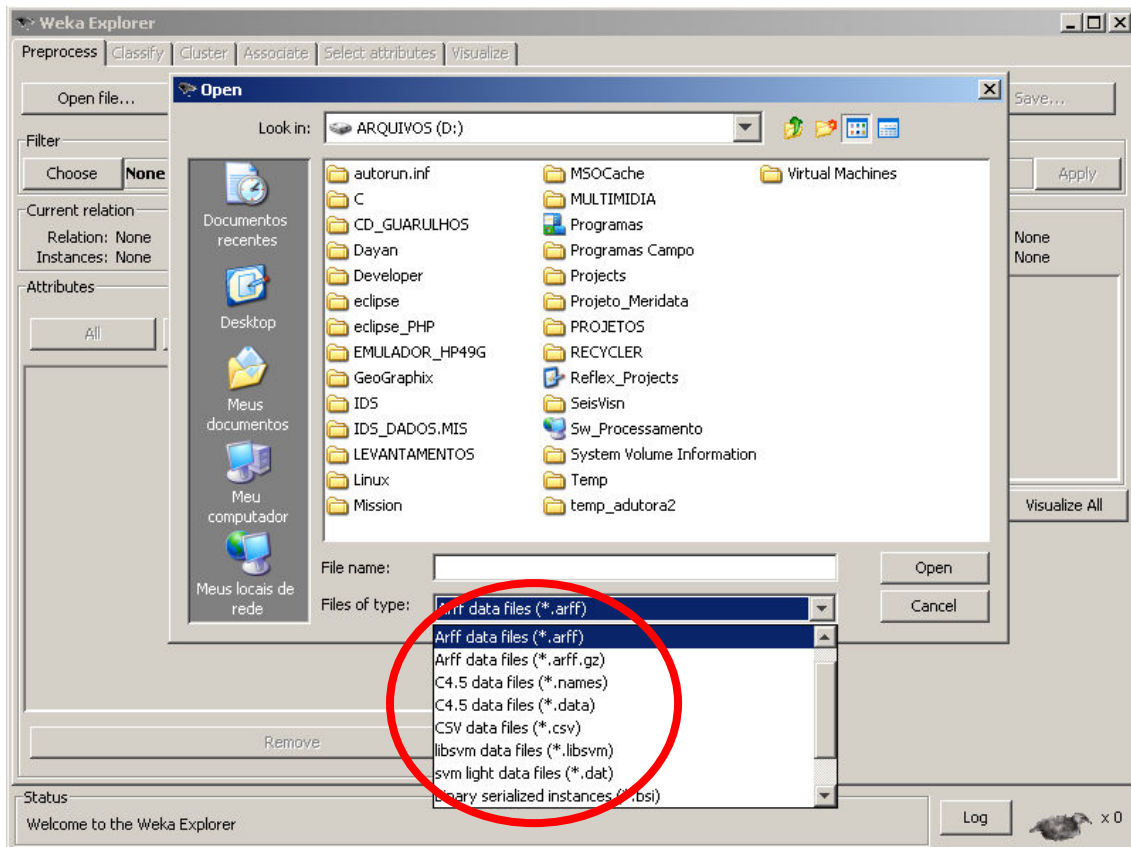
### 3. Abrindo os dados no Weka

Tela inicial do Weka:



