

Marcelo d'Almeida  
[md@id.uff.br](mailto:md@id.uff.br)

# Trabalho Ferramenta Weka

*Tennis Major Tournaments Match Statistics Dataset e Contact Lens Dataset*

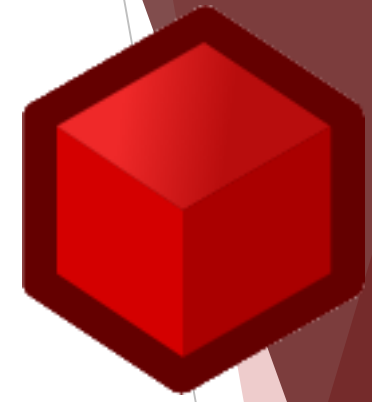
Mineração de Dados (2015.1)



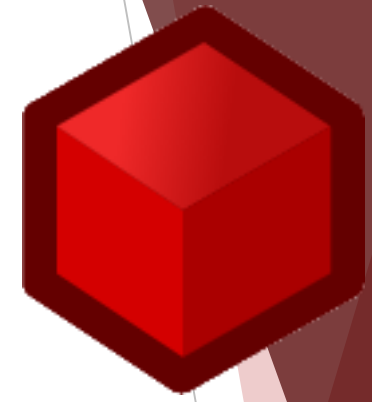
CCD-UFF  
Clube de Ciência de Dados

# Conteúdo

- ▶ Ferramenta Weka
- ▶ Base de Dados
- ▶ Classificação
- ▶ Associação
- ▶ Clusterização
- ▶ Trabalhos Futuros

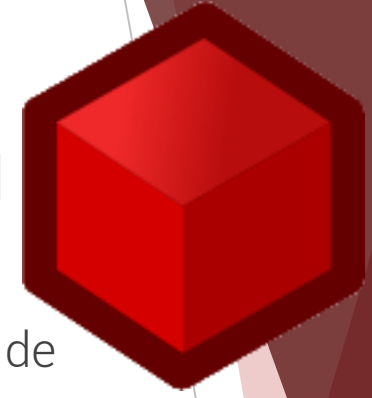


Ferramenta Weka



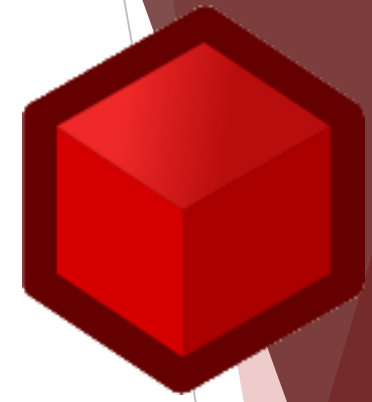
# Ferramenta Weka

[1]

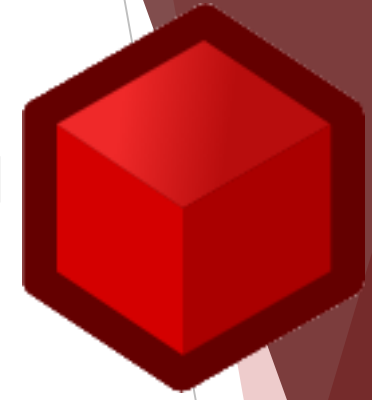


- ▶ Coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Algoritmos que podem ser aplicados diretamente aos datasets ou serem chamados através de código Java
- ▶ Possui ferramentas de pré-processamento, classificação, regressão, clusterização, regras de associação e visualização

Base de Dados



# Base de Dados – Tennis Major Tournaments Match Statistics Dataset [2]



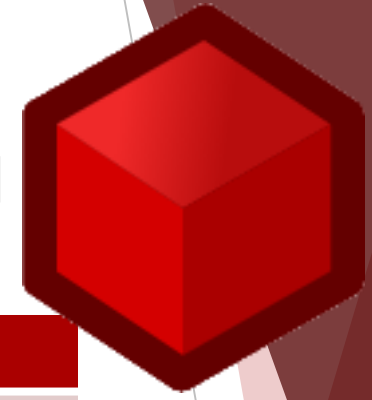
- ▶ Retirado do *UCI Machine Learning Repository*
- ▶ Coleção de 8 arquivos contendo estatísticas de partidas masculinas e femininas nos 4 maiores torneios de tênis, no ano de 2013
  - ▶ Australian Open
  - ▶ French Open - Roland Garros
  - ▶ Wimbledon
  - ▶ US Open
- ▶ Cada arquivo contém 42 colunas e mínimo de 76 linhas

# Base de Dados – Tennis Major Tournaments Match Statistics Dataset [2]



Variável	Significado	Tipo
Player (1 ou 2)	Nome do jogador	Nominal
Result	'1' se Jogador 1 venceu, '0' se Jogador 2 venceu	(0/1)

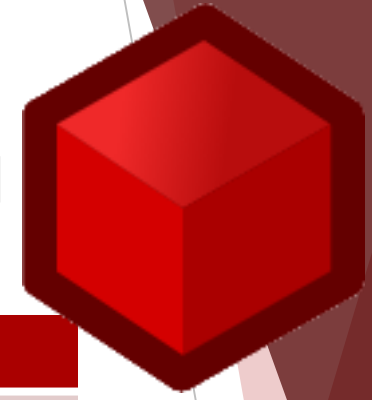
# Base de Dados – Tennis Major Tournaments Match Statistics Dataset [2]



Variável	Significado	Tipo
FSP (.1 ou .2)	<b>Primeiro Serviço</b> (porcentagem)	Real Number
FSW (.1 ou .2)	<b>Primeiro Serviço</b> (ganho)	Real Number
SSP (.1 ou .2)	<b>Primeiro Serviço</b> (porcentagem)	Real Number
SSW (.1 ou .2)	<b>Segundo Serviço</b> (ganho)	Real Number
ACE (.1 ou .2)	<b>Aces</b> ganhos	Numeric-Integer
DBF (.1 ou .2)	<b>Faltas duplas</b> cometidas	Numeric-Integer
WNR (.1 ou .2)	<b>Winners</b> ganhos	Numeric
UFE (.1 ou .2)	<b>Erros ‘não forçados’</b> cometidos	Numeric



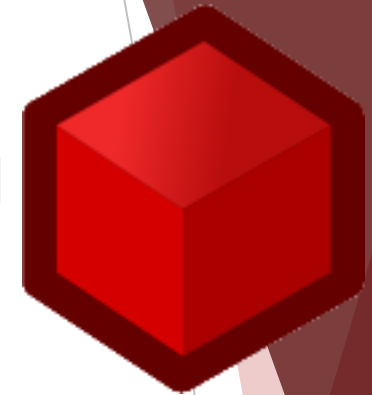
# Base de Dados – Tennis Major Tournaments Match Statistics Dataset [2]



Variável	Significado	Tipo
BPC (.1 ou .2)	<b>Break Points</b> criados	Numeric
BPW (.1 ou .2)	<b>Break Points</b> ganhos	Numeric
NPA (.1 ou .2)	<b>Net Points</b> tentados	Numeric
NPW (.1 ou .2)	<b>Net Points</b> ganhos	Numeric
TPW (.1 ou .2)	Total de pontos ganhos	Numeric
ST1 .. ST5 (.1 ou .2)	Resultado do <b>SET</b>	Numeric-Integer
FNL (.1 ou .2)	Números de <b>Games</b> vencidos	Numeric-Integer
Round	Rodada do torneio	Numeric-Integer

# Base de Dados – Contact Lens Dataset

[3]



- ▶ Retirado do *Sample Weka Data Sets*
- ▶ O dataset reúne informações que determinam se o paciente deveria (ou não) usar lentes de contato e se elas deveriam ser rígidas ou maleáveis.
- ▶ Contém 5 atributos e 24 instâncias

