

Aprendizagem por regras

Jones Granatyr



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
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

Regras

Regra	Resultado
Se renda = >35.000 E história de crédito = BOA	Risco = BAIXO
Se renda = >35.000 e história de crédito = DESCONHECIDA	Risco = BAIXO
Default (padrão)	Risco = ALTO

História = Boa

Dívida = Alta

Garantias = Nenhuma

Renda = > 35

História = Ruim

Dívida = Alta

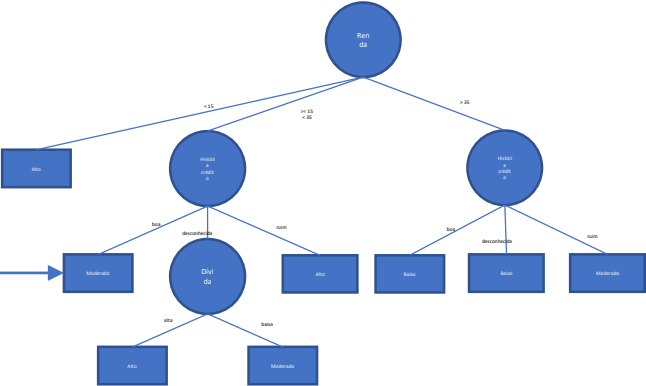
Garantias = Adequada

Renda = < 15



Risco de crédito	História do crédito			Dívida		Garantias		Renda anual		
	Boa	Desconhecida	Ruim	Alta	Baixa	Imóvel	Aluguel	< 15000	15000 - 35000	> 35000
Alta	1/5	2/5	3/4	4/7	2/7	6/11	0	3/3	2/4	1/7
Moderada	1/5	1/5	1/4	1/7	2/7	2/11	1/3	0	2/4	1/7
Baixa	3/5	2/5	0	2/7	3/7	3/11	2/3	0	0	5/7

Naive bayes



Árvore de decisão

Regra	Resultado
Se renda = >35.000 E história_crédito = BOA	Risco = BAIXO
Se renda = >35.000 e história_crédito = DESCONHECIDA	Risco = BAIXO
Default (padrão)	Risco = ALTO

Regras

Registros
% acerto



Algoritmo OneR

- Testar coisas simples primeiro
- Um atributo faz todo o trabalho

Atributo	Quantidade	Regras	Erro	Total erro
História de crédito	Boa 5	SE Boa ENTÃO Baixo	2/5	6/14
	Desconhecida 5	SE Desconhecida ENTÃO Alto	3/5	
	Ruim 4	SE Ruim ENTÃO Alto	1/4	
Dívida	Alta 7	SE Alta ENTÃO Alto	3/7	7/14
	Baixa 7	SE Baixa ENTÃO Baixo	4/7	
Garantias	Nenhuma 11	SE Nenhuma ENTÃO Alto	5/11	6/14
	Adequada 3	SE Adequada ENTÃO Baixo	1/3	
Renda	< 15000 3	SE <15000 ENTÃO Alto	0	4/14
	>= 15000 <= 35000 4	SE >=15000 a <= 35000 ENTÃO Alto	2/4	
	> 35000 7	SE >35000 ENTÃO Baixo	2/7	

Atributo	Quantidade	Regras	Erro	Total erro
História de crédito	Boa	5	SE Boa ENTÃO Baixo	2/5
	Desconhecida	5	SE Desconhecida ENTÃO Alto	3/5
	Ruim	4	SE Ruim ENTÃO Alto	1/4
				6/14

- SE Boa ENTÃO Alto = 1

SE Boa ENTÃO Moderado = 1

SE Boa ENTÃO Baixo = 3
- SE Ruim ENTÃO Alto = 3

SE Ruim ENTÃO Moderado = 1

SE Ruim ENTÃO Baixo = 0
- SE Desconhecida ENTÃO Alto = 2

SE Desconhecida ENTÃO Moderado = 1

SE Desconhecida ENTÃO Baixo = 2

História do crédito	Risco
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

Atributo	Quantidade	Regras	Erro	Total erro
História de crédito	Boa	5	SE Boa ENTÃO Baixo	2/5
	Desconhecida	5	SE Desconhecida ENTÃO Alto	3/5
	Ruim	4	SE Ruim ENTÃO Alto	1/4
Dívida	Alta	7	SE Alta ENTÃO Alto	3/7
	Baixa	7	SE Baixa ENTÃO Baixo	4/7

SE Alta ENTÃO Alto = 4
SE Alta ENTÃO Moderado = 1
SE Alta ENTÃO Baixo = 2

SE Baixa ENTÃO Alto = 2
SE Baixa ENTÃO Moderado = 2
SE Baixa ENTÃO Baixo = 3

Dívida	Risco
Alta	Alto
Alta	Alto
Baixa	Moderado
Baixa	Alto
Baixa	Baixo
Baixa	Baixo
Baixa	Alto
Baixa	Moderado
Baixa	Baixo
Alta	Baixo
Alta	Alto
Alta	Moderado
Alta	Baixo
Alta	Alto

Atributo	Quantidade	Regras	Erro	Total erro
História de crédito	Boa	5	SE Boa ENTÃO Baixo	2/5
	Desconhecida	5	SE Desconhecida ENTÃO Alto	3/5
	Ruim	4	SE Ruim ENTÃO Alto	1/4
Dívida	Alta	7	SE Alta ENTÃO Alto	3/7
	Baixa	7	SE Baixa ENTÃO Baixo	4/7
Garantias	Nenhuma	11	SE Nenhuma ENTÃO Alto	5/11
	Adequada	3	SE Adequada ENTÃO Baixo	1/3