#### 3.1. *Abrindo arquivo de banco de dados*





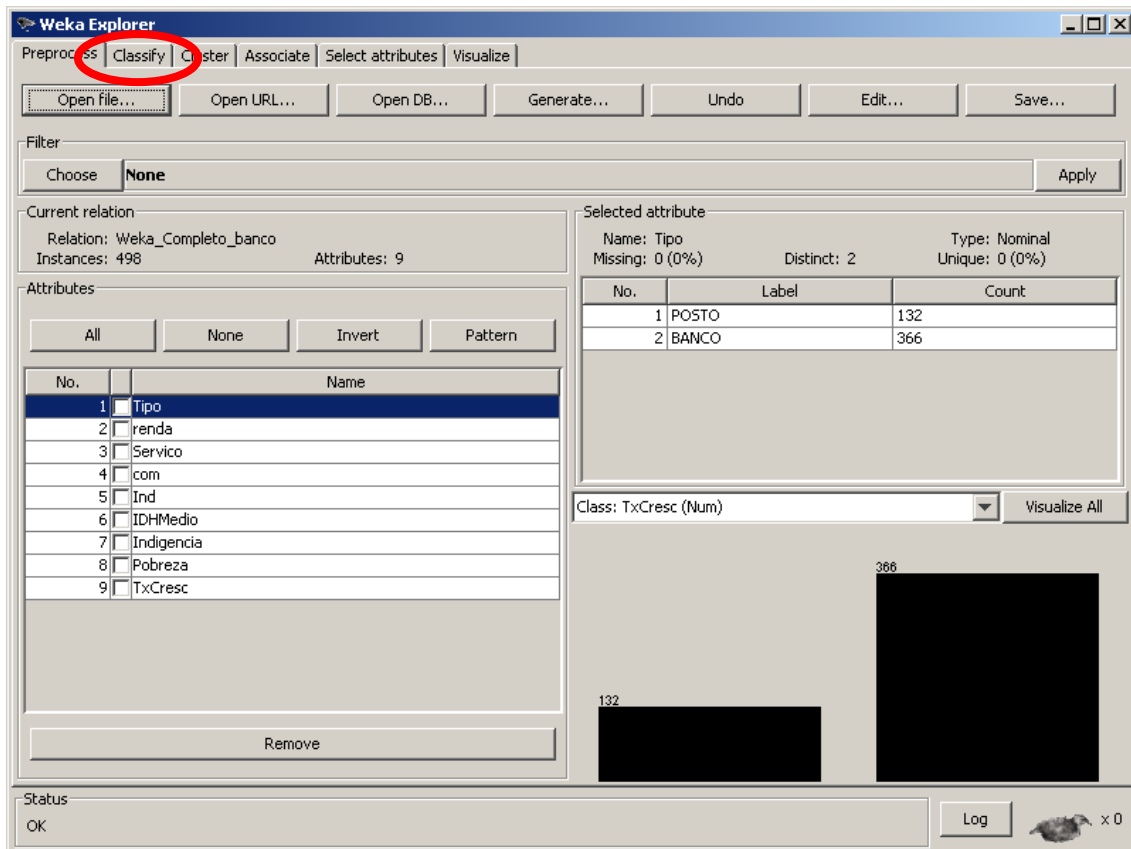
### 3.2. Selecionando o classificador a ser usado



# Prática sobre Heurística e Mineração de Dados

## Dayan Magalhães Castro, Orientação Ana Clara M. Moura

Página 9 de 13



**Weka Explorer**

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

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

Filter: Choose **None** Apply

Current relation:  
Relation: Weka\_Completo\_banco  
Instances: 498 Attributes: 9

Attributes:

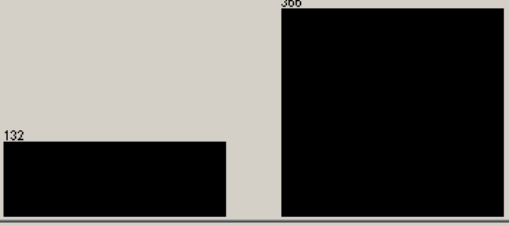
No.	Name
1	<input checked="" type="checkbox"/> Tipo
2	<input type="checkbox"/> renda
3	<input type="checkbox"/> Servico
4	<input type="checkbox"/> com
5	<input type="checkbox"/> Ind
6	<input type="checkbox"/> IDHMedio
7	<input type="checkbox"/> Indigencia
8	<input type="checkbox"/> Pobreza
9	<input type="checkbox"/> TxCresc

