

MO826A/MC936A/MC931S
Tópicos Especiais em Sistemas de Informação
eHealth - Informática e Saúde
Professor: André Santanchè

Lucas Ismailly Bezerra Freitas
Marcos Felipe de Menezes Mota

Universidade Estadual de Campinas
Instituto de Computação

20 de Junho de 2018

Agenda

Introdução

Definição do problema

Técnica e Ferramenta escolhida

Base de dados

Resultados

Agenda

Introdução

Definição do problema

Técnica e Ferramenta escolhida

Base de dados

Resultados

Introdução

Aterosclerose

- ▶ Aterosclerose é uma doença vascular crônica e progressiva que normalmente manifesta-se na idade adulta ou idade avançada

Introdução

Aterosclerose

- ▶ Aterosclerose é uma doença vascular crônica e progressiva que normalmente manifesta-se na idade adulta ou idade avançada
- ▶ Muito comum, mais de 2 milhões casos por ano no Brasil

Introdução

Aterosclerose

- ▶ Aterosclerose é uma doença vascular crônica e progressiva que normalmente manifesta-se na idade adulta ou idade avançada
- ▶ Muito comum, mais de 2 milhões casos por ano no Brasil
- ▶ O tratamento pode ajudar, mas essa doença não tem cura

Introdução

Aterosclerose

- ▶ Aterosclerose é uma doença vascular crônica e progressiva que normalmente manifesta-se na idade adulta ou idade avançada
- ▶ Muito comum, mais de 2 milhões casos por ano no Brasil
- ▶ O tratamento pode ajudar, mas essa doença não tem cura
- ▶ Requer um diagnóstico médico

Introdução

Aterosclerose

- ▶ Aterosclerose é uma doença vascular crônica e progressiva que normalmente manifesta-se na idade adulta ou idade avançada
- ▶ Muito comum, mais de 2 milhões casos por ano no Brasil
- ▶ O tratamento pode ajudar, mas essa doença não tem cura
- ▶ Requer um diagnóstico médico
- ▶ Crônico: pode durar anos ou a vida inteira

Introdução

Aterosclerose



Agenda

Introdução

Definição do problema

Técnica e Ferramenta escolhida

Base de dados

Resultados

Agenda

Introdução

Definição do problema

Técnica e Ferramenta escolhida

Base de dados

Resultados

Problema

Sintomas da aterosclerose

Problema

Sintomas da aterosclerose

- ▶ Dilatações de algumas áreas dos vasos sanguíneos (aneurismas)

Problema

Sintomas da aterosclerose

- ▶ Dilatações de algumas áreas dos vasos sanguíneos (aneurismas)
- ▶ Dor no peito tipo facada (angina ou infarto)

Problema

Sintomas da aterosclerose

- ▶ Dilatações de algumas áreas dos vasos sanguíneos (aneurismas)
- ▶ Dor no peito tipo facada (angina ou infarto)
- ▶ Dores fortes na cabeça (derrame cerebral)

Problema

Sintomas da aterosclerose

- ▶ Dilatações de algumas áreas dos vasos sanguíneos (aneurismas)
- ▶ Dor no peito tipo facada (angina ou infarto)
- ▶ Dores fortes na cabeça (derrame cerebral)
- ▶ Dores em braços e pernas (trombose)

Problema

Sintomas da aterosclerose

- ▶ Dilatações de algumas áreas dos vasos sanguíneos (aneurismas)
- ▶ Dor no peito tipo facada (angina ou infarto)
- ▶ Dores fortes na cabeça (derrame cerebral)
- ▶ Dores em braços e pernas (trombose)
- ▶ Dores no corpo

Problema

Sintomas da aterosclerose

- ▶ Dilatações de algumas áreas dos vasos sanguíneos (aneurismas)
- ▶ Dor no peito tipo facada (angina ou infarto)
- ▶ Dores fortes na cabeça (derrame cerebral)
- ▶ Dores em braços e pernas (trombose)
- ▶ Dores no corpo
- ▶ Cansaço

Problema

Sintomas da aterosclerose

- ▶ Dilatações de algumas áreas dos vasos sanguíneos (aneurismas)
- ▶ Dor no peito tipo facada (angina ou infarto)
- ▶ Dores fortes na cabeça (derrame cerebral)
- ▶ Dores em braços e pernas (trombose)
- ▶ Dores no corpo
- ▶ Cansaço

Exame Angio Tomografia de Coronárias

Problema

Sintomas da aterosclerose

- ▶ Dilatações de algumas áreas dos vasos sanguíneos (aneurismas)
- ▶ Dor no peito tipo facada (angina ou infarto)
- ▶ Dores fortes na cabeça (derrame cerebral)
- ▶ Dores em braços e pernas (trombose)
- ▶ Dores no corpo
- ▶ Cansaço