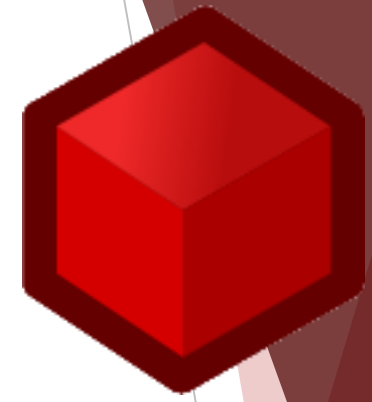
# Base de Dados – Contact Lens Dataset

[3]



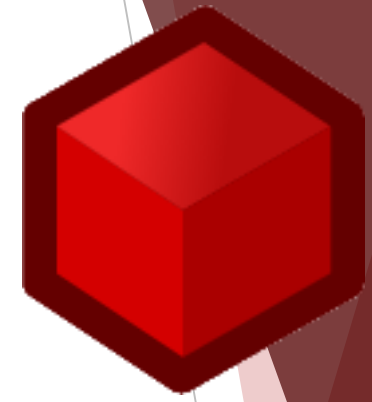
Variável	Significado	Valores possíveis
Age	Idade	Young, Pre-presbyopic, Presbyopic
Spectacle prescription	Se a pessoa possui <b>Miopia</b> ou <b>Hipermetropia</b>	Myope, Hypermetrope
Astigmatic	<b>Astigmatismo</b>	Yes, No
Tear production rate	<b>Taxa de produção de lágrima</b>	Reduced, Normal
Contact lenses	<b>Tipo de lente de contato</b>	Soft, Hard, None

# Classificação



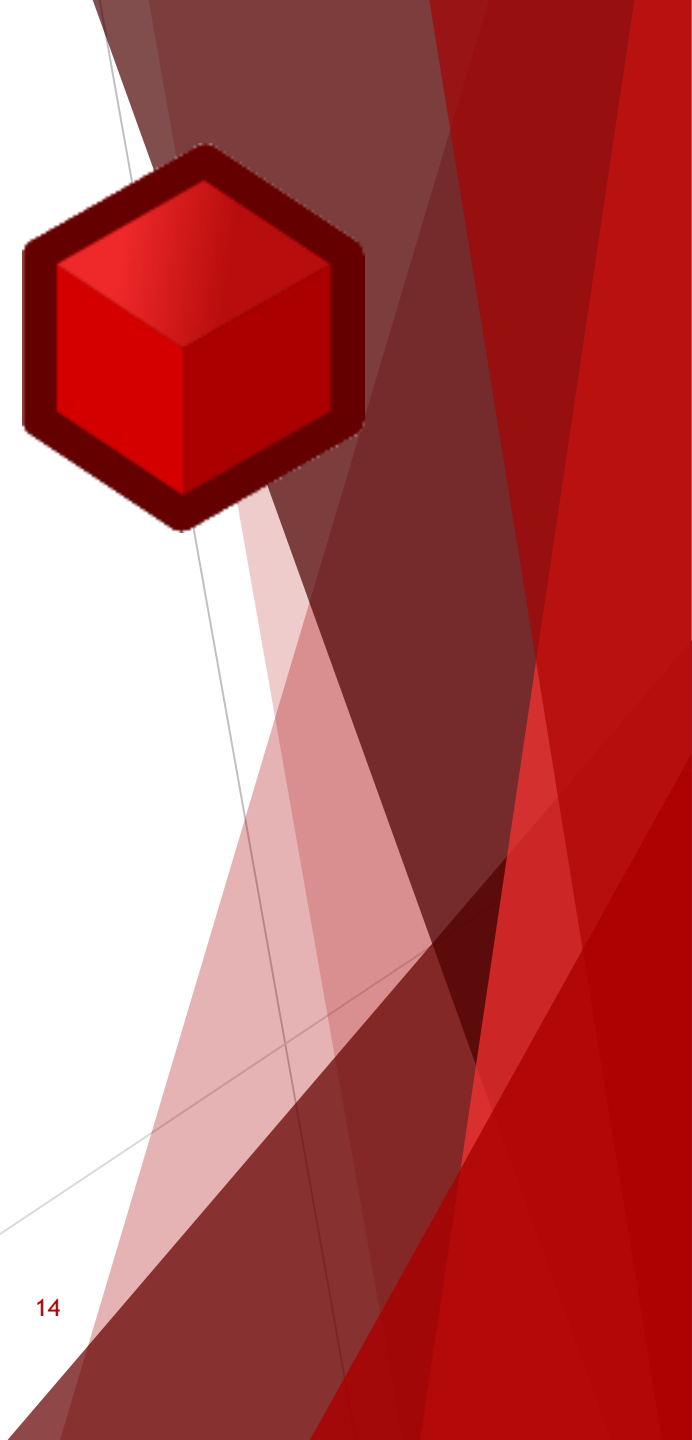
# Classificação

- ▶ Base utilizada: Tennis Major Tournaments Match Statistics Dataset
  - ▶ Torneio: AustraliaOpen
  - ▶ Modalidade: Masculino (men)
- ▶ Pré-processamento
- ▶ Algoritmos usados
  - ▶ NaiveBayes
  - ▶ IBk (k-NN)
  - ▶ J48 (Árvore de decisão)
- ▶ Leave-one-out
- ▶ Comparação



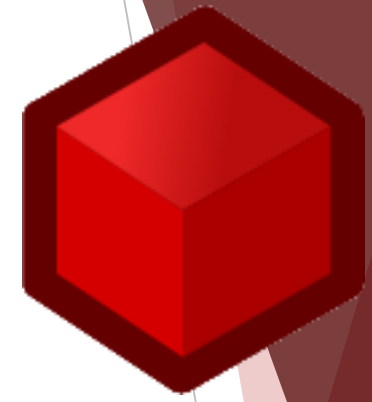
# Classificação – Pré-processamento

- ▶ Duas configurações foram utilizadas
  - ▶ 1 - Todos os atributos
  - ▶ 2 – Atributos 'trivialmente óbvios' removidos
    - ▶ Número total de *Games* vencidos
    - ▶ Resultados dos *Sets*
    - ▶ Número total de pontos
    - ▶ Números de *Winners* obtidos
- ▶ O atributo 'Result' convertido como binário e usado como atributo classe



# Classificação – NaiveBayes

- ▶ Parâmetros testados
  - ▶ Default
  - ▶ useKernelEstimator = True
  - ▶ useSupervisedDiscretization = True



# Classificação – NaiveBayes

## Resultados

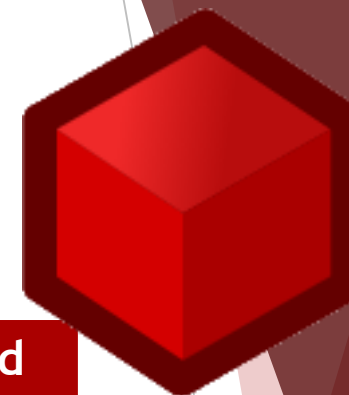


Todos os atributos	Default			Use Kernel Estimator			Use Supervised Discretization		
Classificadas corretamente	98.4127 %			99.2063 %			97.619 %		
	a	b		a	b		a	b	
Matriz de confusão	57	2	a	58	1	a	58	1	a
	0	67	b	0	67	b	2	65	b



# Classificação – NaiveBayes

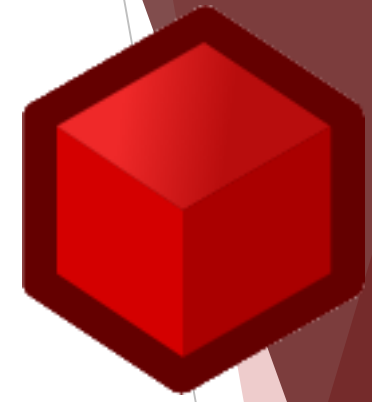
## Resultados



'Trivialmente óbvios' removidos	Default			Use Kernel Estimator			Use Supervised Discretization		
Classificadas corretamente	85.7143 %			88.8889 %			77.7778 %		
	a	b		a	b		a	b	
Matriz de confusão	55	4	a	56	3	a	48	11	a
	14	53	b	11	56	b	17	50	b