Remove

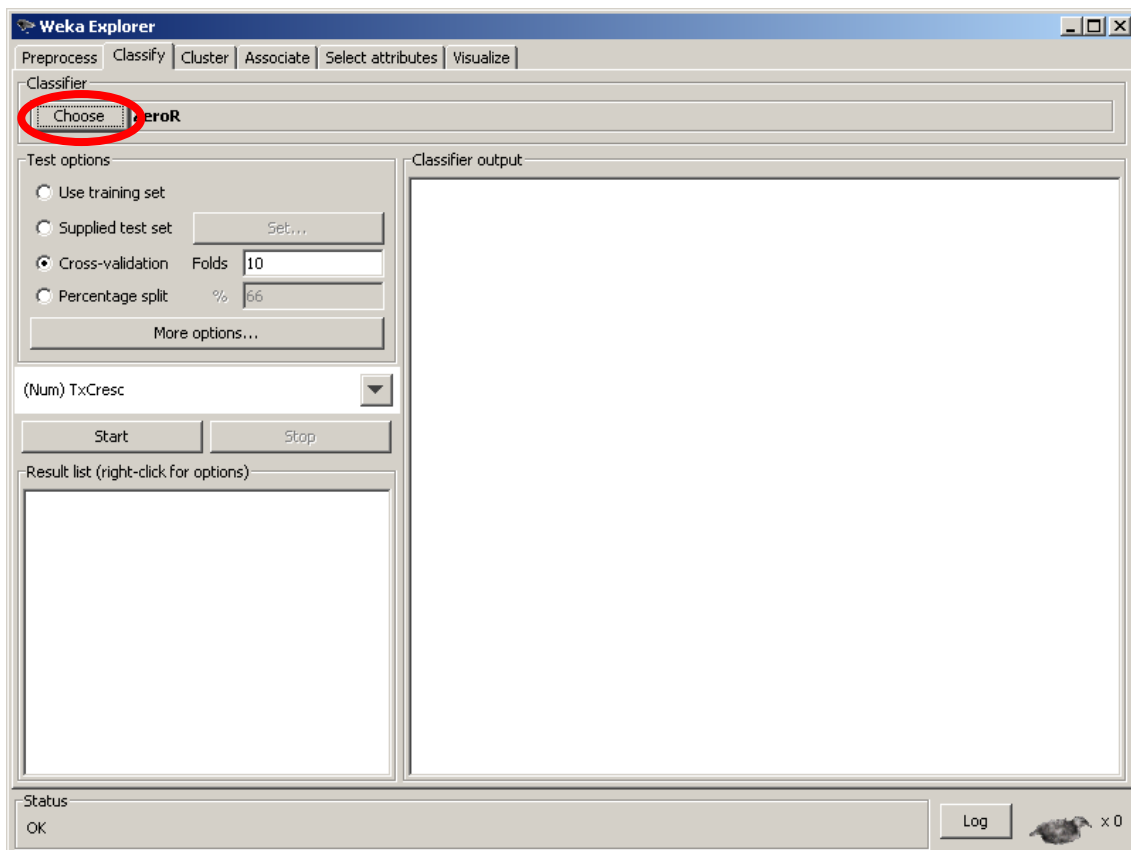
Selected attribute:  
Name: Tipo  
Missing: 0 (0%) Distinct: 2 Type: Nominal  
Unique: 0 (0%)

No.	Label	Count
1	POSTO	132
2	BANCO	366

Class: TxCresc (Num) Visualize All



Status: OK Log x 0



**Weka Explorer**

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

Classifier: Choose **ZeroR**

Test options:

- ☐ Use training set
- ☐ Supplied test set Set...
- ☒ Cross-validation Folds: 10
- ☐ Percentage split %: 66

More options...

(Num) TxCresc

Start Stop

Result list (right-click for options)

