



PL03 - Introdução ao Weka

AEC - Mestrado em Engenharia Biomédica

<https://hpeixoto.me/class/aec>

Plano de Aula - PL03



Introdução ao Weka



Exemplo Prático



Ficha Exercícios (fe03)



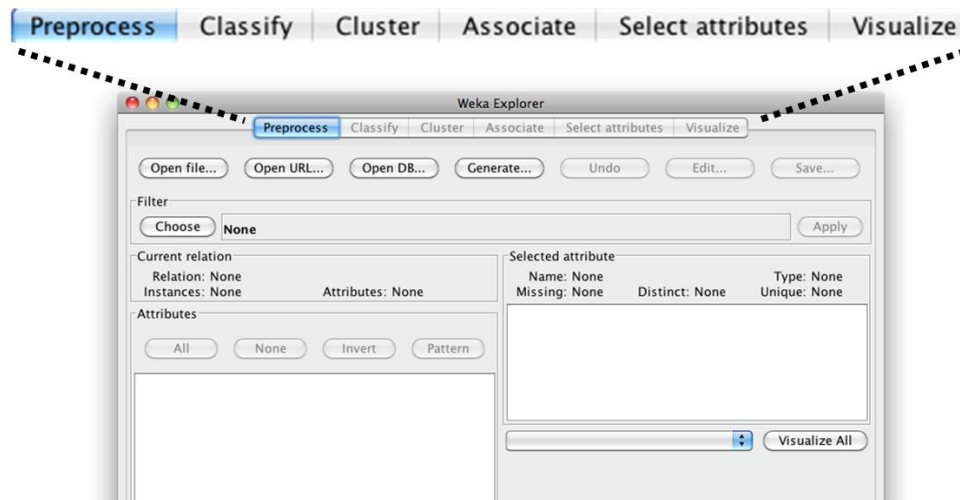
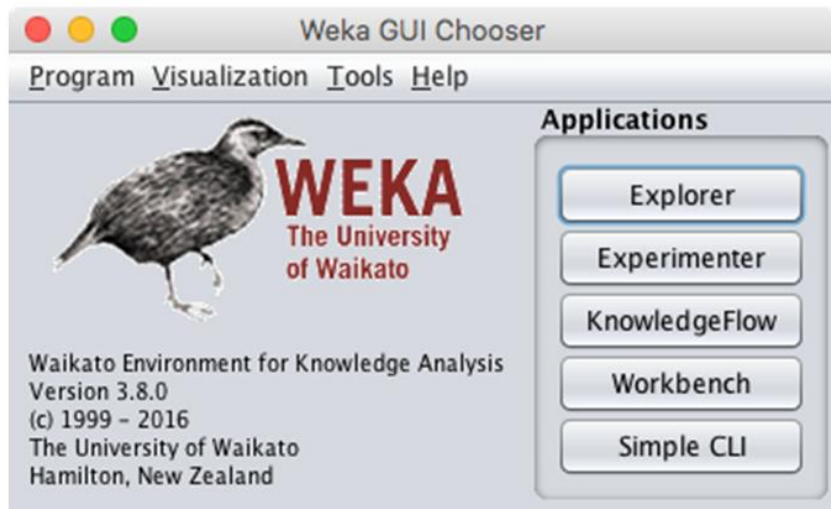
Introdução ao Weka

A ferramenta **Weka** possui uma coleção de algoritmos de machine learning para execução de tarefas de Data Mining. É um software que permite pré-processar grandes volumes de dados, aplicar diferentes algoritmos de Machine Learning e comparar vários outputs.

