

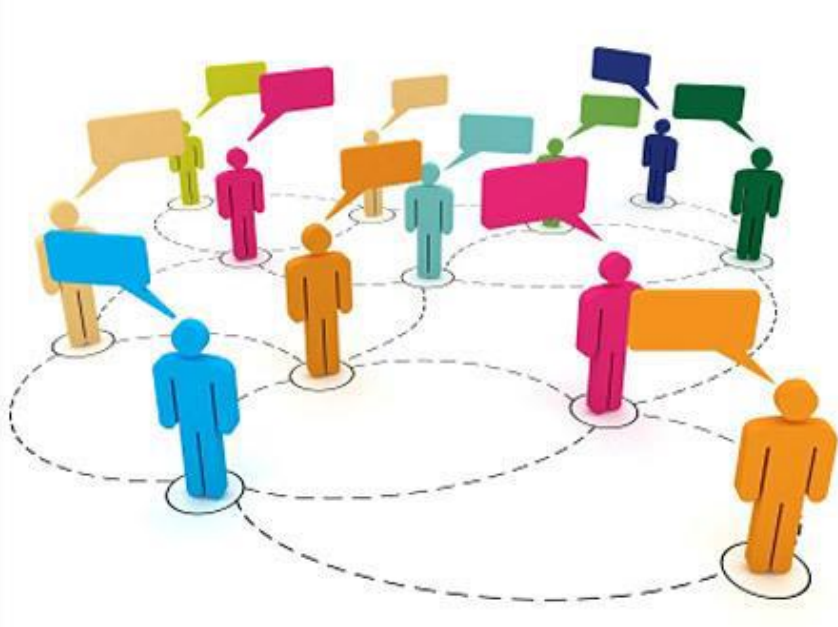


Universidade Federal do Sul e Sudeste do Pará
Faculdade de Computação e Engenharia Elétrica
Inteligência Artificial

**Aula 12 - Capítulo 5: Introdução à
Mineração de Dados e Aprendizagem de
Máquina utilizando o WEKA (Regressão e
Classificação)**