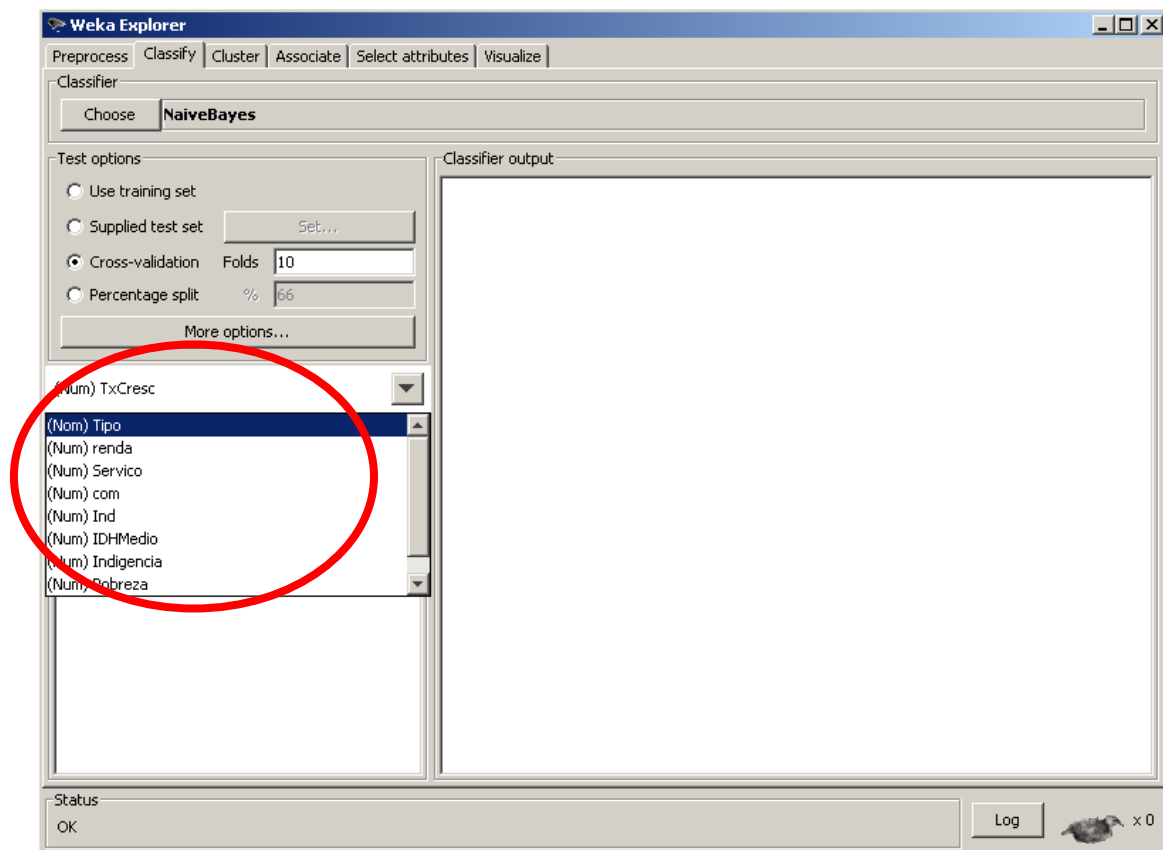
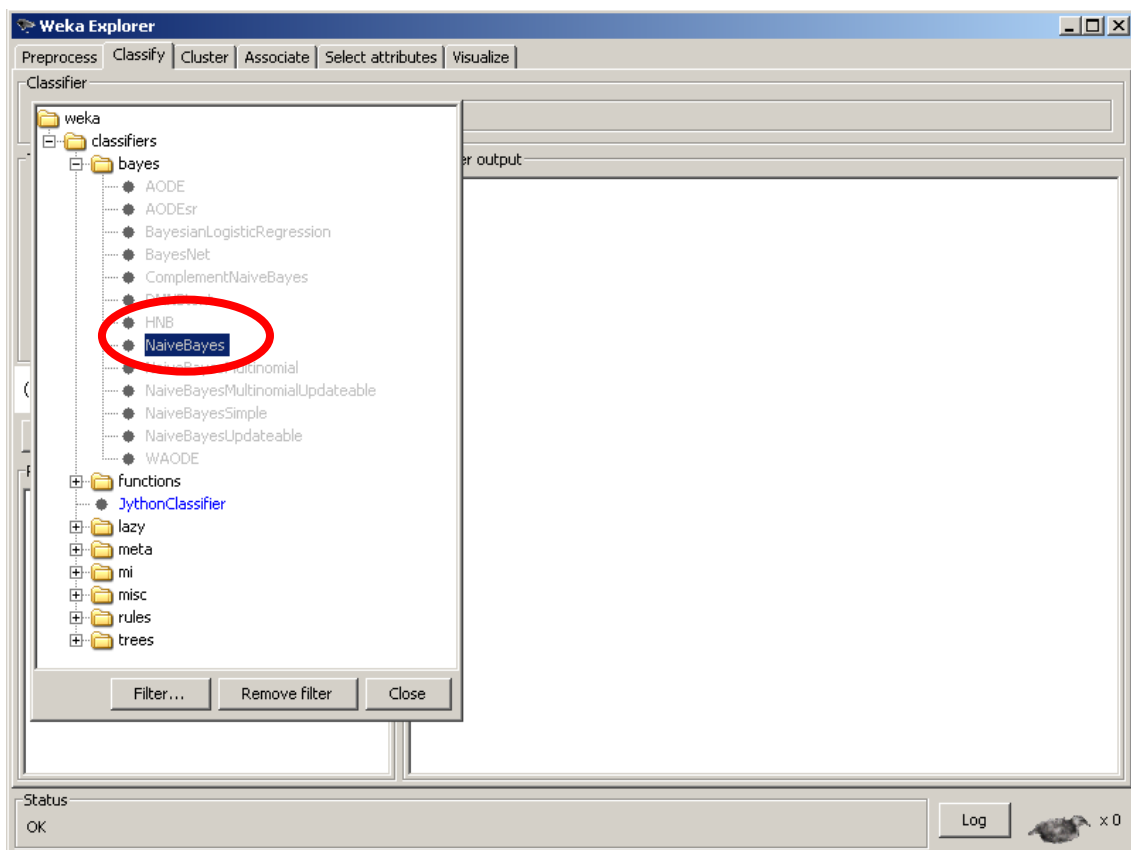
Classifier output

