



Árvore de decisão

usaremos esta base para testes

Histórico	Divida	Garantias	Renda	Risco
ruim	alta	nenhuma	0_15	alto
desconhecida	alta	nenhuma	15_35	alto
desconhecida	baixa	nenhuma	15_35	moderado
desconhecida	baixa	nenhuma	acima_35	alto
desconhecida	baixa	nenhuma	acima_35	baixo
desconhecida	baixa	adequada	acima_35	baixo
ruim	baixa	nenhuma	0_15	alto
ruim	baixa	adequada	acima_35	moderado
Boa	baixa	nenhuma	acima_35	baixo
Boa	alta	adequada	acima_35	baixo
Boa	alta	nenhuma	0_15	alto
Boa	alta	nenhuma	15_35	moderado
Boa	alta	nenhuma	acima_35	baixo
ruim	alta	nenhuma	15_35	alto

Como o algoritmo constrói a árvore?

$$Entropy(s) = \sum_{i=1}^c -P_i \log_2 (P_i)$$

$$Gain(S, A) = Entropy(s) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

1. Definir as possibilidades das classes

Risco
alto
alto
moderado
alto
baixo
baixo
alto
moderado
baixo
baixo
alto
moderado
baixo
alto

$$Alto = \frac{6}{14}$$

$$Moderado = \frac{3}{14}$$

$$Baixo = \frac{5}{14}$$

Necessita-se agora o cálculo da entropia, que basicamente vai **medir o quão organizado e desorganizado** os dados estão na base de dados.

$$Entropy(S) = -\frac{6}{14} * \log_2 \left(\frac{6}{14} \right) - \frac{3}{14} * \log_2 \left(\frac{3}{14} \right) - \frac{5}{14} * \log_2 \left(\frac{5}{14} \right)$$

$$Entropy(s) = 1,53$$

Diante da entropia, precisamos definir agora a posição das colunas na hierarquia da árvore, para definir o atributo que ficará na raiz da árvore, vamos aplicar:

$$Gain(Historia)$$

$$Gain(renda)$$

$$Gain(Garantia)$$

$$Gain(Divida)$$

vamos iniciar com o atributo história de crédito

Histórico	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

$$Boa = \frac{5}{14}$$

$$Desconhecida = \frac{5}{14}$$

$$Ruim = \frac{4}{14}$$

História de credito	Risco (Alto)	Risco (Moderado)	Risco (Baixo)
Boa = 5/14	1/5	1/5	3/5
Desconhecida = 5/14	2/5	1/5	2/5
Ruim = 4/14	3/4	1/5	0

Vamos agora calcular a entropia do risco para o atributo escolhido

- Para o histórico(Boa)

$$Entropy(Alto) = -\frac{1}{5} * \log_2 \left(\frac{1}{5} \right) - \frac{1}{5} * \log_2 \left(\frac{1}{5} \right) - \frac{3}{5} * \log_2 \left(\frac{4}{5} \right)$$

$$Entropy(Alto) = 1,37$$

- Para o Histórico(Desconhecida)

$$Entropy(Moderado) = -\frac{2}{5} * \log_2 \left(\frac{2}{5} \right) - \frac{1}{5} * \log_2 \left(\frac{1}{5} \right) - \frac{2}{5} * \log_2 \left(\frac{2}{5} \right)$$

$$Entropy(Desconhecida) = 1,52$$

- Para o Histórico(Ruim)

$$Entropy(Baixo) = -\frac{3}{4} * \log_2 \left(\frac{3}{4} \right) - \frac{1}{5} * \log_2 \left(\frac{1}{4} \right) - 0$$

$$Entropy(Ruim) = 0,81$$

Agora conseguimos calcular o ganho de informação do histórico

$$Gain(S, A) = Entropy(s) - \sum_{v \in Values(A)} -\frac{|S_v|}{|S|} Entropy(S_v)$$

onde

$$Entropy(s) = 1,53$$

então

$$Gain(Histórico) = 1,53 - \left(\frac{5}{14} * 1,37 \right) - \left(\frac{5}{14} * 1,52 \right) - \left(\frac{4}{14} * 0,81 \right)$$

$$Gain(Histórico) = 0,26$$

vamos iniciar dar continuidade com atributo Dívida

Dívida	Risco
alta	alto
alta	alto
baixa	moderado
baixa	alto
baixa	baixo
baixa	baixo
baixa	alto
baixa	moderado
baixa	baixo
alta	baixo
alta	alto

$$Alta = \frac{7}{14}$$

$$Baixa = \frac{7}{14}$$

alta	moderado
alta	baixo
alta	alto

Dívida	Risco (Alto)	Risco (Moderado)	Risco (Baixo)
Alta = 7/14	4/7	1/7	2/7
Baixa = 7/14	2/7	2/7	3/7

Calculando a entropia para o atributo escolhido

- Para o Dívida(Alta)

$$Entropy(Alto) = -\frac{4}{7} * \log_2 \left(\frac{4}{7} \right) - \frac{1}{7} * \log_2 \left(\frac{1}{7} \right) - \frac{2}{7} * \log_2 \left(\frac{2}{7} \right)$$

$$Entropy(Alto) = 1,38$$

- Para Dívida(Baixa)

$$Entropy(Baixa) = -\frac{2}{7} * \log_2 \left(\frac{2}{7} \right) - \frac{2}{7} * \log_2 \left(\frac{2}{7} \right) - \frac{3}{7} * \log_2 \left(\frac{3}{7} \right)$$

$$Entropy(Baixa) = 1,56$$

Agora conseguimos calcular o ganho da dívida

$$Gain(S, A) = Entropy(s) - \sum_{v \in Values(A)} -\frac{|S_v|}{|S|} Entropy(S_v)$$

onde

$$Entropy(s) = 1,53$$

então

$$Gain(Dívida) = 1,53 - \left(\frac{7}{14} * 1,38 \right) - \left(\frac{7}{14} * 1,56 \right)$$

$$Gain(Dívida) = 0,06$$

vamos iniciar dar continuidade com atributo Garantia

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

$$Nenhuma = \frac{11}{14}$$

$$Adequada = \frac{3}{14}$$

Dívida	Risco (Alto)	Risco (Moderado)	Risco (Baixo)
Nenhuma= 11/14	6/11	2/11	3/11
Adequada = 3/14	0	1/3	2/3