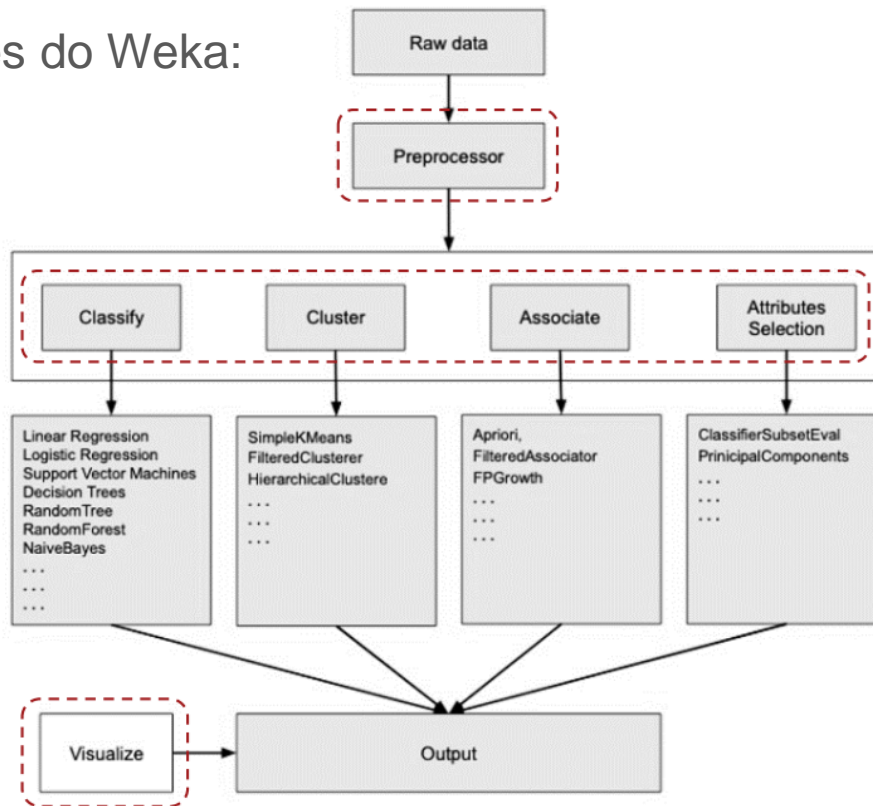
Waikato Environment for Knowledge Analysis



https://waikato.github.io/weka-wiki/downloading_weka/



Principais funcionalidades do Weka:





Exemplo Prático

Exemplo:

De forma a poder descobrir se os seus filhos podem ou não ir brincar para o parque, o António construiu um objetivo para o seu processo de DM. Os dados recolhidos são os seguintes:

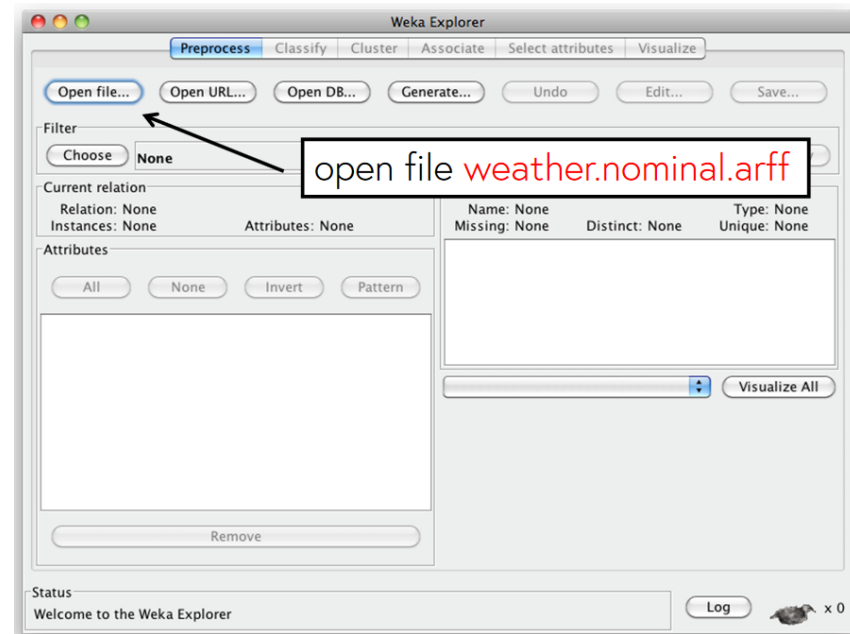
- Aspeto exterior (outlook)
- Temperatura (temp)
- Humidade (humidade)
- Vento (windy)
- Brincar (Play)



Exemplo:

Importar o ficheiro weather.nominal.arff

- Open file....
- Navegar até à pasta data do weka e seleccionar o ficheiro.