Prof. Dr. Elton Alves

Introdução à Aprendizagem de Máquina



NETFLIX



amazon.com

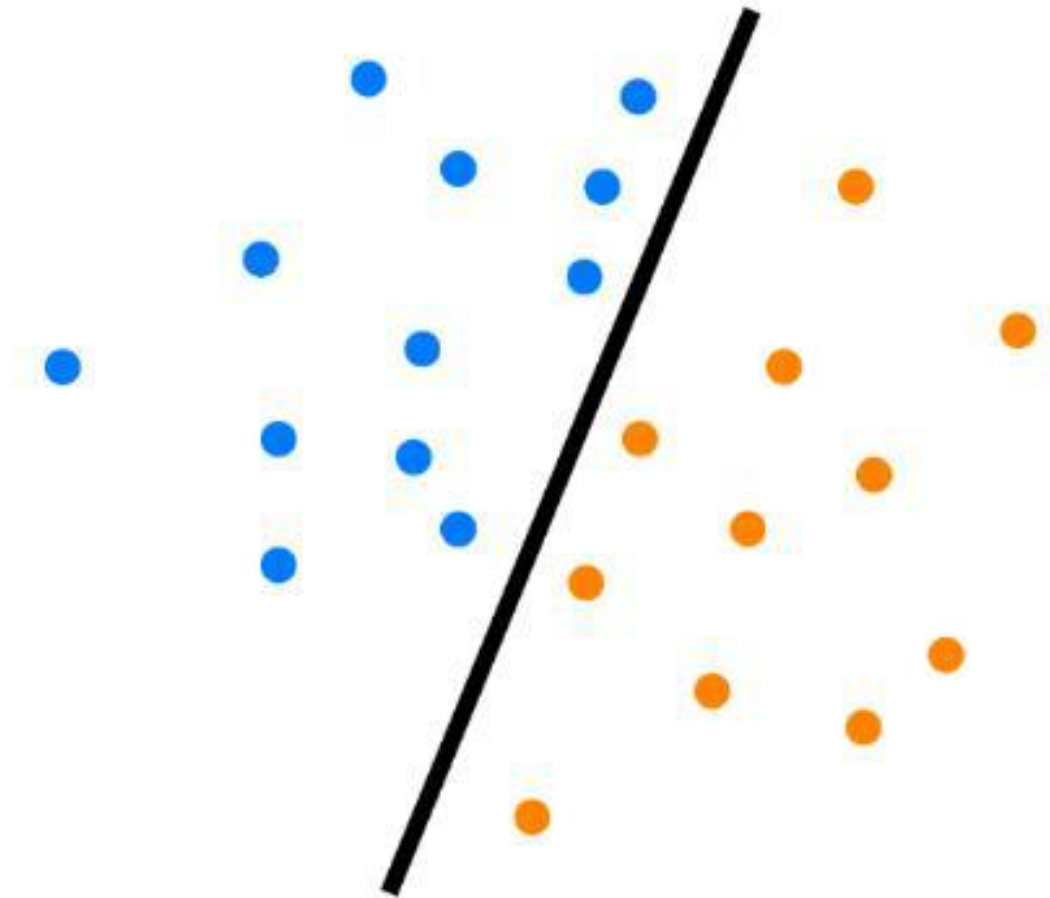
eHarmony

Aprendizagem de Máquina

Aprendizagem de Máquina	
Métodos Preditivos	Métodos Descritivos
Classificação	Associação
Regressão	Agrupamento
	Detecção de desvios
	Padrões sequenciais
	Sumarização

Classificação

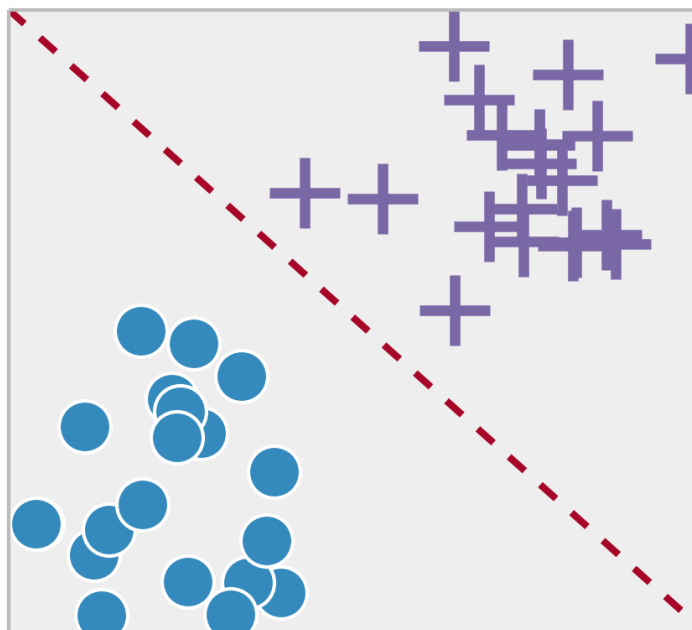
- ❑ Insatisfação de clientes;
- ❑ Risco de crédito;
- ❑ Separação de notícias;
- ❑ Reconhecimento de voz;
- ❑ Reconhecimento de face;
- ❑ Previsão de doenças.



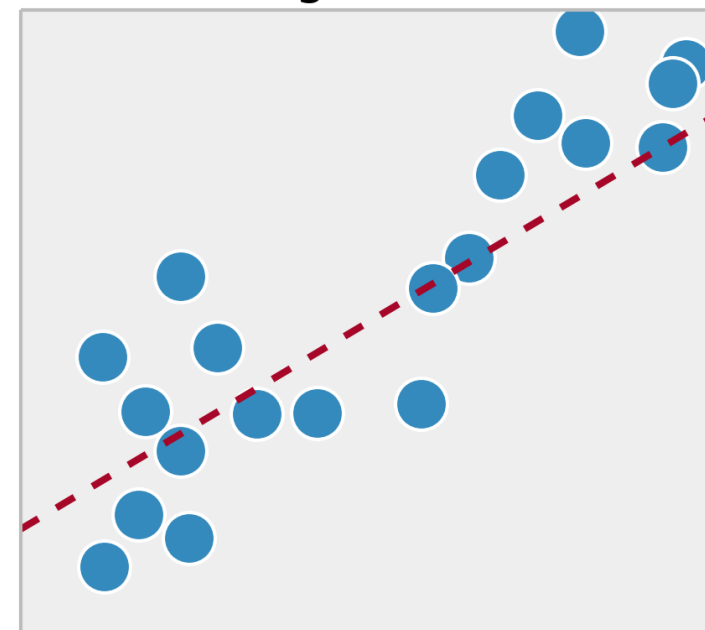
Regressão

- ❑ Gastos propaganda → **valor de venda**;
- ❑ Temperatura, umidade e pressão do ar → **velocidade do vento**;
- ❑ Resultados do exame → **probabilidade de um paciente sobreviver**;
- ❑ Fatores externos → **valor do dólar**