# Classificação – IBk

- ▶ Parâmetros testados
  - ▶ Variação do  $k$  de 1 - 20



# Classificação – IBk

## Resultados



Todos os atributos	Classificados corretamente		Classificados corretamente
k = 1	92.8571 %	k = 11	96.8254 %
k = 2	93.6508 %	k = 12	96.8254 %
k = 3	95.2381 %	k = 13	97.619 %
k = 4	95.2381 %	k = 14	97.619 %
k = 5	94.4444 %	k = 15	97.619 %
k = 6	95.2381 %	k = 16	97.619 %
k = 7	96.8254 %	k = 17	97.619 %
k = 8	97.619 %	k = 18	<b>98.4127 %</b>
k = 9	96.0317 %	k = 19	97.619 %
k = 10	97.619 %	k = 20	97.619 %

# Classificação – IBk

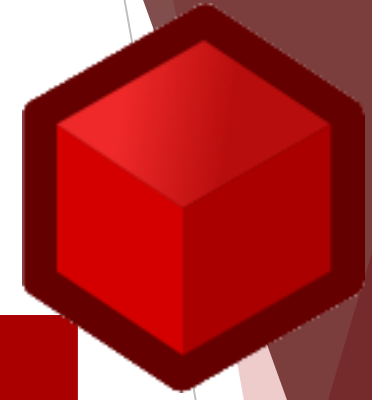
## Resultados



Todos os atributos	k = 1			k = 2			k = 18		
Classificadas corretamente	92.8571 %			93.6508 %			98.4127 %		
	a	b		a	b		a	b	
Matriz de confusão	53	6	a	58	1	a	58	1	a
	3	64	b	7	60	b	1	66	b

# Classificação – IBk

## Resultados



'Trivialmente óbvios' removidos	Classificados corretamente		Classificados corretamente
k = 1	72.2222 %	k = 11	84.127 %
k = 2	78.5714 %	k = 12	86.5079 %
k = 3	76.1905 %	k = 13	85.7143 %
k = 4	81.746 %	k = 14	85.7143 %
k = 5	80.9524 %	k = 15	87.3016 %
k = 6	79.3651 %	k = 16	86.5079 %
k = 7	82.5397 %	k = 17	86.5079 %
k = 8	84.127 %	k = 18	86.5079 %
k = 9	81.746 %	k = 19	88.0952 %
k = 10	83.3333 %	k = 20	88.0952 %

# Classificação – IBk

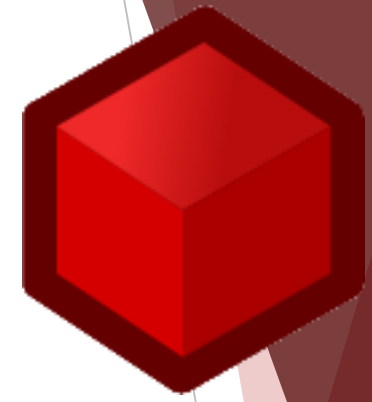
## Resultados



'Trivialmente óbvios' removidos	k = 1			k = 2			k = 20		
Classificadas corretamente	72.2222 %			78.5714 %			88.0952 %		
Matriz de confusão	a	b		a	b		a	b	
	38	21	a	51	8	a	50	9	a
	14	53	b	19	48	b	6	61	b

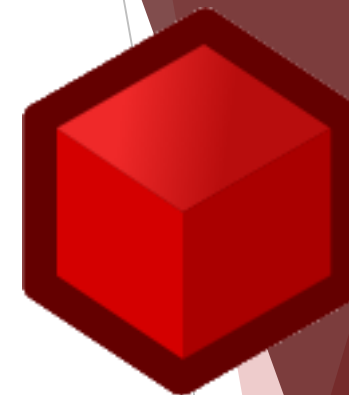
# Classificação – J48

- ▶ Parâmetros testados
  - ▶ Default
  - ▶ ReducedErrorPruning = True



# Classificação – J48

## Resultados

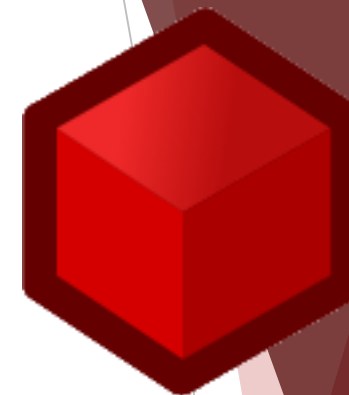


Todos os atributos	Default			Reduced Error Pruning		
Classificadas corretamente	97.619 %			98.4127 %		
	a	b		a	b	
Matriz de confusão	57	2	a	57	2	a
	1	66	b	0	67	b



# Classificação – J48

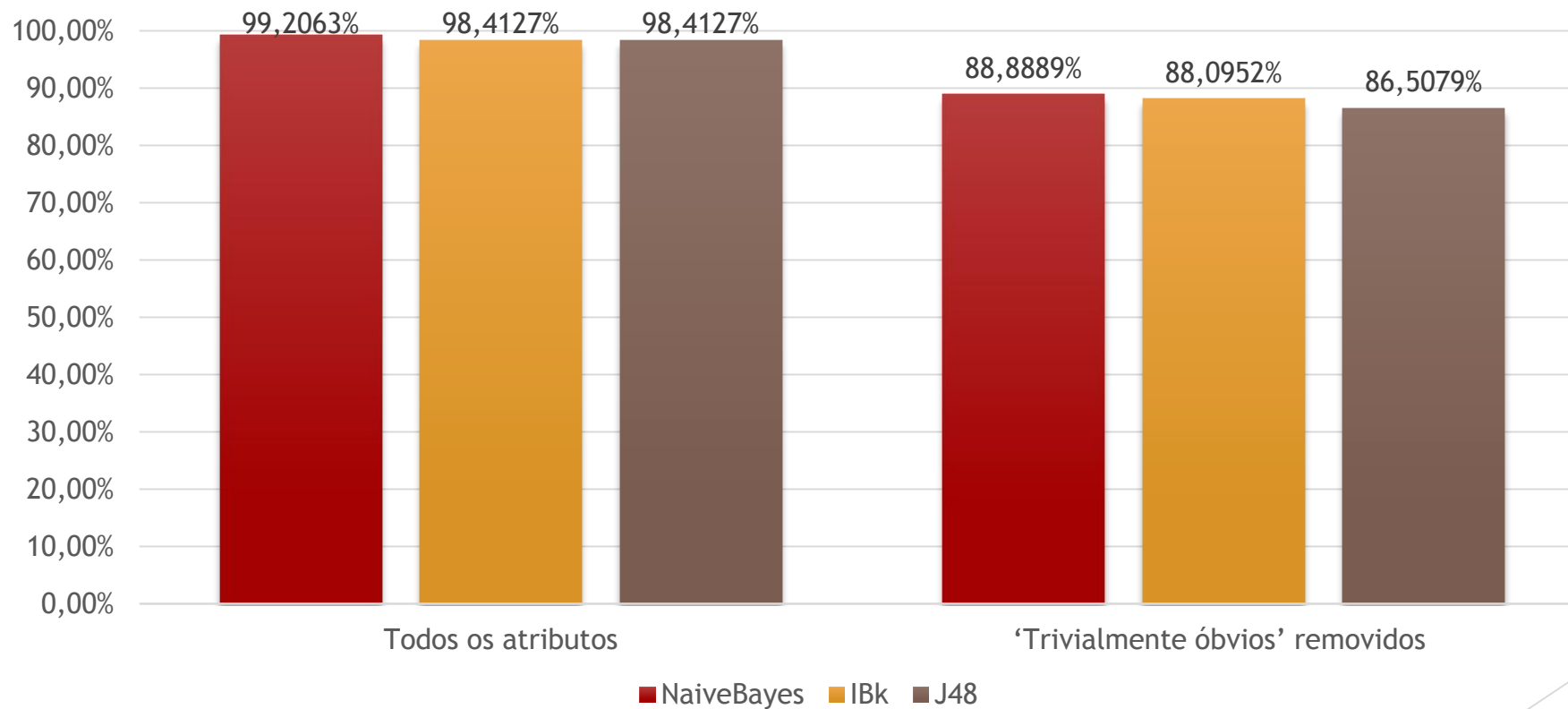
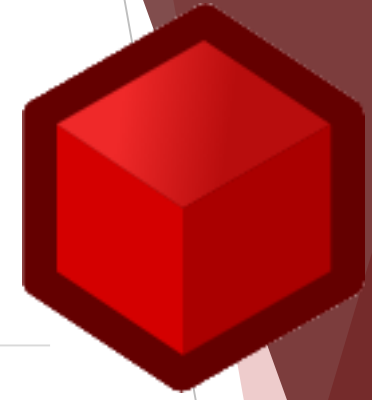
## Resultados