SE Nenhuma ENTÃO Alto = 6

SE Nenhuma ENTÃO Moderado = 2

SE Nenhuma ENTÃO Baixo = 3

SE Adequada ENTÃO Alto = 0

SE Adequada ENTÃO Moderado = 1

SE Adequada ENTÃO Baixo = 2

Garantias	Risco
Nenhuma	Alto
Nenhuma	Alto
Nenhuma	Moderado
Nenhuma	Alto
Nenhuma	Baixo
Adequada	Baixo
Nenhuma	Alto
Adequada	Moderado
Nenhuma	Baixo
Adequada	Baixo
Nenhuma	Alto
Nenhuma	Moderado
Nenhuma	Baixo
Nenhuma	Alto

Atributo	Quantidade		Regras	Erro	Total erro
História de crédito	Boa	5	SE Boa ENTÃO Baixo	2/5	6/14
	Desconhecida	5	SE Desconhecida ENTÃO Alto	3/5	
	Ruim	4	SE Ruim ENTÃO Alto	1/4	
Dívida	Alta	7	SE Alta ENTÃO Alto	3/7	7/14
	Baixa	7	SE Baixa ENTÃO Baixo	4/7	
Garantias	Nenhuma	11	SE Nenhuma ENTÃO Alto	5/11	6/14
	Adequada	3	SE Adequada ENTÃO Baixo	1/3	
Renda	< 15000	3	SE < 15000 ENTÃO Alto	0	4/14
	>= 15000 < 35000	4	SE >= 15000 < 35000 ENTÃO Alto	2/4	
	> 35000	7	SE > 35000 ENTÃO Baixo	2/7	

SE < 15000 ENTÃO Alto = 3
SE < 15000 ENTÃO Moderado = 0
SE < 15000 ENTÃO Baixo = 0

SE > 35000 ENTÃO Alto = 1
SE > 35000 ENTÃO Moderado = 1
SE > 35000 ENTÃO Baixo = 5

SE >= 15000 < 35000 ENTÃO Alto = 2
SE >= 15000 < 35000 ENTÃO Moderado = 2
SE >= 15000 < 35000 ENTÃO Baixo = 0

Renda anual	Risco
< 15.000	Alto
>= 15.000 a <= 35.000	Alto
>= 15.000 a <= 35.000	Moderado
> 35.000	Alto
> 35.000	Baixo
> 35.000	Baixo
< 15.000	Alto
> 35.000	Moderado
> 35.000	Baixo
> 35.000	Baixo
< 15.000	Alto
>= 15.000 a <= 35.000	Moderado
> 35.0000	Baixo
>= 15.000 a <= 35.000	Alto

Algoritmo PRISM - Se ? Então Risco = Alto

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	> 35.000	Alto
Ruim	Baixa	Nenhuma	< 15.000	Alto
Boa	Alta	Nenhuma	< 15.000	Alto
Ruim	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto

Se ? Então Risco = Alto

Se Garantias = Nenhuma Então Risco = Alto

Atributo/valor	Abrangência
História de crédito = Boa	1/5
História de crédito = Desconhecida	2/5
História de crédito = Ruim	3/4
Dívida = Alta	4/7
Dívida = Baixa	2/7
Garantias = Nenhuma	6/11
Garantias = Adequada	0/3
Renda = <15	3/3
Renda = 15-35	2/4
Renda = >35	1/7

Se Garantias = Nenhuma E ? Então Risco = Alto

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	> 35.000	Alto
Ruim	Baixa	Nenhuma	< 15.000	Alto
Boa	Alta	Nenhuma	< 15.000	Alto
Ruim	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto

Se Garantias = Nenhuma E ? Então Risco = Alto

Se Garantias = Nenhuma E Dívida = Alta Então Risco = Alto

Atributo/valor	Abrangência
História de crédito = Boa	1/4
História de crédito = Desconhecida	2/4
História de crédito = Ruim	3/3
Dívida = Alta	4/6
Dívida = Baixa	2/5
Renda = <15	3/3
Renda = 15-35	2/4
Renda = >35	1/4

Se Garantias = Nenhuma E Dívida = Alta E ?
Então Risco = Alto

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

Se Garantias = Nenhuma E Dívida = Alta E ? Então Risco = Alto

Se Garantias = Nenhuma E Dívida = Alta E História = Ruim Então Risco = Alto

Atributo/valor	Abrangência
História de crédito = Boa	1/3
História de crédito = Desconhecida	1/1
História de crédito = Ruim	2/2
Renda = <15	2/2
Renda = 15-35	2/3
Renda = >35	0/1

Se Garantias = Nenhuma E Dívida = Alta E História
= Ruim E ? Então Risco = Alto

História do crédito	Dívida	Garantias	Renda anual	Risco
Ruim	Alta	Nenhuma	< 15.000	Alto
Ruim	Alta	Nenhuma	>= 15.000 a <= 35.000	Alto

Se Garantias = Nenhuma E Dívida = Alta E História = Ruim
E ? Renda <15 Então Risco = Alto

Se Garantias = Nenhuma E Dívida = Alta E História = Ruim
E Renda < 15 Então Risco = Alto

Atributo/valor	Abrangência
Renda = <15	1/1
Renda = 15-35	1/1
Renda = >35	0

Árvores x Regras

- Compreensibilidade
- Pouco espaço de armazenamento
- Mais lento (regras)
- Em geral, regras não apresentam melhores resultados do que árvores de decisão

Conclusão