Classification



Regression

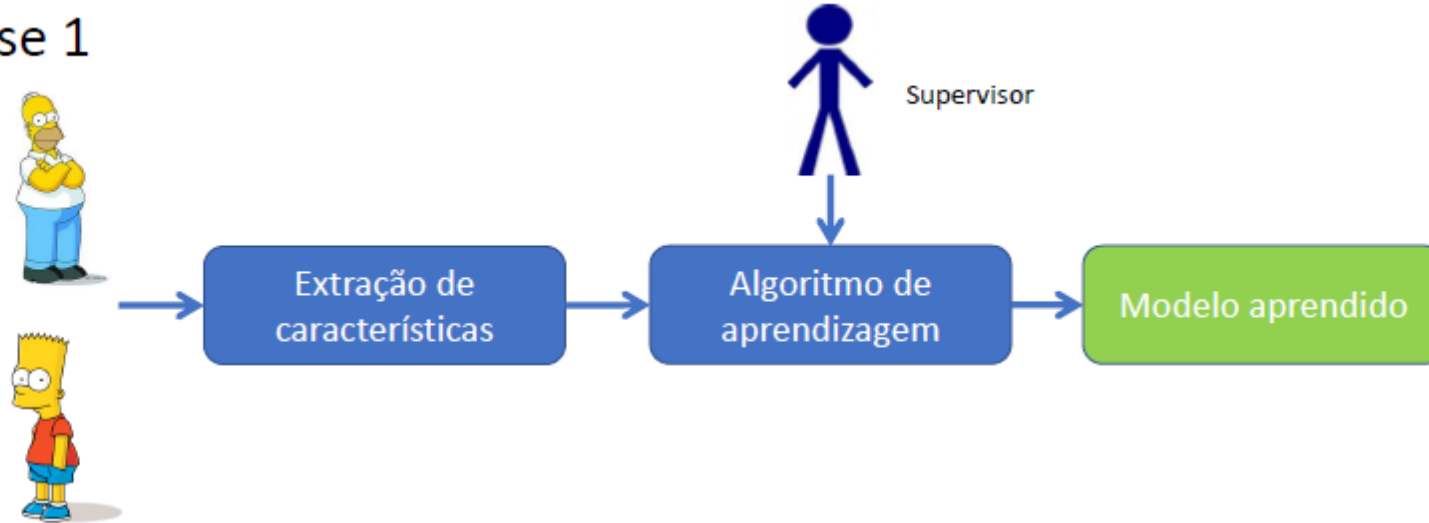


Tipos de Aprendizagem Máquina

Aprendizagem de Máquina		
Supervisionada	Não supervisionada	Reforço
Classificação	Associação	
Regressão	Agrupamento	
	Detecção de desvios	
	Padrões sequenciais	
	Sumarização	

