



INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE FEDERAL DE ALAGOAS - UFAL

LÍVIA ANDRESSA DA SILVA SANTOS

Lista 1

MACEIÓ

2024

Questão 1 -

	História de Crédito	Dívida	Garantia	Renda	Risco
E1	Ruim	Alta	Nenhuma	\$0 a \$15k	Alto
E2	Desconhecida	Alta	Nenhuma	\$0 a \$15k	Alto
E3	Desconhecida	Baixa	Nenhuma	\$15 a \$35k	Moderado
E4	Desconhecida	Baixa	Nenhuma	\$0 a \$15k	Alto
E5	Desconhecida	Baixa	Nenhuma	Acima de \$35k	Baixo
E6	Desconhecida	Baixa	Adequada	Acima de \$35k	Baixo
E7	Ruim	Baixa	Nenhuma	\$0 a \$15k	Alto
E8	Ruim	Baixa	Adequada	Acima de \$35k	Moderado
E9	Boa	Baixa	Nenhuma	Acima de \$35k	Baixo
E10	Boa	Alta	Adequada	Acima de \$35k	Baixo
E11	Boa	Alta	Nenhuma	\$0 a \$15k	Alto
E12	Boa	Alta	Nenhuma	\$15 a \$35k	Moderado
E13	Boa	Alta	Nenhuma	Acima de \$35k	Baixo
E14	Ruim	Alta	Nenhuma	\$15 a \$35k	Alto

$$Entropia(S) = Entropia(Risco) = Entropia[6 \text{ alto} , 3 \text{ moderado} , 5 \text{ baixo}]$$

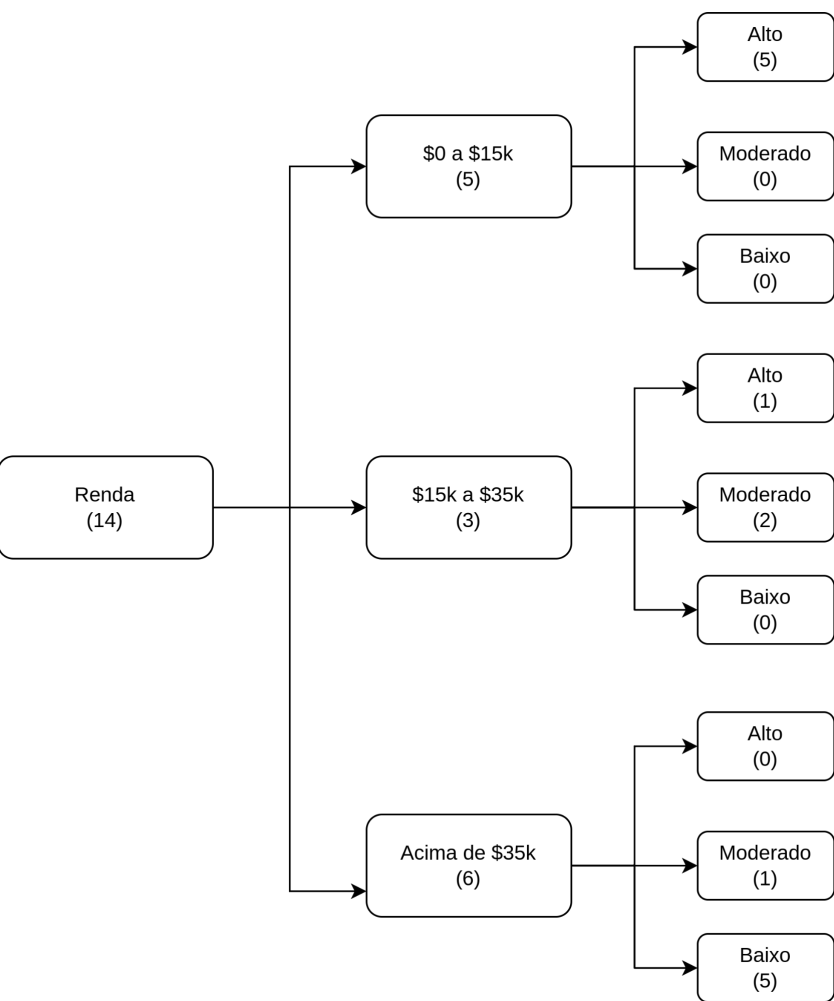
Passo 1:

	Risco
E1	Alto
E2	Alto
E3	Moderado
E4	Alto
E5	Baixo
E6	Baixo
E7	Alto
E8	Moderado
E9	Baixo
E10	Baixo
E11	Alto
E12	Moderado
E13	Baixo
E14	Alto