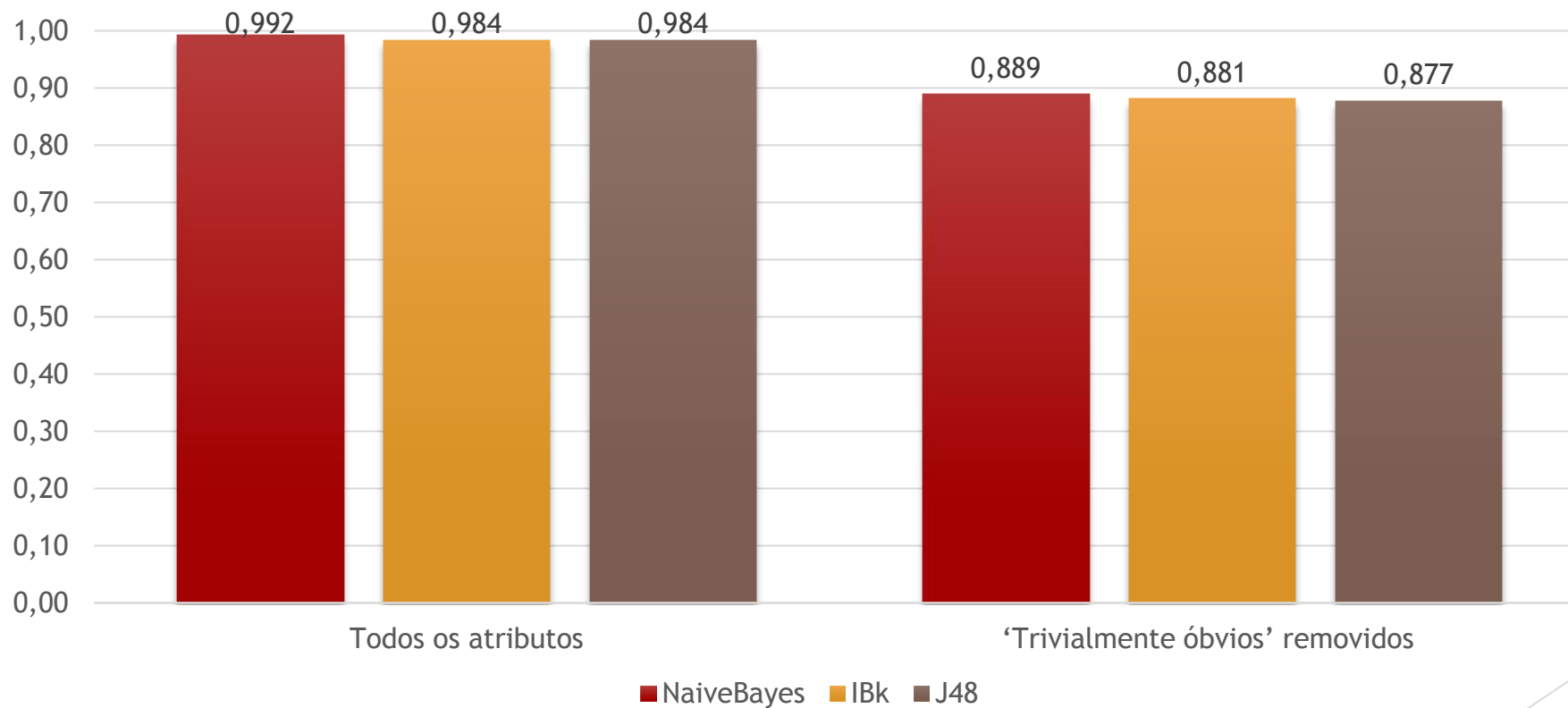
'Trivialmente óbvios' removidos	Default			Reduced Error Pruning		
Classificadas corretamente	84.9206 %			86.5079 %		
	a	b		a	b	
Matriz de confusão	48	11	a	53	6	a
	8	59	b	11	56	b

# Classificação – Comparação

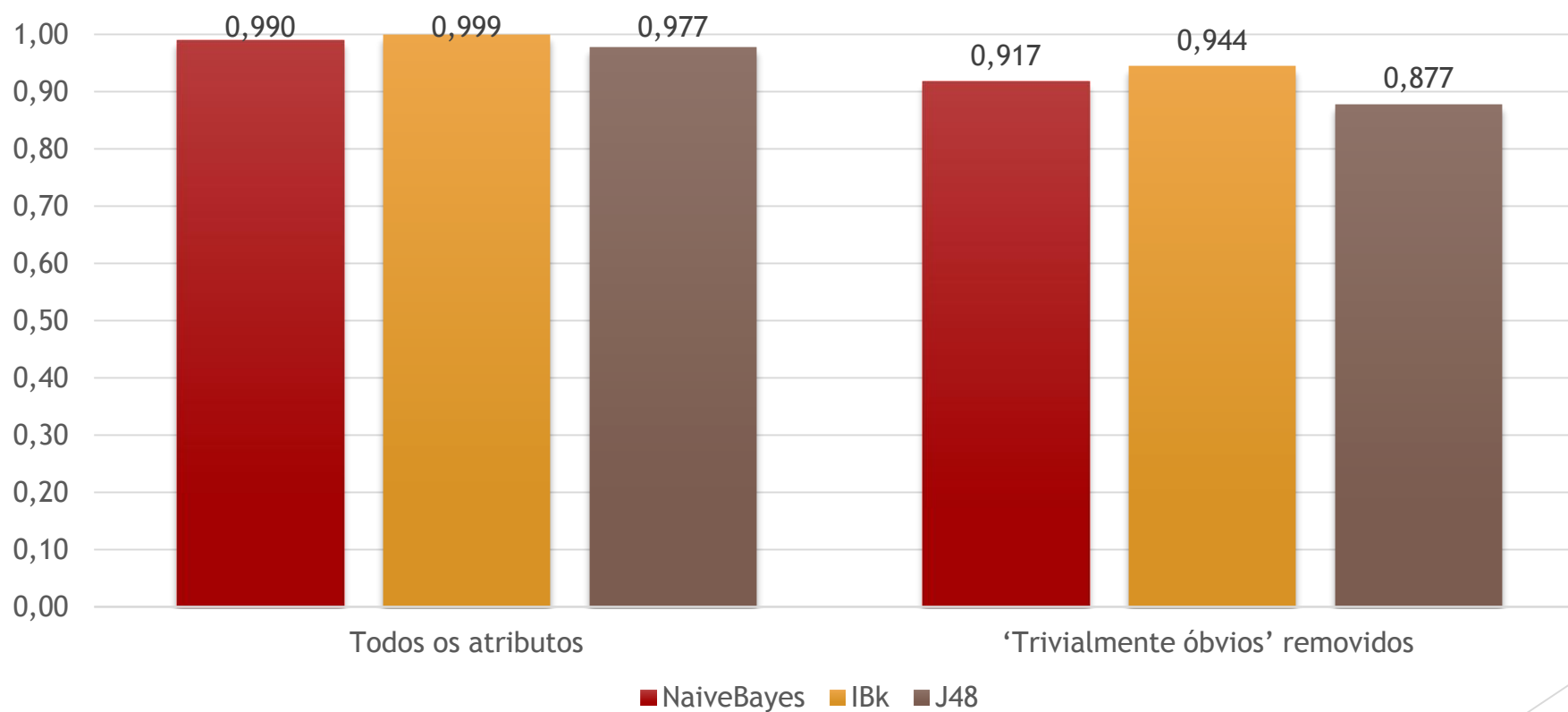
## Instâncias classificadas corretamente