Aprendizagem Supervisionada

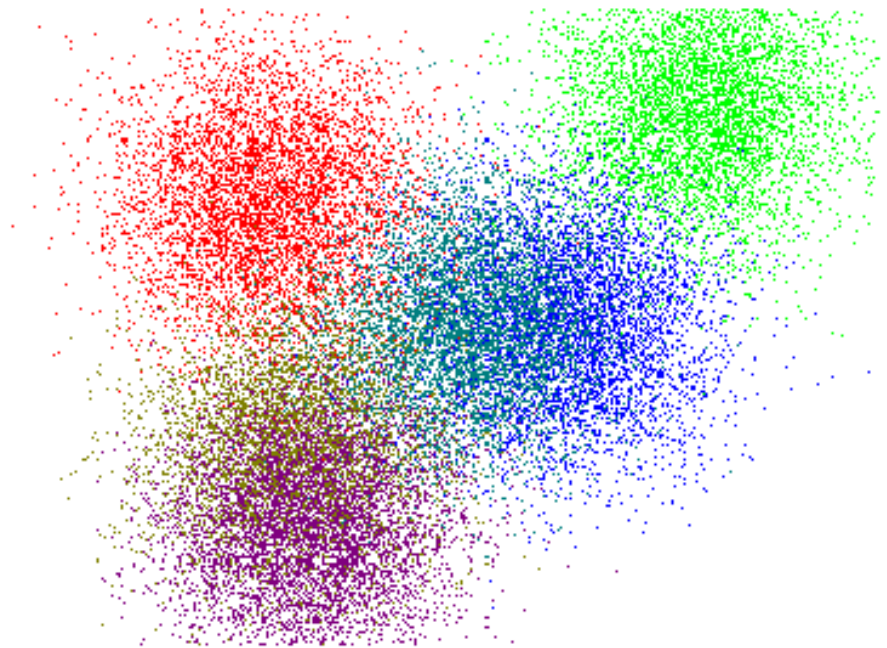
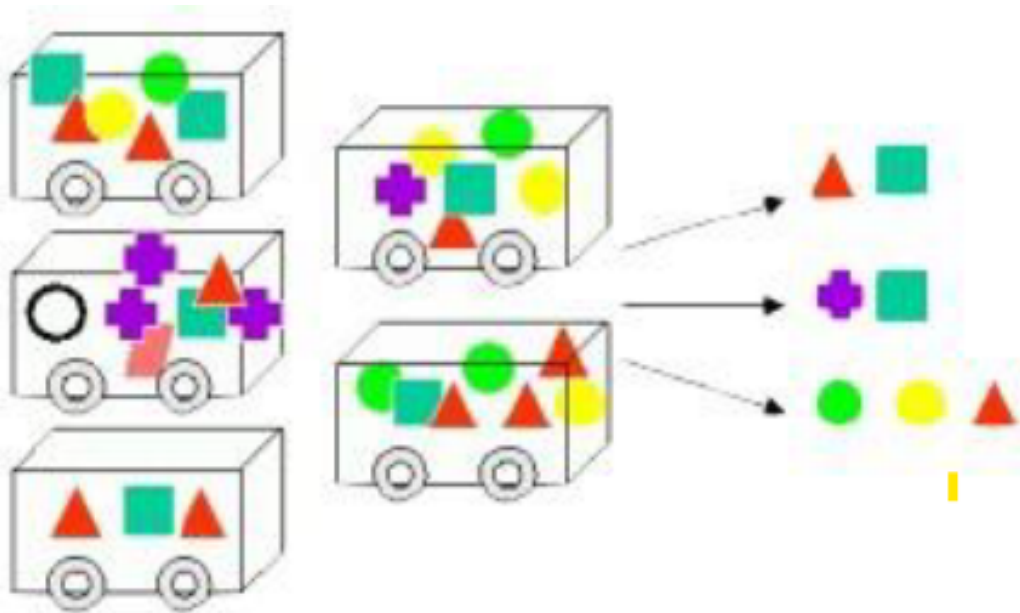
Fase 1



Fase 2



Aprendizagem Não-Supervisionada



Aprendizagem por Reforço

- ❑ Aprender com as interações com o ambiente (causa e efeito);
- ❑ Aprender com sua própria experiência;
- ❑ Exemplos:
 - Robô coletando lixo aprendendo a andar em um ambiente;
 - Controle automatizado de elevadores.

WEKA

cs.waikato.ac.nz/ml/weka/index.html



[Project](#)

[Software](#)

[Book](#)

[Courses](#)

[Publications](#)

[People](#)

[Related](#)

Weka 3: Machine Learning Software in Java

Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization.

Found only on the islands of New Zealand, the Weka is a flightless bird with an inquisitive nature. The name is pronounced like **this**, and the bird sounds like **this**.

Weka is open source software issued under the **GNU General Public License**.

We have put together several **free online courses** that teach machine learning and data mining using Weka. The videos for the courses are available **on Youtube**.

Weka supports **deep learning**!



Getting started

- **Requirements**
- **Download**
- **Documentation**
- **FAQ**
- **Getting Help**

Further information

- **Citing Weka**
- **Datasets**
- **Related Projects**
- **Miscellaneous Code**
- **Other Literature**

Developers

- **Development**
- **History**
- **Subversion**
- **Contributors**
- **Commercial licenses**

Tempo	Temperatura	Humidade	Vento	Jogar ténis
Ensolarado	Quente	Alta	Fraco	Não
Ensolarado	Quente	Alta	Forte	Não
Nublado	Quente	Alta	Fraco	Sim
Chuvoso	Moderada	Alta	Fraco	Sim
Chuvoso	Agradável	Normal	Fraco	Sim
Chuvoso	Agradável	Normal	Forte	Não
Nublado	Agradável	Normal	Forte	Sim
Ensolarado	Moderada	Alta	Fraco	Não
Ensolarado	Agradável	Normal	Fraco	Sim
Chuvoso	Moderada	Normal	Fraco	Sim
Ensolarado	Moderada	Normal	Forte	Sim
Nublado	Moderada	Alta	Fraco	Sim
Nublado	Quente	Normal	Fraco	Sim
Chuvoso	Moderado	Alta	Forte	Não

```
@relation playtennis
```

```
@attribute outlook {sunny, overcast, rainy}  
@attribute temperature {hot, mild, cool}  
@attribute humidity {high, normal}  
@attribute wind {weak, strong}  
@attribute play {yes, no}
```

```
@data  
sunny, hot, high, weak, no  
sunny, hot, high, strong, no  
overcast, hot, high, weak, yes  
rainy, mild, high, weak, yes  
rainy, cool, normal, weak, yes  
rainy, cool, normal, strong, no  
overcast, cool, normal, strong, yes  
sunny, mild, high, weak, no  
sunny, cool, normal, weak, yes  
rainy, mild, normal, weak, yes  
sunny, mild, normal, strong, yes  
overcast, mild, high, strong, yes  
overcast, hot, normal, weak, yes  
rainy, mild, high, strong, no
```

