Naive Bayes

Jones Granatyr



Introdução

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

Baixa

Baixa

Baixa

Baixa

Baixa

Alta

Alta

Alta

Alta

Alta

Desconhecida

Desconhecida

Ruim

Ruim

Boa

Boa

Boa

Boa

Boa

Ruim

Base original											
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							

> 35.000

> 35.000

< 15.000

> 35.000

> 35.000

> 35.000

< 15.000

> 35.0000

>= 15.000 a <= 35.000

>= 15.000 a <= 35.000

Baixo

Baixo

Alto

Baixo

Baixo

Alto

Baixo

Alto

Moderado

Moderado

Nenhuma

Adequada

Nenhuma

Adequada

Nenhuma

Adequada

Nenhuma

Nenhuma

Nenhuma

Nenhuma

Maive Raves

1/6

1/3

3/5

2/6

1/3

2/5

3/6

1/3

0

Alto

6/14

Moderado

3/14

Baixo

5/14

IVaiv		iyes								
	Hist	ória do cr	rédito	Dív	ida	Gara	ntias	Re	nda anı	ual
Risco de crédito	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	< 15000 3	>= 15000 <= 35000	> 35

2/6

3/5

1/3 2/3

4/6

2/5

6/6

2/3

3/5

0

1/3

2/5

3/6

0

2/6

2/3

0

> 35000

1/6

1/3

5/5

	História do crédito		Dívida		Garantias		R	al		
Risco de crédito	Boa 5	Desconhecida 5	Ruim 4							
Alto 6/14	1/6	2/6	3/6							
Moderado 3/14	1/3	1/3	1/3							
_{Ваіхо} 5/14	3/5	2/5	0							

História do crédito
Ruim
Desconhecida
Ruim
Ruim
Boa
Воа
Boa
Boa
Воа
Ruim

Alto

Alto

Alto

Baixo

Baixo

Alto

Baixo

Baixo

Alto

Baixo

Alto

Moderado

Moderado

Moderado

Risco de crédito 5 Desconhecida Ruim Alta 7 7 Alto 1/6 2/6 3/6 4/6 2/6 6/14 1/3 1/3 1/3 1/3 2/3 Baixo 5/14 3/5 2/5 0 2/5 3/5	História do crédito			Dív	rida	Gara	ntias	Renda anual			
6/14 Moderado 3/14 1/3 1/3 1/3 1/3 1/3 2/3 8aixo 3/5 2/5 0 2/5 3/5											
3/14 1/3 1/3 1/3 2/3 Baixo 3/5 2/5 0 2/5 3/5	1/6	2/6	3/6	4/6	2/6						
3/5 2/5 0 2/5 3/5	1/3	1/3	1/3	1/3	2/3						
	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

Alto

Baixo

Baixo

Alto

Baixo

Alto

Moderado

Moderado

Moderado

	Hi	stória do cré	dito	Dív	vida	Gara	ntias	Renda anual		Renda anual		Renda anual		Risco
											Nenhuma	Alto		
Risco de											Nenhuma	Alto		
crédito	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3				Nenhuma	Moderado		
			·								Nenhuma	Alto		
Alto	4 /6	2/5	0.45	1/6	2/5	6.16					Nenhuma	Baixo		
6/14	1/6	2/6	3/6	4/6	2/6	6/6	0				Adequada	Baixo		
3, 2 :											Nenhuma	Alto		
Moderado	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 Dívida Garantias Renda anual		istória do crédito Dívida		stória do crédito Dívida Garantias Renda anual					História do crédito		Dívida		Garantias Renda anual			Renda anual		Renda anual	Risco
												< 15.000	Alto							
									>= 15			>= 15.000 a <= 35.000	Alto							
Risco de crédito	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma	Adequada 3	< 15 3	<= 35 4	> 35 7		>= 15.000 a <= 35.000	Moderado							
		J	·					Ū				> 35.000	Alto							
Alto	1 10	2/2	2.45	1/6	2/6	6.16		2/5	2.46	1 /6		> 35.000	Baixo							
6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6		> 35.000	Baixo							
37 = 1												< 15.000	Alto							
Moderado	1 /2	1 /2	1 /2	1/2	2/2	2/2	1 /2	0	2/2	1 /2		> 35.000	Moderado							
3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3		> 35.000	Baixo							
												> 35.000	Baixo							
Baixo	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5		< 15.000	Alto							
5/14												>= 15.000 a <= 35.000	Moderado							
												> 35.0000	Baixo							
												>= 15.000 a <= 35.000	Alto							

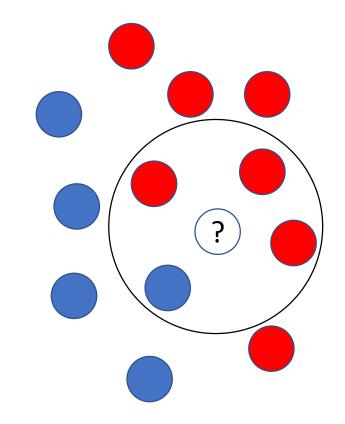
	História do crédito			al								
Risco de crédito	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	< 15 3	>= 15 <= 35 4	> 35 7	História = Boa Dívida = Alta Garantias = Nenhuma Renda = > 35	
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6	Soma: 0,0079 + 0,0052 + 0,0514 = 0,0645	
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3		
_{Ваіхо} 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5		
P(Alto) =	P(Alto) = 6/14 * 1/6 * 4/6 * 6/6 * 1/6 P(Moderado) = 3/14 * 1/3 * 1/3 * 2/3 * 1/3 P(Baixo) = 5/14 * 3/5 * 2/5 * 3/5 * 5/5 P(Alto) = 0,0079 P(Moderado) = 0,0052 P(Baixo) = 0,0514 P(Alto) = 0,0079 / 0,0645 * 100 = 12,24% P(Moderado) = 0,0052 / 0,0645 * 100 = 8,06% P(Baixo) = 0,0514 / 0,0645 * 100 = 79,68%											

	Hi	istória do cré										
Risco de crédito	Boa 5	Desconhecida 5	Ruim 4	Alta 7	Baixa 7	Nenhuma 11	Adequada 3	< 15 3	>= 15 <= 35 4	> 35 7	História = Ruim Dívida = Alta Garantias = Adequada Renda = < 15	
Alto 6/14	1/6	2/6	3/6	4/6	2/6	6/6	0	3/6	2/6	1/6	Correção Laplaciana	
Moderado 3/14	1/3	1/3	1/3	1/3	2/3	2/3	1/3	0	2/3	1/3		
_{Ваіхо} 5/14	3/5	2/5	0	2/5	3/5	3/5	2/5	0	0	5/5		
P(Alto) = 6/14 * 3/6 * 4/6 * 0 * 3/6 P(Moderado) = 3/14 * 1/3 * 1/3 * 1/3 * 0 P(Baixo) = 5/14 * 0 * 2/5 * 2/5 * 0												

$$P(vermelho) = 7 / 12$$

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 de características são independentes – nem sempre é verdade

Conclusão