Exame Angio Tomografia de Coronárias

- ▶ Custa em média 4,500 reais

Problema

Sintomas da aterosclerose

- ▶ Dilatações de algumas áreas dos vasos sanguíneos (aneurismas)
- ▶ Dor no peito tipo facada (angina ou infarto)
- ▶ Dores fortes na cabeça (derrame cerebral)
- ▶ Dores em braços e pernas (trombose)
- ▶ Dores no corpo
- ▶ Cansaço

Exame Angio Tomografia de Coronárias

- ▶ Custa em média 4,500 reais
- ▶ Então, a ideia é tentar prever a necessidade de realizar ou não o exame

Agenda

Introdução

Definição do problema

Técnica e Ferramenta escolhida

Base de dados

Resultados

Agenda

Introdução

Definição do problema

Técnica e Ferramenta escolhida

Base de dados

Resultados

Técnica e Ferramenta

Escolher a que tem o melhor resultado entre:

Técnica e Ferramenta

Escolher a que tem o melhor resultado entre:

- ▶ Redes Bayesiana

Técnica e Ferramenta

Escolher a que tem o melhor resultado entre:

- ▶ Redes Bayesiana
- ▶ Floresta de decisão

Técnica e Ferramenta

Escolher a que tem o melhor resultado entre:

- ▶ Redes Bayesiana
- ▶ Floresta de decisão
- ▶ Árvore de decisão

Técnica e Ferramenta

Escolher a que tem o melhor resultado entre:

- ▶ Redes Bayesiana
- ▶ Floresta de decisão
- ▶ Árvore de decisão

Ferramenta

- ▶ Weka

Agenda

Introdução

Definição do problema

Técnica e Ferramenta escolhida

Base de dados

Resultados

Agenda

Introdução

Definição do problema

Técnica e Ferramenta escolhida

Base de dados

Resultados

Base de datos

Base de dados

Base - aterosclerose

- ▶ Base com 14 atributos

Base de dados

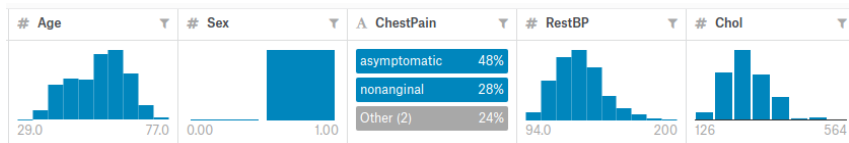
Base - aterosclerose

- ▶ Base com 14 atributos
- ▶ 303 instâncias

Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
63	1	typical	145	233	1	2	150	0	2.3	3	0	fixed	No
67	1	asymptomatic	160	286	0	2	108	1	1.5	2	3	normal	Yes
67	1	asymptomatic	120	229	0	2	129	1	2.6	2	2	reversible	Yes
37	1	nonanginal	130	250	0	0	187	0	3.5	3	0	normal	No
41	0	nontypical	130	204	0	2	172	0	1.4	1	0	normal	No
56	1	nontypical	120	236	0	0	178	0	0.8	1	0	normal	No
62	0	asymptomatic	140	268	0	2	160	0	3.6	3	2	normal	Yes
57	0	asymptomatic	120	354	0	0	163	1	0.6	1	0	normal	No
63	1	asymptomatic	130	254	0	2	147	0	1.4	2	1	reversible	Yes
53	1	asymptomatic	140	203	1	2	155	1	3.1	3	0	reversible	Yes
57	1	asymptomatic	140	192	0	0	148	0	0.4	2	0	fixed	No
56	0	nontypical	140	294	0	2	153	0	1.3	2	0	normal	No
56	1	nonanginal	130	256	1	2	142	1	0.6	2	1	fixed	Yes
44	1	nontypical	120	263	0	0	173	0	0	1	0	reversible	No
52	1	nonanginal	172	199	1	0	162	0	0.5	1	0	reversible	No
57	1	nonanginal	150	168	0	0	174	0	1.6	1	0	normal	No
48	1	nontypical	110	229	0	0	168	0	1	3	0	reversible	Yes
54	1	asymptomatic	140	239	0	0	160	0	1.2	1	0	normal	No

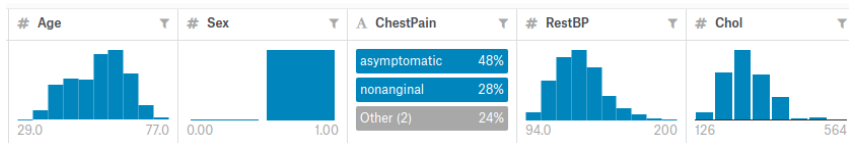
Figura: Dados de dados.