ARFF

Abrir no Weka

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: playtennis Attributes: 5
Instances: 14 Sum of weights: 14

Attributes
All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> outlook
2	<input type="checkbox"/> temperature
3	<input type="checkbox"/> humidity
4	<input type="checkbox"/> wind
5	<input type="checkbox"/> play

Remove

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

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

Class: play (Nom) Visualize All

Outlook	Blue (play=no)	Red (play=yes)
sunny	3	2
overcast	4	0
rainy	3	2

Status: OK Log x 0

Naive bayes

□ Abordagem Probabilística (Teorema de Bayes)

$$P(\text{vermelho}) = 7 / 12$$

$$P(\text{azul}) = 5 / 12$$

Probabilidades apriori

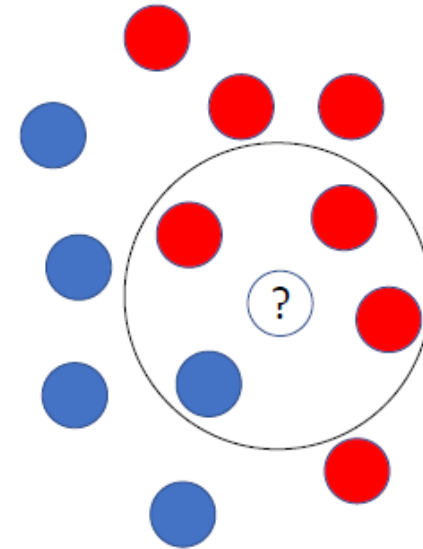
$$P'(\text{vermelho}) = 3 / 7$$

$$P'(\text{azul}) = 1 / 5$$

Probabilidades posteori

$$P''(\text{vermelho}) = 7 / 12 * 3 / 7 = 21 / 84 = \mathbf{0,25}$$

$$P''(\text{azul}) = 5 / 12 * 1 / 5 = 5 / 60 = \mathbf{0,08}$$



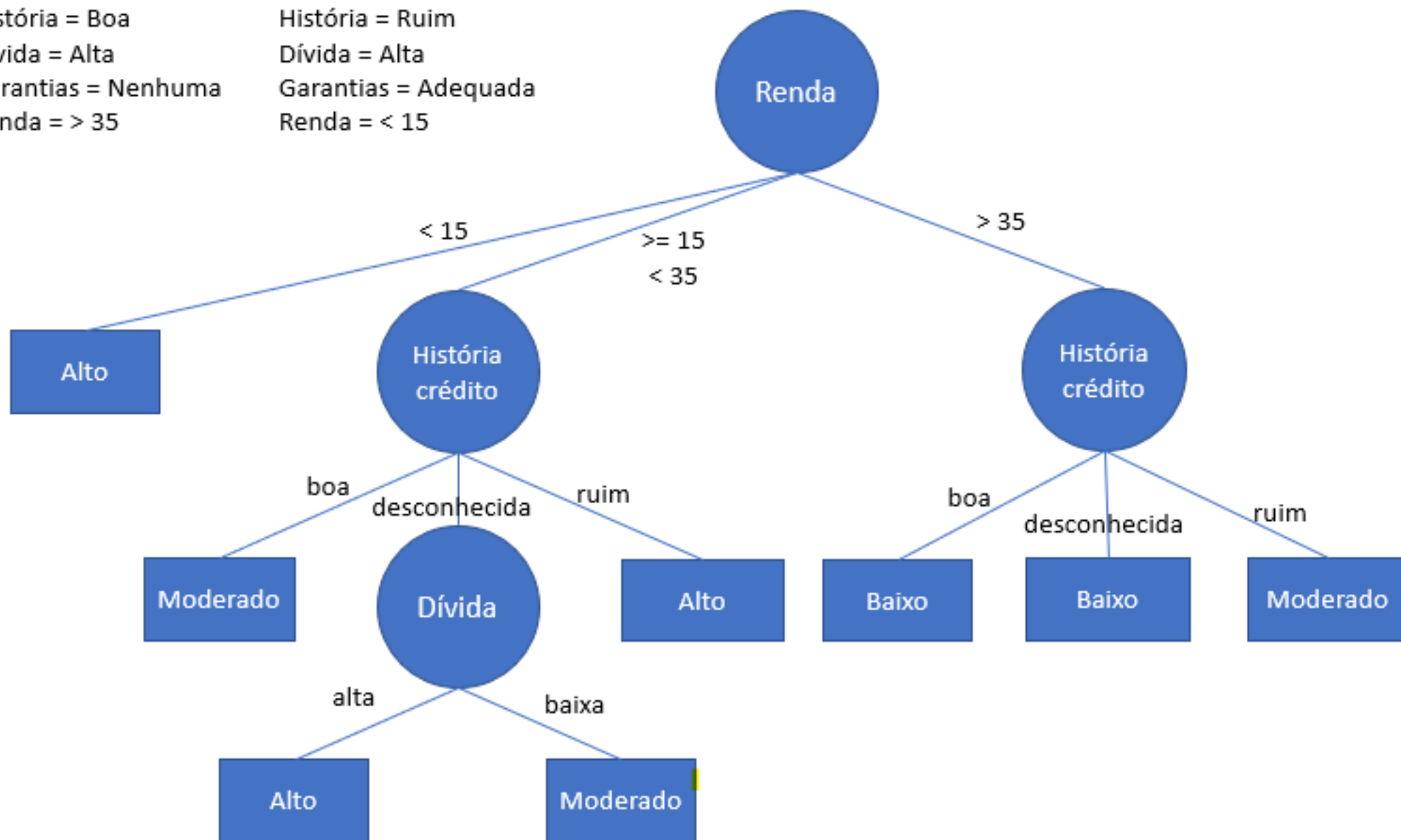
Árvore de Decisão