Calculando a entropia para o atributo escolhido

- Para o Garantia(Nenhuma)

$$Entropy(Nenhuma) = -\frac{6}{11} * \log_2 \left(\frac{2}{11} \right) - \frac{2}{11} * \log_2 \left(\frac{3}{11} \right) - \frac{2}{7} * \log_2 \left(\frac{4}{11} \right)$$

$$Entropy(Nenhuma) = 1,44$$

- Para o Garantia(Adequada)

$$Entropy(Adequada) = 0 - \frac{1}{3} * \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} * \log_2 \left(\frac{2}{3} \right)$$

$$Entropy(Adequada) = 0,92$$

Agora conseguimos calcular a Garantia

$$Gain(S, A) = Entropy(s) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

onde

$$Entropy(s) = 1,53$$

então

$$Gain(Garantia) = 1,53 - \left(\frac{11}{14} * 1,44\right) - \left(\frac{3}{14} * 0,92\right)$$

$$Gain(Dívida) = 0,20$$

vamos iniciar dar continuidade com atributo Renda

Renda	Risco
0_15	alto
15_35	alto
15_35	moderado
acima_35	alto
acima_35	baixo
acima_35	baixo
0_15	alto
acima_35	moderado
acima_35	baixo
acima_35	baixo
0_15	alto
15_35	moderado
acima_35	baixo
15_35	alto

$$< 15 = \frac{3}{14}$$

$$\geq 15 < 35 = \frac{4}{14}$$

$$> 35 = \frac{7}{14}$$

Renda	Risco (Alto)	Risco (Moderado)	Risco (Baixo)
<15	3/3	0	0
≥15<35	2/4	2/4	0
>35	1/7	5/7	1/7

Calculando a entropia para o atributo escolhido

- Para o Renda(<15)

$$Entropy(< 15) = -\frac{3}{3} * \log_2 \left(\frac{3}{3} \right) - 0 - 0$$

$$Entropy(< 15) = 0$$

- Para o Renda($\geq 15 < 35$)

$$Entropy(\geq 15 < 35) = -\frac{2}{4} * \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} * \log_2 \left(\frac{2}{4} \right) - 0$$

$$Entropy(< 15) = 1,00$$

- Para o Renda(>35)

$$Entropy(> 35) = -\frac{1}{7} * \log_2 \left(\frac{1}{7} \right) - \frac{5}{7} * \log_2 \left(\frac{5}{7} \right) - \frac{1}{7} * \log_2 \left(\frac{1}{7} \right)$$

$$Entropy(< 15) = 1,15$$

$$Gain(Renda) = 1,53 - \left(\frac{3}{14} * 0 \right) - \left(\frac{4}{14} * 1 \right) - \left(\frac{7}{14} * 1,15 \right)$$

$$Gain(Renda) = 0,66$$

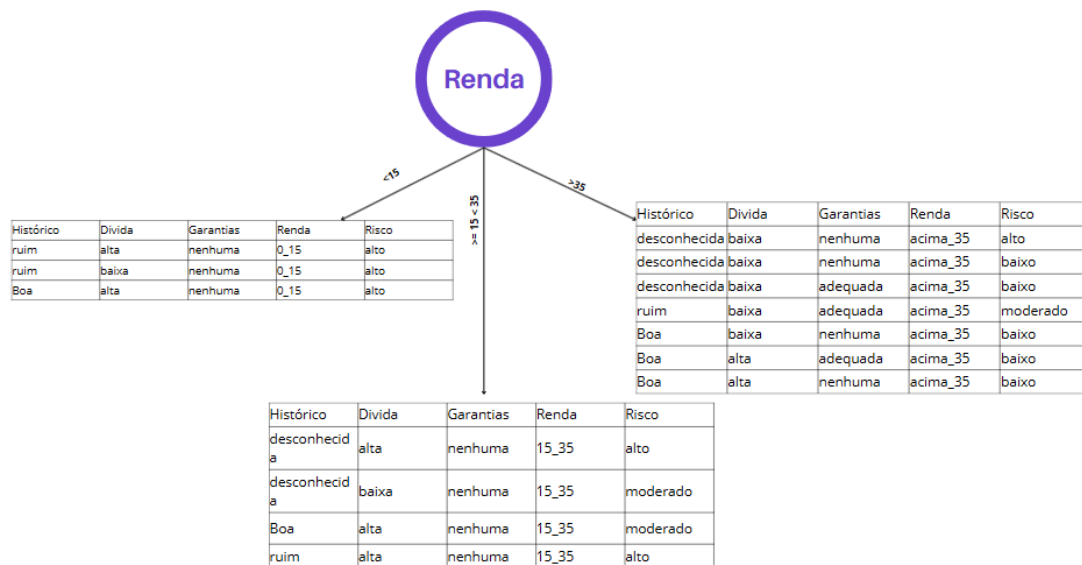
Com todos os ganhos agora podemos tomar a decisão

$$Gain(Historia) = 0,26$$

$$Gain(Dívida) = 0,06$$

$$Gain(Garantia) = 0,20$$

$$Gain(Renda) = 0,66$$



desse modo criamos mais três bases que precisará passar pelo algoritmo novamente