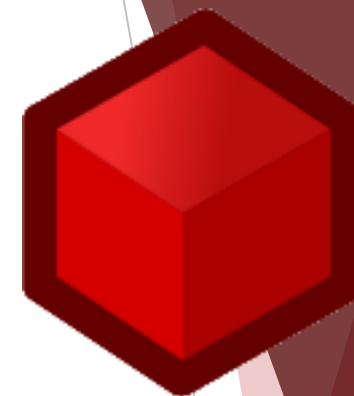
# Classificação – Comparação F-Measure (média ponderada)



# Classificação – Comparação ROC area (média ponderada)

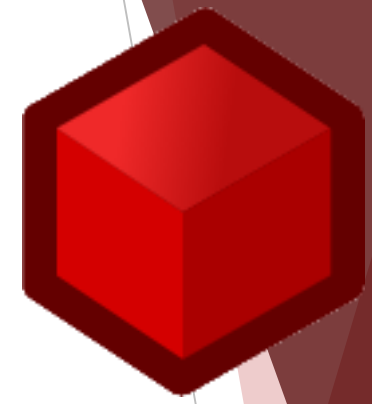


Associação



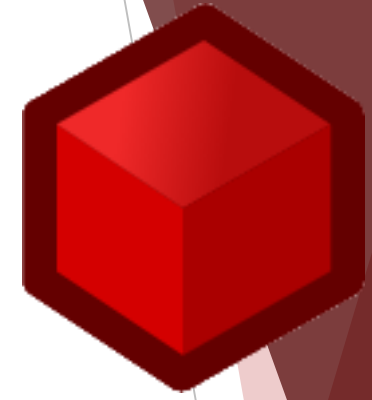
# Associação

- ▶ Base utilizada: Contact Lens Dataset
- ▶ Algoritmo usado
  - ▶ Apriori
- ▶ Parâmetros testados
  - ▶ Suporte mínimo: 0.1, 0.2 e 0.3
  - ▶ Confiança mínima: 0.75, 0.8 e 1
  - ▶ Lift: 1.5, 2.0, 4.0 e 6.0
  - ▶ Car: True; False



# Associação – Apriori

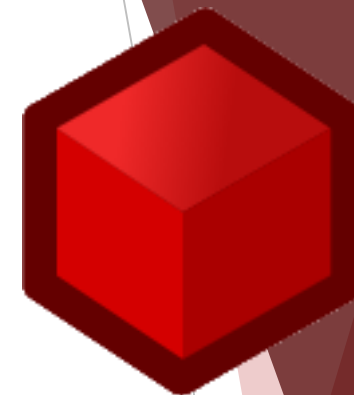
## Total de regras mineiradas



Conf \ MinSup (‘car’=True)	0.1	0.2	0.3
0.75	108 (23)	18 (8)	2 (-)
0.8	89 (16)	15 (7)	2 (-)
1.00	83 (14)	10 (5)	1 (-)

# Associação – Apriori

## Total de regras mineiradas

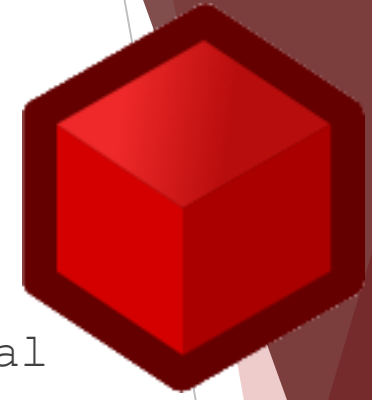


Lift \ MinSup	0.1	0.2	0.3
1.5	320	28	2
2.0	166	10	-
4.0	26	2	-
6.0	4	-	-



# Associação – Apriori

## Regras interessantes



contact-lenses=soft ==> astigmatism=no tear-prod- rate=normal

**minsup: (0.2) conf: (1) <lift: (4)>**

astigmatism=no tear-prod-rate=normal ==> contact- lenses=soft

**minsup: (0.2) conf: (0.83) <lift: (4)>**

tear-prod-rate=reduced ==> contact-lenses=none

**minsup: (0.3) conf: (1) <lift: (1.6)>**

spectacle-prescrip=myope astigmatism=yes tear-prod-  
rate=normal ==> contact-lenses=hard

**minsup: (0.1) conf: (1) <lift: (6)>**

age=young astigmatism=yes tear-prod-rate=normal ==> contact-  
lenses=hard

**minsup: (0.1) conf: (1) <lift: (6)>**

# Associação – Apriori

## Regras interessantes

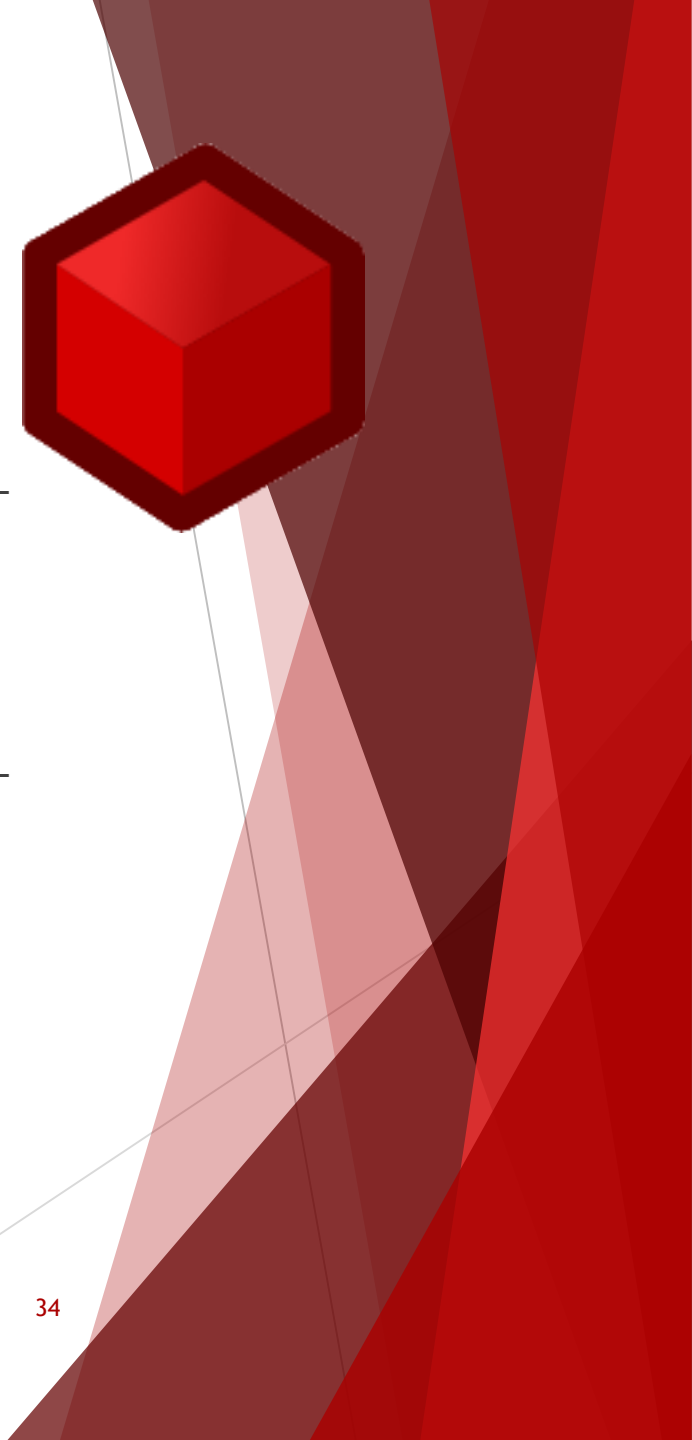
spectacle-prescrip=hypermetrope astigmatism=yes ==> contact-lenses=none      **minsup: (0.2) conf: (0.83)**

age=presbyopic ==> contact-lenses=none  
                         **minsup: (0.2) conf: (0.75)**