História do crédito	Dívida	Garantias	Renda anual	Risco
Ruim	Alta	Nenhuma	< 15.000	Alto
Desconhecida	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto
Desconhecida	Baixa	Nenhuma	>= 15.000 a <= 35.000	Moderado
Desconhecida	Baixa	Nenhuma	> 35.000	Alto
Desconhecida	Baixa	Nenhuma	> 35.000	Baixo
Desconhecida	Baixa	Adequada	> 35.000	Baixo
Ruim	Baixa	Nenhuma	< 15.000	Alto
Ruim	Baixa	Adequada	> 35.000	Moderado
Boa	Baixa	Nenhuma	> 35.000	Baixo
Boa	Alta	Adequada	> 35.000	Baixo
Boa	Alta	Nenhuma	< 15.000	Alto
Boa	Alta	Nenhuma	>= 15.000 a <= 35.000	Moderado
Boa	Alta	Nenhuma	> 35.0000	Baixo
Ruim	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto

Árvore de Decisão

História = Boa
Dívida = Alta
Garantias = Nenhuma
Renda = > 35

História = Ruim
Dívida = Alta
Garantias = Adequada
Renda = < 15



Árvore de Decisão

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

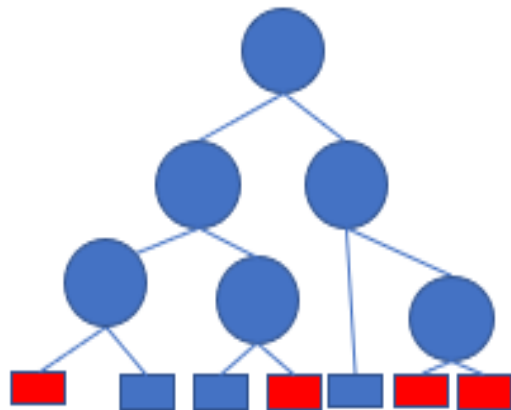
Random Forest (floresta randômica)



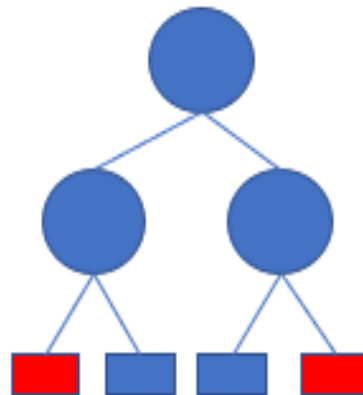
Random Forest (floresta randômica)

❑ Ensemble learning (**aprendizagem em conjunto**)