Exemplo Prático

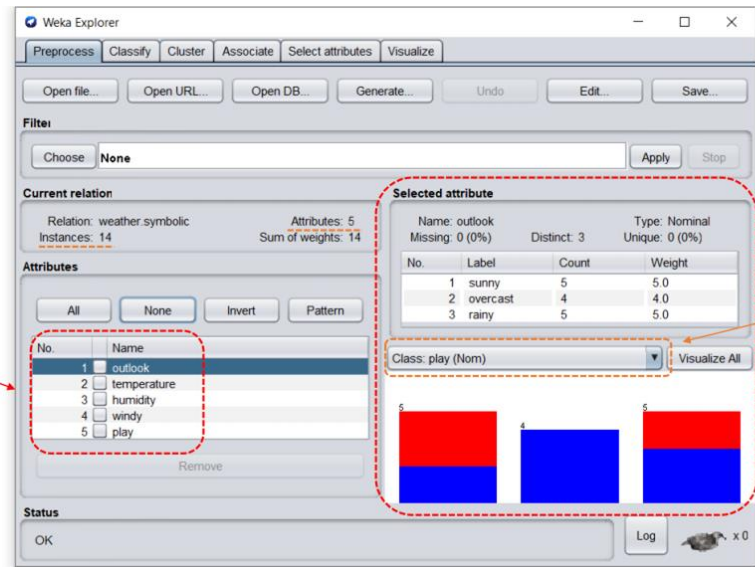
Exemplo:

- Analisar de forma detalhada a informação Presente na fase "Preprocess"
- Identificar os atributos
- Identificar a classe
- Validar os gráficos de distribuição por classe

attributes					
	Outlook	Temp	Humidity	Windy	Play
instances 1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
3	Overcast			False	Yes
4	Rainy			False	Yes
5	Rainy		Normal	False	Yes
6	Rainy		Normal	True	No
7	Overcast		Normal	True	Yes
8	Sunny	Cool	Normal	False	Yes
9	Sunny	Cool	Normal	True	Yes
10	Rainy	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Rainy	Mild	High	True	No

Classification problem:
predict the "class" value

Atributos



Weka Explorer

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

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

Filter: Choose: None | Apply | Stop

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

Attributes: All | None | Invert | Pattern

No.	Name
1	outlook
2	temperature
3	humidity
4	windy
5	play

Remove

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

Status: OK | Log | x0

Classe

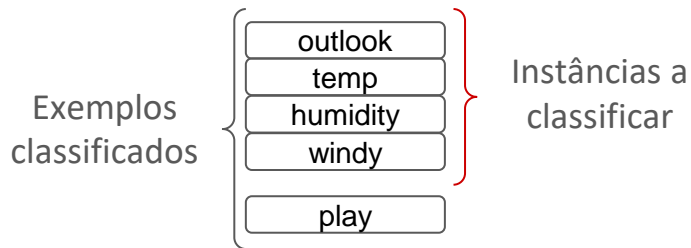
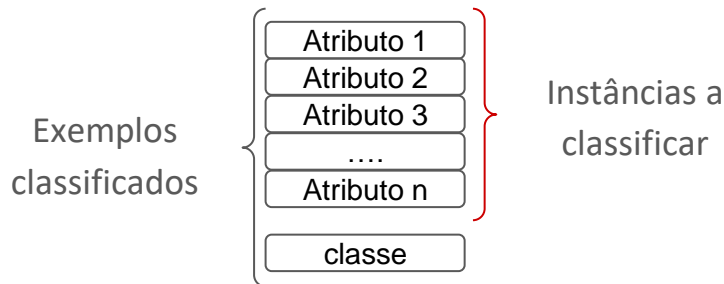
Exemplo Prático

Exemplo:

PROBLEMA DE CLASSIFICAÇÃO (supervised learning)

Dataset -> exemplos classificados

Criar modelos que classifiquem novos exemplos



Discreto -> nominal -> problema de classificação
 Contínuo -> numérico -> problema de regressão