Status: OK Log x 0

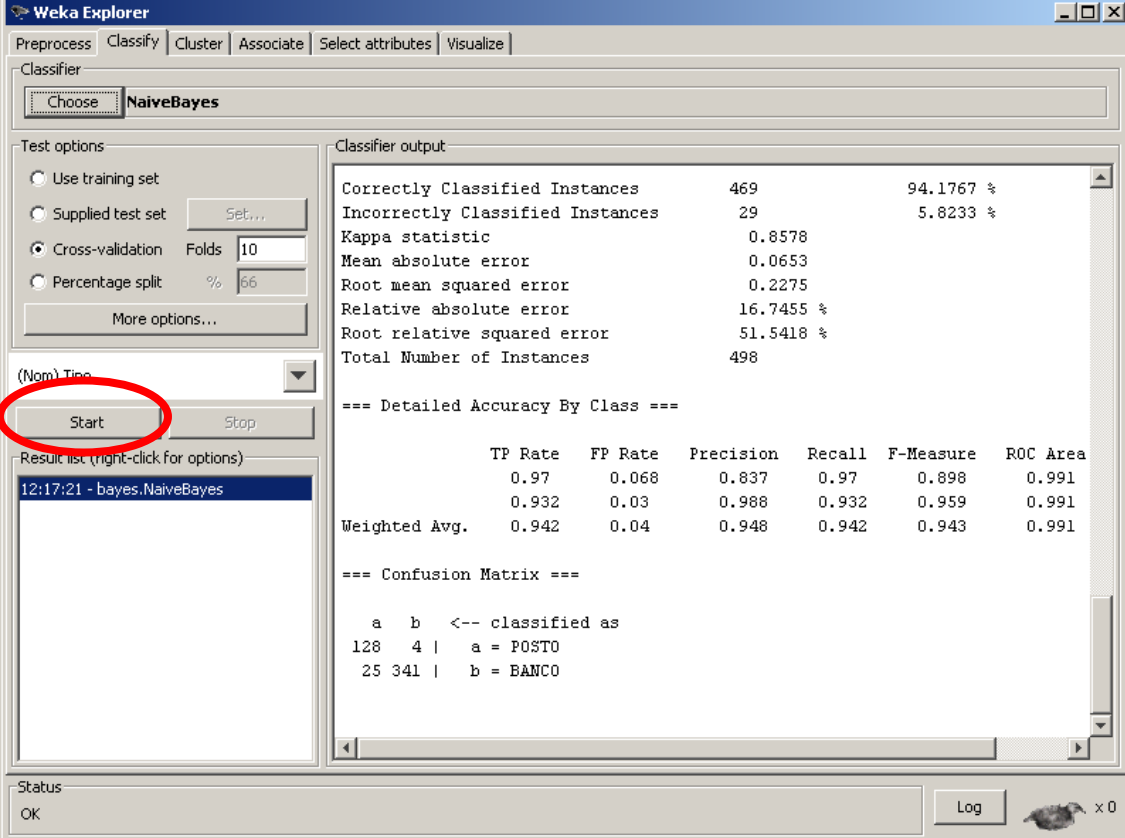
# Prática sobre Heurística e Mineração de Dados