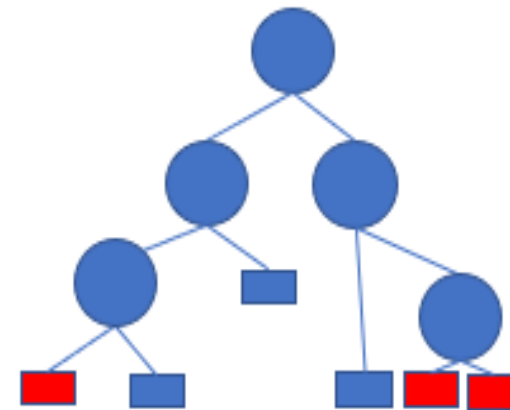
- “Consultar diversos profissionais para tomar uma decisão”
- Vários algoritmos juntos para construir um algoritmo mais “forte”
- Usa a média (regressão) ou votos da maioria (classificação) para dar a resposta final.



Risco = Alto



Risco = Baixo



Risco = Baixo

Naive Bayes no Weka

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: playtennis
Instances: 14
Attributes: 5
Sum of weights: 14

Attributes

All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> outlook
2	<input type="checkbox"/> temperature
3	<input type="checkbox"/> humidity
4	<input type="checkbox"/> wind
5	<input type="checkbox"/> play

Remove

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

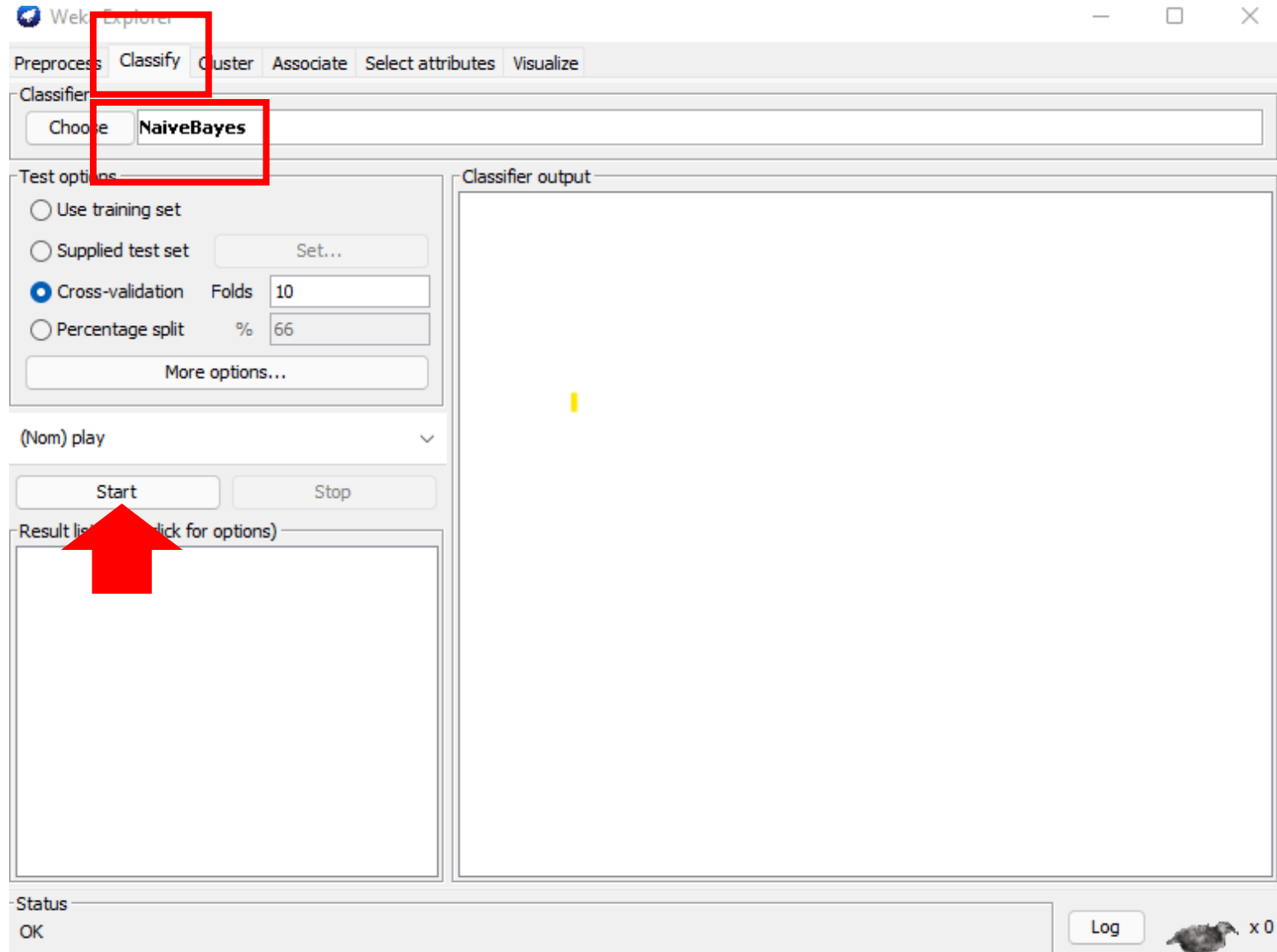
No.	Label	Count	Weight
1	sunny	5	5
2	overcast	4	4
3	rainy	5	5