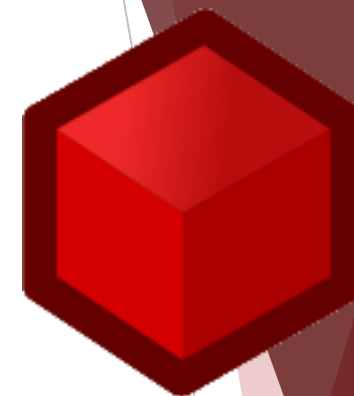
spectacle-prescrip=hypermetrope astigmatism=yes ==> contact-lenses=none      **minsup: (0.1) conf: (0.83)**

age=presbyopic ==> contact-lenses=none  
                         **minsup: (0.1) conf: (0.75)**

**'car' = True ==> No Lift measured**

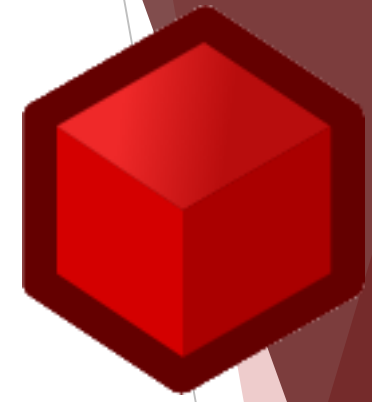


# Clusterização



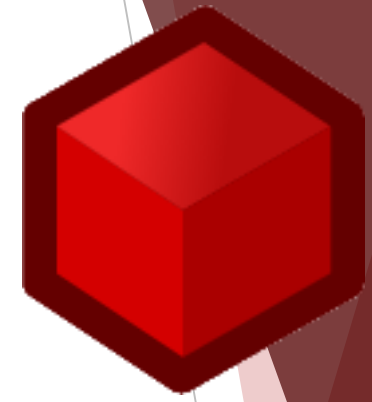
# Clusterização

- ▶ Base utilizada: Tennis Major Tournaments Match Statistics Dataset
  - ▶ Torneio: AustraliaOpen
  - ▶ Modalidade: Masculino (men)
- ▶ Pré-processamento
  - ▶ Retirado o atributo 'Result' (tido como atributo classe)
- ▶ Algoritmos usados
  - ▶ SimpleKMeans (k-means)
  - ▶ DBSCAN



# Clusterização – SimpleKMeans

- ▶ Parâmetros testados
  - ▶ Número de clusters: 2, 3, 4 e 5



# Clusterização – SimpleKMeans

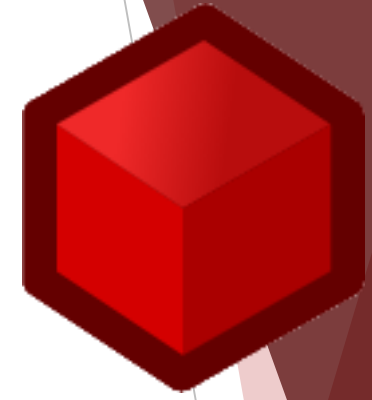
## Resultado



k - cluster \ Distância	Euclidiana
2 - 0	49 (39%)
- 1	77 (61%)
3 - 0	21 (17%)
- 1	33 (26%)
- 2	72 (57%)
4 - 0	23 (18%)
- 1	33 (26%)
- 2	32 (25%)
- 3	38 (30%)

# Clusterização – SimpleKMeans

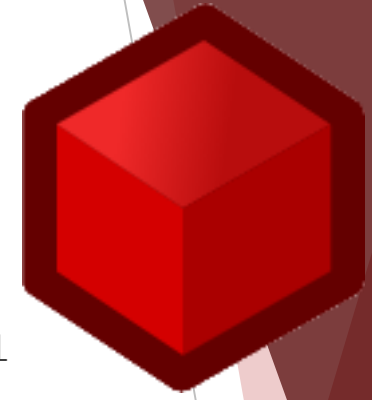
## Resultado



k - cluster \ Distância	Euclidiana
5 - 0	17 (13%)
- 1	25 (20%)
- 2	21 (17%)
- 3	17 (13%)
- 4	46 (37%)

# Clusterização – SimpleKMeans

## k = 2

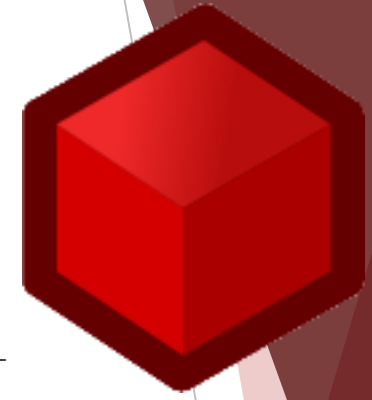


Attribute	Full Data (126.0)	0 (49.0)	1 (77.0)
=====			
<b>Player1</b>	Rafael Nadal	<b>Tommy Robredo</b>	<b>Rafael Nadal</b>
<b>Player2</b>	Roger Federer	<b>David Ferrer</b>	<b>Roger Federer</b>
<b>FNL1</b>	1.881	<b>2.3265</b>	<b>1.5974</b>
<b>FNL2</b>	1.7063	<b>2.0816</b>	<b>1.4675</b>
FSP.1	61.3254	59.5306	62.4675
FSP.2	61.4444	61.2245	61.5844
FSW.1	48.9444	60.9388	41.3117
FSW.2	49	62.4286	40.4545
SSP.1	38.6746	40.4694	37.5325
SSP.2	38.5556	38.7755	38.4156
SSW.1	21.754	28.7347	17.3117
SSW.2	21.5635	27.7755	17.6104



# Clusterização – SimpleKMeans

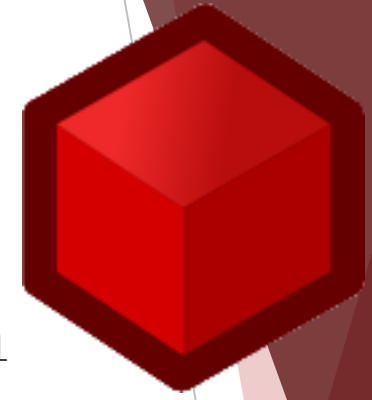
## k = 2



Attribute	Full Data (126.0)	0 (49.0)	1 (77.0)
Player1	Rafael Nadal	Tommy Robredo	Rafael Nadal
Player2	Roger Federer	David Ferrer	Roger Federer
<b>ACE.1</b>	10.5317	<b>14.1633</b>	<b>8.2208</b>
<b>ACE.2</b>	8.9762	<b>10.6735</b>	<b>7.8961</b>
<b>DBF.1</b>	4.344	<b>5.7755</b>	<b>3.433</b>
<b>DBF.2</b>	4.608	<b>5.8776</b>	<b>3.8001</b>
WNR.1	33.8333	45.3061	26.5325
WNR.2	32.127	39.449	27.4675
<b>UFE.1</b>	32.3968	<b>42.7959</b>	<b>25.7792</b>
<b>UFE.2</b>	33.9286	<b>41.7959</b>	<b>28.9221</b>