Base - atherosclerosis

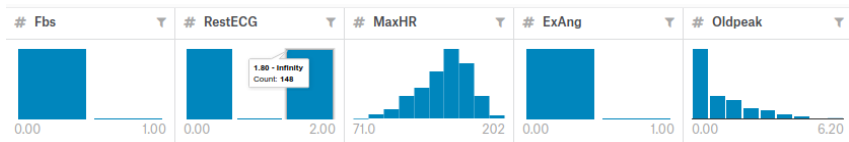


(a)

Base - atherosclerosis

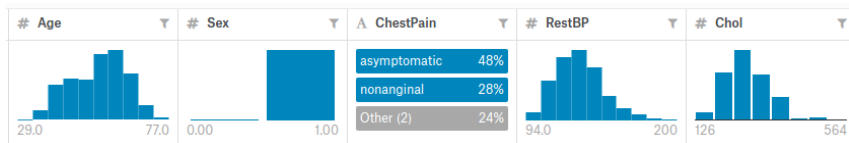


(a)

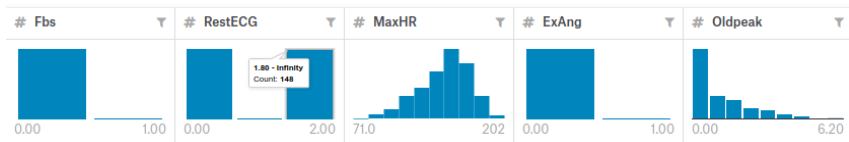


(b)

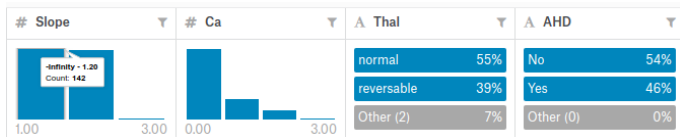
Base - atherosclerosis



(a)



(b)



(c)

Agenda

Introdução

Definição do problema

Técnica e Ferramenta escolhida

Base de dados

Resultados

Agenda

Introdução

Definição do problema

Técnica e Ferramenta escolhida

Base de dados

Resultados

Resultados

Árvore de decisão

Size of the tree : 171

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	229	75.5776 %
Incorrectly Classified Instances	74	24.4224 %
Kappa statistic	0.5098	
Mean absolute error	0.2462	
Root mean squared error	0.4947	
Relative absolute error	49.5639 %	
Root relative squared error	99.2677 %	
Total Number of Instances	303	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.756	0.245	0.785	0.756	0.77	0.753	No
	0.755	0.244	0.724	0.755	0.739	0.753	Yes
Weighted Avg.	0.756	0.244	0.757	0.756	0.756	0.753	

=== Confusion Matrix ===

a	b	<-- classified as
124	40	a = No
34	105	b = Yes

Resultados

Floresta de decisão

Random forest of 100 trees, each constructed while considering 4 random features.
Out of bag error: 0.2013

Time taken to build model: 0.32 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	246	81.1881 %
Incorrectly Classified Instances	57	18.8119 %
Kappa statistic	0.6201	
Mean absolute error	0.2688	
Root mean squared error	0.356	
Relative absolute error	54.1256 %	
Root relative squared error	71.4319 %	
Total Number of Instances	303	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.841	0.223	0.817	0.841	0.829	0.902	No
	0.777	0.159	0.806	0.777	0.791	0.902	Yes
Weighted Avg.	0.812	0.193	0.812	0.812	0.812	0.902	

=== Confusion Matrix ===

a	b	<-- classified as
138	26	a = No
31	108	b = Yes

Resultados

Rede Bayesiana

```
=== Stratified cross-validation ===  
=== Summary ===
```

Correctly Classified Instances	251	82.8383 %
Incorrectly Classified Instances	52	17.1617 %
Kappa statistic	0.6537	
Mean absolute error	0.1924	
Root mean squared error	0.365	
Relative absolute error	38.7314 %	
Root relative squared error	73.2407 %	
Total Number of Instances	303	

```
=== Detailed Accuracy By Class ===
```

