Naive Bayes

Introdução

- Abordagem probabilística (Teorema de Bayes)
- Exemplos
 - Filtros de spam
 - Mineração de emoções
 - Separação de documentos

Base original

Baixa

Baixa

Baixa

Alta

Alta

Alta

Alta

Alta

Ruim

Ruim

Boa

Boa

Boa

Boa

Boa

Ruim

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

< 15.000

> 35.000

> 35.000

> 35.000

< 15.000

> 35.0000

>= 15.000 a <= 35.000

>= 15.000 a <= 35.000

Nenhuma

Adequada

Nenhuma

Adequada

Nenhuma

Nenhuma

Nenhuma

Nenhuma

Alto

Baixo

Baixo

Baixo

Alto

Alto

Moderado

Moderado

Naive Bayes

Moderado

3/14

Baixo

5/14

1/3

3/5

1/3

2/5

1/3

0

1/3

2/5

	His	stória d	lo cré	ditoDí	vida	Gar	antias	Renda anual			
Risco o crédito		Desconhecio 5	da Ruim 4	Alta 7	Baixa 7	Nenhuma 11	a Adequada 3	a < 1500 3	>= 150 <= 350 4	00 00 3500 7	
Alto	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6	

6/14

2/3 2/3

3/5

3/5

1/3

2/5

2/3

1/3

5/5

	Н	istória do	crédito	o Dí	vida	Gar	antias	Renda anual			
Risco d crédito	DOG	Desconhecid 5	a Ruim 4								
Alto 6/14	1/6	2/6	3/6								
Moderad 3/14	1/3	1/3	1/3								
Baixo 5/14	3/5	2/5	0								

História do créo	iRi sco
Ruim	Alto
Desconhecida	Alto
Desconhecida	Moderado
Desconhecida	Alto
Desconhecida	Baixo
Desconhecida	Baixo
Ruim	Alto
Ruim	Moderado
Boa	Baixo
Boa	Baixo
Boa	Alto
Boa	Moderado
Boa	Baixo
Ruim	Alto

	História do crédito			Dí	vida	Gar	antias	Renda anual			
Risco d crédito	Doa	Desconhecid	la Ruim 4	Alta 7	Baixa 7						
Alto 6/14	1/6	2/6	3/6	4/6	2/6						
Moderac 3/14	1/3	1/3	1/3	1/3	2/3						
Baixo 5/14	3/5	2/5	0	2/5	3/5						

Dívida Alta Alta Baixa Baixa Baixa Baixa Baixa Baixa Baixa Alta Alta Alta Alta Alta

Risco

Alto

Alto

Alto

Baixo

Baixo

Baixo

Baixo

Baixo

Alto

Alto

Moderado

Moderado

Alto

Moderado

	História do crédit		crédit	o Dívida		Gar	Garantias		Renda anual		Garantias	Risco
											Nenhuma	Alto
Risco d	a -			A.11	Б.						Nenhuma	Alto
crédito	Boa 5	Desconhecio 5	la Ruim 4	Alta 7	Baixa 7	Nenhum 11	aAdequada 3	a			Nenhuma	Moderado
			·	ŕ					Nenhuma	Alto		
Alto	1 /6	276	2/6	A 1C	216	C 1C	0				Nenhuma	Baixo
6/14	1/6	2/6	3/6	4/6	2/6	6/6	0				Adequada	Baixo
0, = .											Nenhuma	Alto
Moderad	1/3	1/3	1/3	1/3	2/3	2/3	1/3				Adequada	Moderado
3/14	1/3	1/3	1/3	1/3	2/3	2/3	1,3				Nenhuma	Baixo
											Adequada	Baixo
Baixo	3/5	2/5	0	2/5	3/5	3/5	2/5				Nenhuma	Alto
5/14											Nenhuma	Moderado
											Nenhuma	Baixo
											Nenhuma	Alto

		História do crédito			vida	Gar —	antias	Renda anual			
Risco d crédito	e Boa 5	Desconhecid	la Ruim 4	Alta 7	Baixa 7	Nenhuma	aAdequad 3	a < 15 3	>= 15 <= 35 4		
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6	
Moderad 3/14	°1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3	
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5	

Renda anual < 15.000 >= 15.000 a <= 35.0<mark>0Alto</mark> >= 15.000 a <= 35.0<mark>0Moderad</mark> > 35.000 35.000 > 35.000 < 15.000 > 35.000 > 35.000 > 35.000 < 15.000 >= 15.000 a <= 35.0<mark>0Moderad</mark> 35.0000 >= 15.000 a <= 35.0<mark>0Alto</mark>

Risco

Alto

Alto

Baixo

Baixo

Alto

Moderado

Baixo

Baixo

Alto

Baixo

	Hi	istória do	crédito	DÍN	Dívida Gar		antias	Renda anual			
Risco c crédit	^e Boa 5	Desconhecida 5	a Ruim 4	Alta 7	Baixa 7	Nenhum 11	Adequada 3	e < 15 3	>= 15 <= 35 4	<u> </u>	História = Boa Dívida = Alta Garantias = Nenhu Renda = > 35
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6	Soma: 0,0079 + 0,0052 - 0,0514 = 0,0645
Modera 3/14	°1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3	
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5	
P(Alto)	= 0.0	079		P(N	/lodera	do) = 0	,0052			P(Baixo	b) = 5/14 * 3/5 * 2/5 * 3/5 * b) = 0,0514 b% P(Baixo) = 0,0514 / 0,06

Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6	Correção Laplacia
Moderad 3/14	°1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3	
Baixo 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5	
P(Alto)	= 6/1	4 * 3/6 *	4/6 * 0	* 3/6P(N	Modera	do) = 3,	/14 * 1/3	3 * 1/3	* 1/3 *	P(Baixo	b) = 5/14 * 0 * 2/5 * 2/5 * 0

Garantias

Baixa Nenhum Adequad < 15

11

3

Renda anual

>= 15 <= 35 4 7

História = Ruim

Garantias = Adequ

Dívida = Alta

Renda = < 15

História do crédito

Desconhecica Ruim

4

5

Risco d

crédito

Boa

5

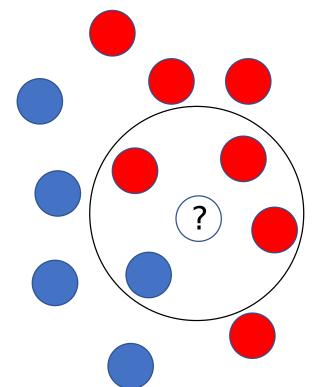
Dívida

Alta

P(vermelho) = 7 / 12

Probabilidades apriori

P(azul) = 5 / 12



P'(vermelho) = 3 / 7

P'(azul) = 1 / 5

Probabilidades posteori

P''(vermelho) = 7 / 12 * 3 / 7 = 21 / 84 =**0,25**

P''(azul) = 5 / 12 * 1 / 5 = 5 / 60 =**0,08**

Vantagens x desvantagens

- Vantagens
 - Rápido
 - Simplicidade de interpretação
 - Trabalha com altas dimensões
 - Boas previsões em bases pequenas
- Desvantagem
 - Combinação de características (atributos independentes) cada par características são independentes – nem sempre é verdade

Conclusão