# Clusterização – SimpleKMeans

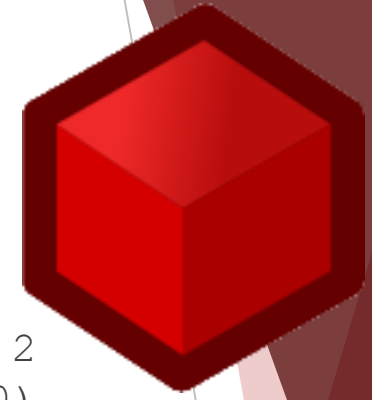
## k = 2



Attribute	Full Data (126.0)	0 (49.0)	1 (77.0)
Player1	Rafael Nadal	Tommy Robredo	Rafael Nadal
Player2	Roger Federer	David Ferrer	Roger Federer
BPC.1	3.8095	4.8163	3.1688
BPC.2	3.4048	4.2245	2.8831
BPW.1	9.5317	12.9184	7.3766
BPW.2	8.5635	10.6122	7.2597
NPA.1	14.4231	17.2002	12.6558
NPA.2	16.2019	19.833	13.8912
NPW.1	22.3846	26.5196	19.7532
NPW.2	23.8846	29.4074	20.3701
<b>TPW.1</b>	112.1587	<b>143.898</b>	<b>91.961</b>
<b>TPW.2</b>	110.0714	<b>140.6531</b>	<b>90.6104</b>

# Clusterização – SimpleKMeans

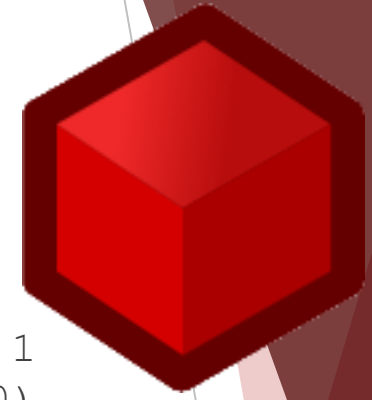
## k = 3



Attribute	0 (21.0)	1 (33.0)	2 (72.0)
Player1	Andreas Seppi	Tommy Robredo	Rafael Nadal
Player2	Jarkko Nieminen	Grigor Dimitrov	Roger Federer
FNL1	2.0476	2.4242	1.5833
FNL2	2.6667	1.7576	1.4028
ACE.1	11.7619	15.0606	8.0972
ACE.2	10	12.0606	7.2639
DBF.1	6.619	5.2424	3.2687
DBF.2	6.9524	4.9697	3.7584
UFE.1	40.8571	39.8485	26.5139
UFE.2	42.9048	40.2727	28.4028
TPW.1	146.6667	139.8485	89.4028
TPW.2	151.3333	132.2424	87.875

# Clusterização – SimpleKMeans

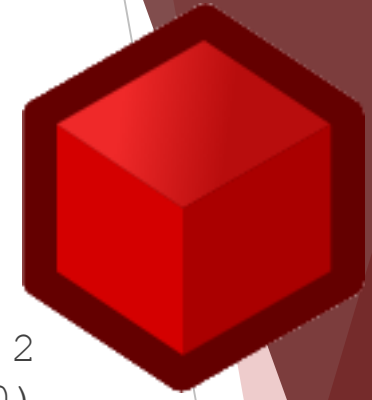
## k = 2 (análise de desvio padrão)



Attribute	Full Data		0		1	
	(126.0)		(49.0)		(77.0)	
=====						
<b>FNL1</b>	1.881	<b>+/-1.2433;</b>	2.3265	<b>+/-0.8263;</b>	1.5974	<b>+/-1.379;</b>
<b>FNL2</b>	1.7063	<b>+/-1.3088;</b>	2.0816	<b>+/-0.8859;</b>	1.4675	<b>+/-1.4742;</b>
<b>ACE.1</b>	10.5317	<b>+/-7.5352;</b>	14.1633	<b>+/-8.4295;</b>	8.2208	<b>+/-5.8818;</b>
<b>ACE.2</b>	8.9762	<b>+/-6.2064;</b>	10.6735	<b>+/-7.4453;</b>	7.8961	<b>+/-5.0304;</b>
<b>DBF.1</b>	4.344	<b>+/-3.1448;</b>	5.7755	<b>+/-3.4777;</b>	3.433	<b>+/-2.5427;</b>
<b>DBF.2</b>	4.608	<b>+/-3.4639;</b>	5.8776	<b>+/-3.8331;</b>	3.8001	<b>+/-2.9587;</b>
<b>UFE.1</b>	32.3968	<b>+/-19.5788;</b>	42.7959	<b>+/-21.9971;</b>	25.7792	<b>+/-14.5182;</b>
<b>UFE.2</b>	33.9286	<b>+/-20.5427;</b>	41.7959	<b>+/-22.6973;</b>	28.9221	<b>+/-17.4188;</b>
<b>TPW.1</b>	112.1587	<b>+/-33.7213;</b>	143.898	<b>+/-23.9811;</b>	91.961	<b>+/-21.0772;</b>
<b>TPW.2</b>	110.0714	<b>+/-34.8692;</b>	140.6531	<b>+/-23.7351;</b>	90.6104	<b>+/-25.6372;</b>

# Clusterização – SimpleKMeans

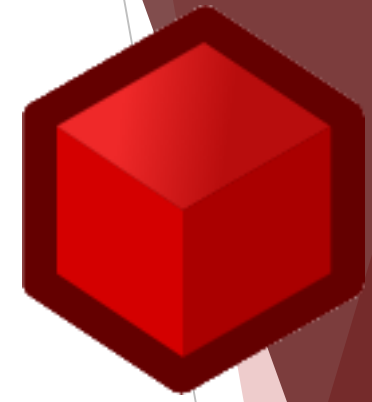
## k = 3 (análise de desvio padrão)



Attribute	0 (21.0)	1 (33.0)	2 (72.0)
<b>FNL1</b>	2.0476 +/-0.74;	2.4242 +/-0.9024;	1.5833 +/-1.4017;
<b>FNL2</b>	2.6667 +/-0.5774;	1.7576 +/-0.9024;	1.4028 +/-1.4791;
<b>ACE.1</b>	11.7619 +/-6.9131;	15.0606 +/-9.0896;	8.0972 +/-5.7682;
<b>ACE.2</b>	10 +/-7.3485;	12.0606 +/-7.437;	7.2639 +/-4.469;
<b>DBF.1</b>	6.619 +/-4.165;	5.2424 +/-2.8288;	3.2687 +/-2.4165;
<b>DBF.2</b>	6.9524 +/-4.0184;	4.9697 +/-3.5221;	3.7584 +/-2.9267;
<b>UFE.1</b>	40.8571 +/-23.7598;	39.8485 +/-21.4551;	26.5139 +/-15.0436;
<b>UFE.2</b>	42.9048 +/-26.6625;	40.2727 +/-22.4809;	28.4028 +/-15.5379;
<b>TPW.1</b>	146.6667 +/-22.283;	139.8485 +/-24.0833;	89.4028 +/-19.1574;
<b>TPW.2</b>	151.3333 +/-14.044;	132.2424 +/-25.0175;	87.875 +/-23.9018;

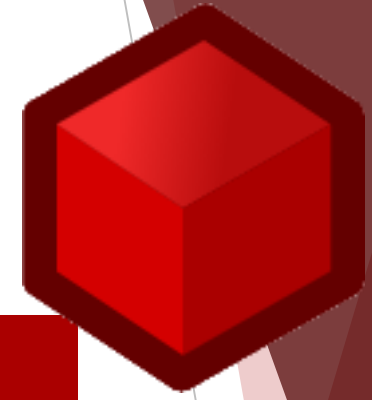
# Clusterização – DBSCAN

- ▶ Parâmetros testados
  - ▶ Epsilon (raio): 2, 2.37 e 2.5
  - ▶ Mínimo de pontos (densidade) 2, 3, 6 e 10
  - ▶ Default = (raio = 0.9, densidade = 6)
  - ▶ (raio = 4, densidade = 20)