Class: play (Nom) Visualize All

Outlook	Blue (play=no)	Red (play=yes)	Total
sunny	3	2	5
overcast	4	0	4
rainy	3	2	5

Status: OK Log x 0

Naive Bayes no Weka



Naive Bayes no Weka

weka explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose NaiveBayes

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)

09:38:35 - bayes.NaiveBayes

Classifier output

Correctly Classified Instances	8	57.1429 %
Incorrectly Classified Instances	6	42.8571 %
Kappa statistic	-0.0244	
Mean absolute error	0.4374	
Root mean squared error	0.4916	
Relative absolute error	91.8631 %	
Root relative squared error	99.6492 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
	0,778	0,800	0,636	0,778	0,700	-0,026
	0,200	0,222	0,333	0,200	0,250	-0,026
Weighted Avg.	0,571	0,594	0,528	0,571	0,539	-0,026

=== Confusion Matrix ===

a b <-- classified as

7 2 | a = yes

4 1 | b = no

Status: OK

Log x 0

Árvore de Decisão no Weka

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)

09:38:35 - bayes.NaiveBayes

09:57:52 - trees.J48

Classifier output

Correctly Classified Instances 7 50 %

Incorrectly Classified Instances 7 50 %

Kappa statistic -0.0426

Mean absolute error 0.4167

Root mean squared error 0.5984

Relative absolute error 87.5 %

Root relative squared error 121.2987 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
	0,556	0,600	0,625	0,556	0,588	-0,043
	0,400	0,444	0,333	0,400	0,364	-0,043
Weighted Avg.	0,500	0,544	0,521	0,500	0,508	-0,043


=== Confusion Matrix ===

a b <-- classified as

5 4 | a = yes

3 2 | b = no

Status OK

Log  x 0

Random Forest no Weka

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose RandomForest -P 100 -print -I 3 -num-slots 1 -K 0 -M 1.0 -V 0.00

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)

09:38:35 - bayes.NaiveBayes

09:57:52 - trees.J48

Status

OK

Class for constructing a forest of random trees.

More

Capabilities

bagSizePercent 100

batchSize 100

breakTiesRandomly False

calcOutOfBag False

computeAttributeImportance False

debug False

doNotCheckCapabilities False

maxDepth 0

numDecimalPlaces 2

numExecutionSlots 1

numFeatures 0

numIterations 3

outputOutOfBagComplexityStatistics False

printClassifiers True

seed 1

storeOutOfBagPredictions False

Open... Save... OK Cancel

Classifier output

Correctly Classified Instances 100.0%

Incorrectly Classified Instances 0.0%

Kappa statistic 1.000

Mean absolute error 0.000

Root mean squared error 0.000

Relative absolute error 0.000

Root relative squared error 0.000

Total Number of Instances 10

=== Detailed Accuracy By Class ===

	TP	Rate
0	5	0.556
1	4	0.400
Weighted Avg.		0.500

=== Confusion Matrix ===

a \ b	<-- classified as
5 4	a = yes
3 2	b = no

Random Forest no Weka

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose RandomForest -P 100 -print -I 3 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

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)

- 09:38:35 - bayes.NaiveBayes
- 09:57:52 - trees.J48
- 10:21:28 - trees.RandomForest
- 10:21:59 - trees.RandomForest

Classifier output

Size of the tree : 8

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	8	57.1429 %
Incorrectly Classified Instances	6	42.8571 %
Kappa statistic	0.0667	
Mean absolute error	0.4296	
Root mean squared error	0.5586	
Relative absolute error	90.2143 %	
Root relative squared error	113.2225 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
	0,667	0,600	0,667	0,667	0,667	0,067
	0,400	0,333	0,400	0,400	0,400	0,067

Status

OK

Log x 0

Exercício

❑ Considerando as bases de dados:

1. **Características;**
2. **Letter;**
3. **Risco de Crédito;**
4. **Zoo;**

❑ Aplique Naive Bayes, Árvore de Decisão e Random Forest, compare os resultados.