## Dayan Magalhães Castro, Orientação Ana Clara M. Moura

Página 10 de 13



### 3.3. Executando o processo



The screenshot shows the Weka Explorer application window. The 'Classify' tab is selected. The 'Classifier' dropdown is set to 'NaiveBayes'. Under 'Test options', 'Cross-validation' is selected with 'Folds' set to 10. The 'Start' button is circled in red. The 'Classifier output' pane displays the following results:

Correctly Classified Instances 469 94.1767 %  
Incorrectly Classified Instances 29 5.8233 %  
Kappa statistic 0.8578  
Mean absolute error 0.0653  
Root mean squared error 0.2275  
Relative absolute error 16.7455 %  
Root relative squared error 51.5418 %  
Total Number of Instances 498

=== Detailed Accuracy By Class ===

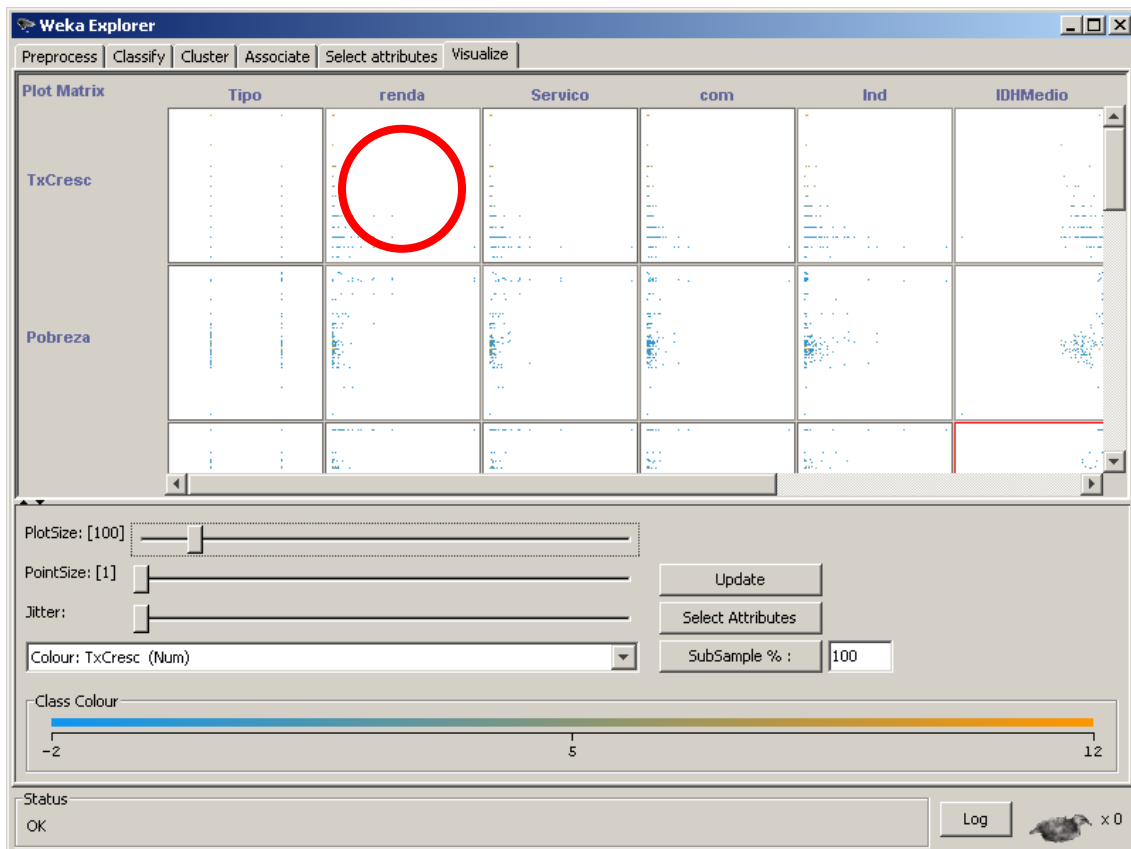
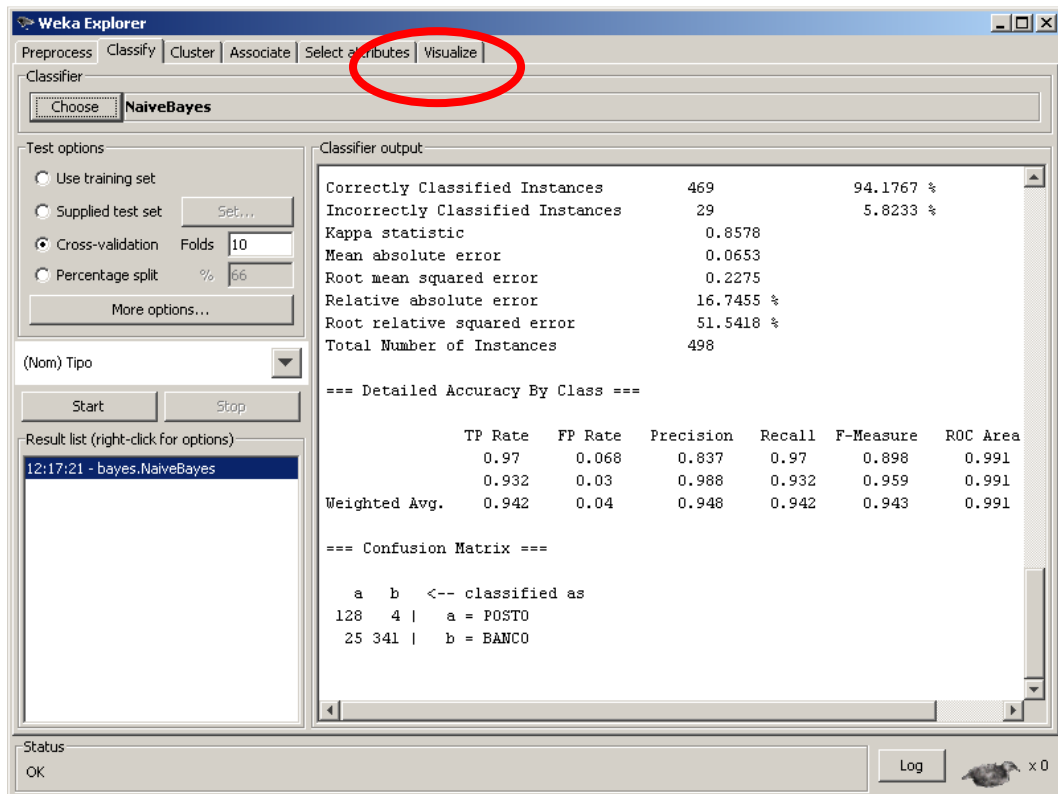
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
	0.97	0.068	0.837	0.97	0.898	0.991
	0.932	0.03	0.988	0.932	0.959	0.991
Weighted Avg.	0.942	0.04	0.948	0.942	0.943	0.991

=== Confusion Matrix ===

```
a b <-- classified as
128 4 | a = POSTO
25 341 | b = BANCO
```

The 'Result list' on the left shows '12:17:21 - bayes.NaiveBayes' selected. The 'Status' bar at the bottom indicates 'OK'.

### 3.4. Visualizando os resultados



# Prática sobre Heurística e Mineração de Dados

## Dayan Magalhães Castro, Orientação Ana Clara M. Moura

Página 13 de 13