# Clusterização – DBSCAN

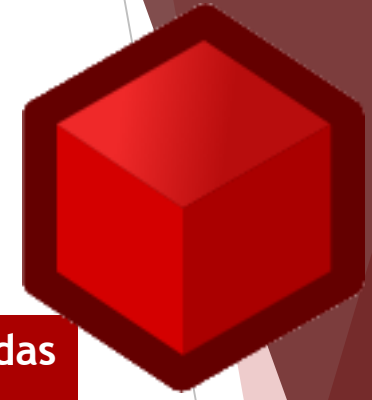
## Resultados



Configuração (raio, densidade) - cluster	Resultado da clusterização	Instâncias não clusterizadas
Default (0.9, 6)	-	126
(2.5, 10) - 0	80 (100%)	46
(2.5, 6) - 0	89 (100%)	37
(2.5, 3) - 0	97 (92%)	21
- 1	5 (5%)	
- 2	3 (3%)	
(2.5, 2) - 0	97 (89%)	17
- 1	5 (5%)	
- 2	2 (2%)	
- 3	3 (3%)	
- 4	2 (2%)	

# Clusterização – DBSCAN

## Resultados

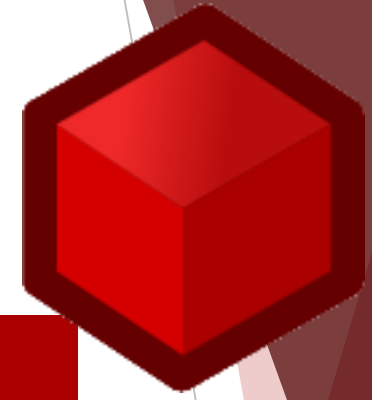


Configuração (raio, densidade) - cluster	Resultado da clusterização	Instâncias não clusterizadas
(2, 10) - 0	22 (100%)	104
(2, 6) - 0 - 1	23 (77%) 7 (23%)	96
(2, 3) - 0 - 1 - 2 - 3	7 (16%) 24 (53%) 11 (24%) 3 (7%)	81
(2, 2) - 0 - 1 - 2 - 3 - 4 - 5 - 6	7 (14%) 24 (24%) 11 (22%) 2 (4%) 2 (4%) 2 (4%) 3 (6%)	75



# Clusterização – DBSCAN

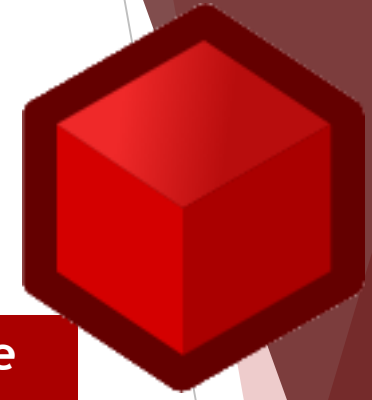
## Resultados



Configuração (raio, densidade) - cluster	Resultado da clusterização	Instâncias não clusterizadas
(2.37, 10) - 0 - 1	33 (49%) 35 (51%)	58
(2.37, 6) - 0 - 1	33 (44%) 42 (56%)	51
(2.37, 3) - 0 - 1	33 (39%) 51 (61%)	42
(2.37, 2) - 0 - 1 - 2 - 3 - 4	33 (37%) 51 (57%) 2 (2%) 2 (2%) 2 (2%)	36
(4, 20) - 0	126 (100%)	-

# Clusterização – DBSCAN

## Comparação entre parâmetros

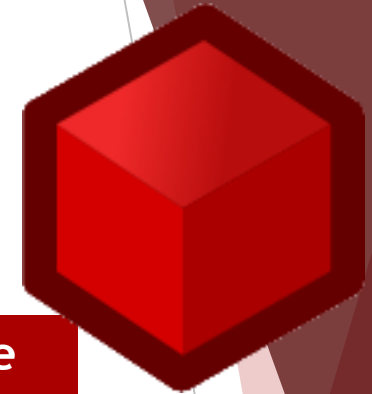


Raio \ Densidade	Relativamente pequena	Supostamente ideal	Relativamente grande
Relativamente pequeno	Não efetua clusterização (muitos 'não clusterizados')	     	Clusterização balanceada, 'não clusterizados' aumentam
Supostamente ideal	— — —	Melhor resultado	— — —
Relativamente grande	Clusterização desbalanceada, mais clusters, 'não clusterizados' diminuem	     	Um cluster contendo a maioria dos elementos

# Clusterização – DBSCAN

## Comparação entre parâmetros

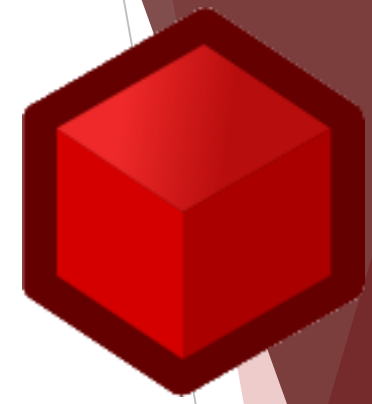
Resultado Bom.  
Mais provável



Raio \ Densidade	Relativamente pequena	Supostamente ideal	Relativamente grande
Relativamente pequeno	Não efetua clusterização (muitos 'não clusterizados')	<div>               </div>	Clusterização balanceada, 'não clusterizados' aumentam
Supostamente ideal	<div> —  — —  — — — </div>	<div> Melhor resultado  (Altamente improvável) </div>	<div> —  — —  — — — </div>
Relativamente grande	Clusterização desbalanceada, mais clusters, 'não clusterizados' diminuem	<div>               </div>	Um cluster contendo a maioria dos elementos

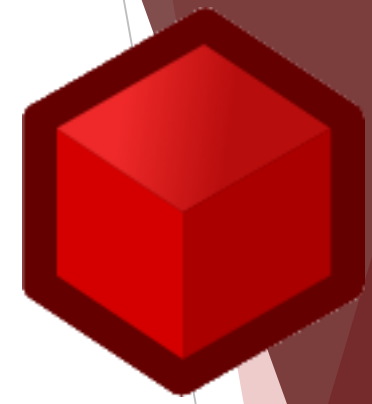
# Trabalhos Futuros

- ▶ Modificar o dataset para avaliar jogadores individualmente invés de partidas
- ▶ Avaliar as diferenças entre os toneios
- ▶ Avaliar a diferença entre as partidas femininas e masculinas
- ▶ Discretizar adequadamente a base das partidas de tênis para avaliar associações adequadas entre os atributos

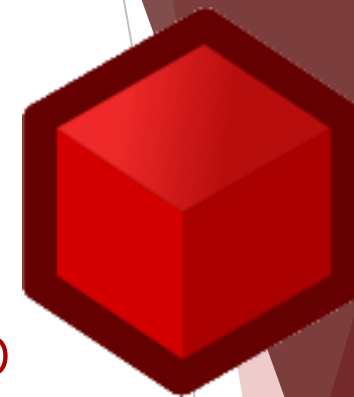


# Referências

- ▶ [1]<http://www.cs.waikato.ac.nz/ml/weka/>
- ▶ [2]<https://archive.ics.uci.edu/ml/datasets/Tennis+Major+Tournament+Match+Statistics>
- ▶ [3]<http://storm.cis.fordham.edu/~gweiss/data-mining/datasets.html>



Esta apresentação possui um documento associado

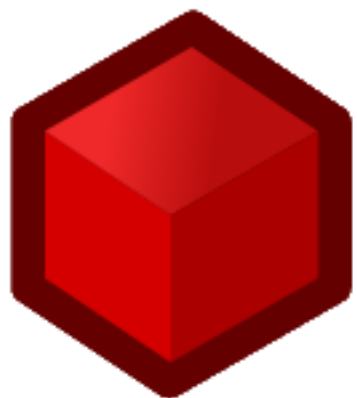


Veja mais em: [github.com/ClubedeCienciaDadosUFF/CCDrepository](https://github.com/ClubedeCienciaDadosUFF/CCDrepository)

"Believe you can and you're halfway there." (Theodore Roosevelt)

Dúvidas?  
Sugestões?

Marcelo d'Almeida  
[md@id.uff.br](mailto:md@id.uff.br)



CCD-UFF  
Clube de Ciência de Dados

09/06/2015