Exemplo Prático

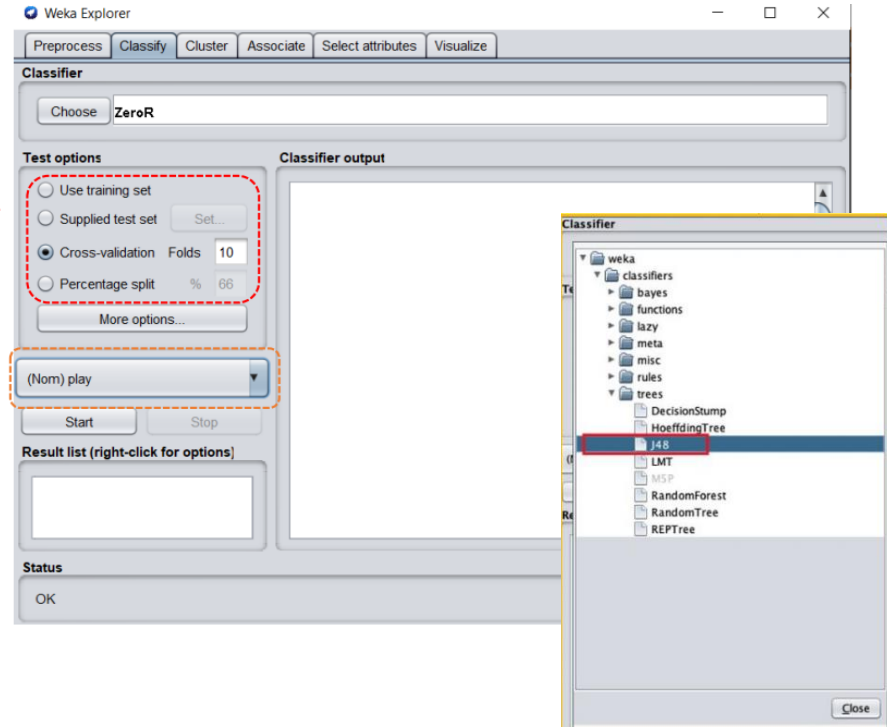
Exemplo:

Algoritmo de classificação

Selecionar: Classifiers -> Trees -> J48

OPÇÕES DE TESTE

CLASSE



Exemplo Prático

Exemplo:

Resultados

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

☐ Use training set

☐ Supplied test set **Set...**

☒ Cross-validation **Folds 10**

☐ Percentage split **% 66**

More options...

(Nom) play

Start **Stop**

Result list (right-click for options)

02:25:18 - trees.J48

Classifier output

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    weather-weka.filters.supervised.attribute.Discretize
Instances:   14
Attributes:  3
  outlook
  windy
  play
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----
outlook = sunny: no (5)
outlook = overcast: yes (2)
outlook = rainy
| windy = TRUE: no (2.0)
| windy = FALSE: yes (3.0)
  
```

=== Confusion Matrix ===

a \ b	classified as a = yes	classified as b = no
6 3	6	3
5 0	5	0

=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	6	42.8571 %
Incorrectly Classified Instances	8	57.1429 %
Kappa statistic	-0.3659	
Mean absolute error	0.4571	
Root mean squared error	0.5589	
Relative absolute error	95.9918 %	
Root relative squared error	113.2761 %	
Total Number of Instances	14	

Number of Leaves : 4

Size of the tree : 6

Exemplo Prático

Exemplo:

Matriz de confusão

	a	b	<= Classificado como
5		4	a
3		2	b

Exemplo Prático

Exemplo:

Abrir o dataset glass.arff;

Escolher o algoritmo J48;

Analisar os resultados e visualizar a árvore;

Carregar em cima do algoritmo J48;

Examinar as diferentes opções;

Usar uma árvore não podada - 'unpruned tree';

Colocar a propriedade 'minNumObj' igual a 15 para evitar folhas pequenas;

Comparar com os resultados obtidos anteriormente.

Exemplo Prático

Prunning Decision Trees:

é uma técnica que reduz o tamanho das árvores de decisão ao remover secções da árvore que fornecem pouco poder para classificar as instâncias.

A poda reduz a complexidade do classificador final e, portanto, melhora a precisão da previsão através da redução do excesso de ajustes – overfitting.



Ficha de Exercícios 03



PL03 - Introdução ao Weka

AEC - Mestrado em Engenharia Biomédica

<https://hpeixoto.me/class/aec>