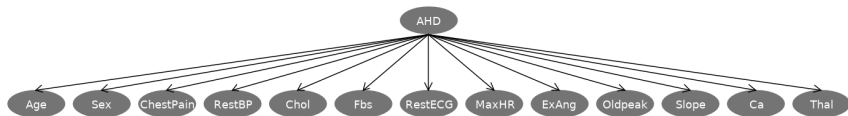
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.854	0.201	0.833	0.854	0.843	0.906	No
	0.799	0.146	0.822	0.799	0.81	0.906	Yes
Weighted Avg.	0.828	0.176	0.828	0.828	0.828	0.906	

```
=== Confusion Matrix ===
```

a	b	<-- classified as
140	24	a = No
28	111	b = Yes

Resultados

Rede Bayesiana - Grafo



Resultados

Escolher a que tem o melhor resultado entre:

- ▶ Floresta de decisão ✗
- ▶ Árvore de decisão ✗
- ▶ Redes Bayesiana ✓

Resultados

Escolher a que tem o melhor resultado entre:

- ▶ Floresta de decisão ✗
- ▶ Árvore de decisão ✗
- ▶ Redes Bayesiana ✓

Atenção

- ▶ Para esse conjunto de dados
- ▶ Cada contexto impõe restrições e dados diferentes

Resultados - Melhorando Redes Bayesianas

Rede Bayesiana - TAN

```
=== Stratified cross-validation ===  
=== Summary ===
```

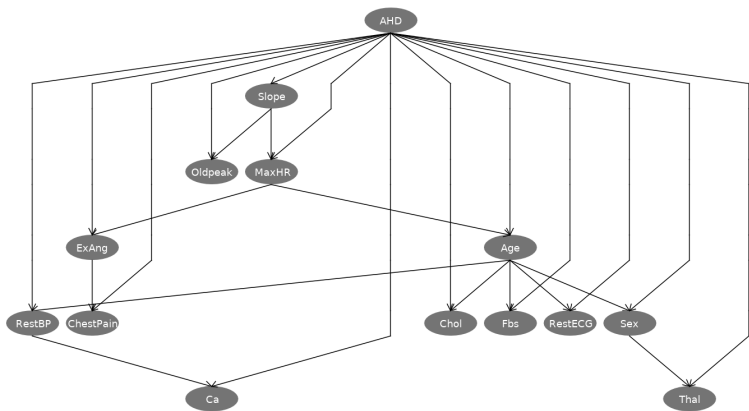
Correctly Classified Instances	254	83.8284 %
Incorrectly Classified Instances	49	16.1716 %
Kappa statistic	0.6724	
Mean absolute error	0.2049	
Root mean squared error	0.348	
Relative absolute error	41.262 %	
Root relative squared error	69.8326 %	
Total Number of Instances	303	

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.884	0.216	0.829	0.884	0.855	0.912	No
	0.784	0.116	0.852	0.784	0.816	0.912	Yes
Weighted Avg.	0.838	0.17	0.839	0.838	0.838	0.912	

Resultados - Melhorando Redes Bayesianas

Rede Bayesiana - TAN



Resultados - Melhorando Redes Bayesianas

Rede Bayesiana - Inductive Causation

```
=== Stratified cross-validation ===  
=== Summary ===
```

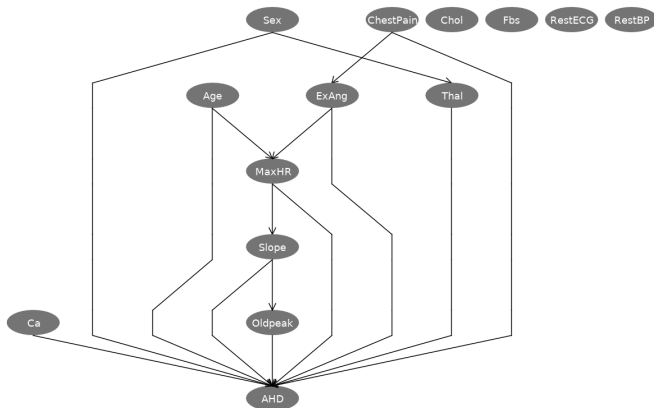
Correctly Classified Instances	221	72.9373 %
Incorrectly Classified Instances	82	27.0627 %
Kappa statistic	0.4403	
Mean absolute error	0.3449	
Root mean squared error	0.4016	
Relative absolute error	69.4378 %	
Root relative squared error	80.5881 %	
Total Number of Instances	303	

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.896	0.468	0.693	0.896	0.782	0.848	No
	0.532	0.104	0.813	0.532	0.643	0.848	Yes
Weighted Avg.	0.729	0.301	0.748	0.729	0.718	0.848	

Resultados - Melhorando Redes Bayesianas

Rede Bayesiana - Inductive Causation



Dúvidas?