$$Entropia(Risco) = - (6/14) * \log_2(6/14) - (3/14) * \log_2(3/14) - (5/14)$$

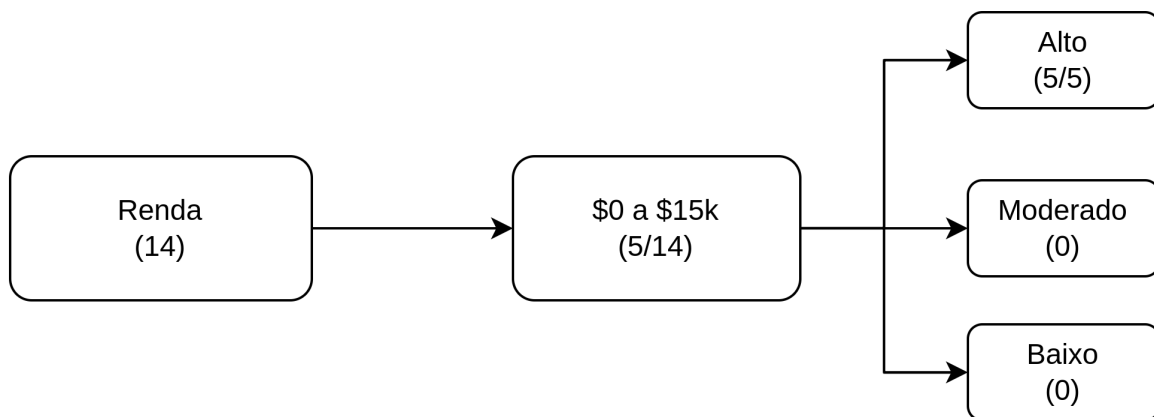
$$* \log_2(5/14) = 1.530$$

Ganho de informação com o atributo *Renda*:



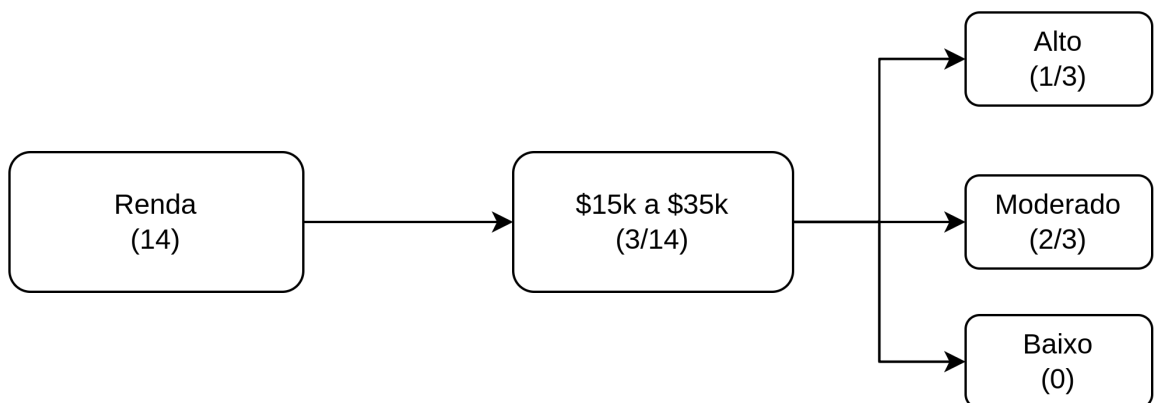
Renda	Risco
\$0 a \$15k	Alto
\$0 a \$15k	Alto
\$15 a \$35k	Moderado
\$0 a \$15k	Alto
Acima de \$35k	Baixo
Acima de \$35k	Baixo
\$0 a \$15k	Alto
Acima de \$35k	Moderado
Acima de \$35k	Baixo
Acima de \$35k	Baixo
\$0 a \$15k	Alto
\$15 a \$35k	Moderado
Acima de \$35k	Baixo
\$15 a \$35k	Alto

Renda: de \$0 a \$15k



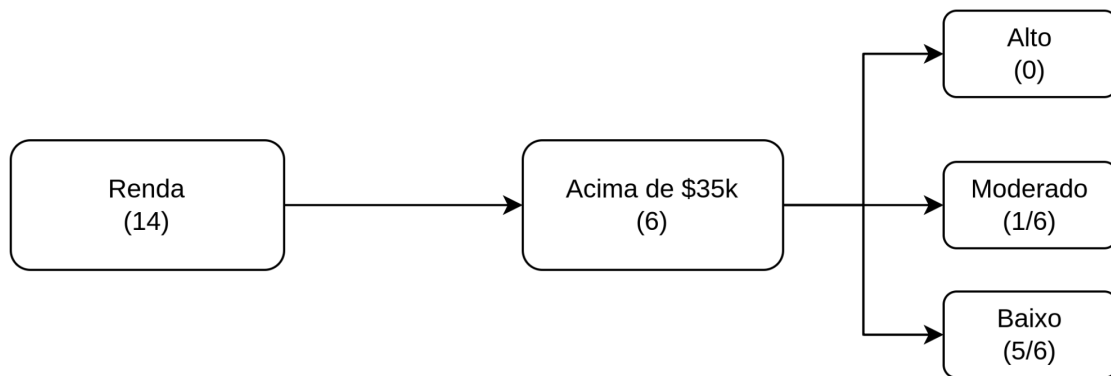
$Renda = \$0 \text{ a } \$15k \text{ (5/14)}, \text{ tendo } S \$0 \text{ a } \$15k \leftarrow [5 \text{ alto}, 0 \text{ moderado}, 0 \text{ baixo}]$ $Entropia(S \$0 \text{ a } 15k) = - (5/5) \log_2 (5/5) = 0$

Renda: \$15 a \$35k



$Renda = \$15 \text{ a } \$30k \text{ (3/14)}, \text{ tendo } S \$0 \text{ a } \$15k \leftarrow [1 \text{ alto}, 2 \text{ moderado}, 0 \text{ baixo}]$ $Entropia(s \$15 \text{ a } 30k) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0.91829$

Renda: Acima de 35k



$Renda = Acima\ de\ \$30k\ (6/14)$, tendo S acima de $30k \leftarrow [0\ alto, 1\ moderado, 5\ baixo]$ Entropia(s Acima de $30k$) = $-(1/6)\log_2(1/6) - (5/6)\log_2(5/6) = 0.65002$

Entropia para o atributo Renda:

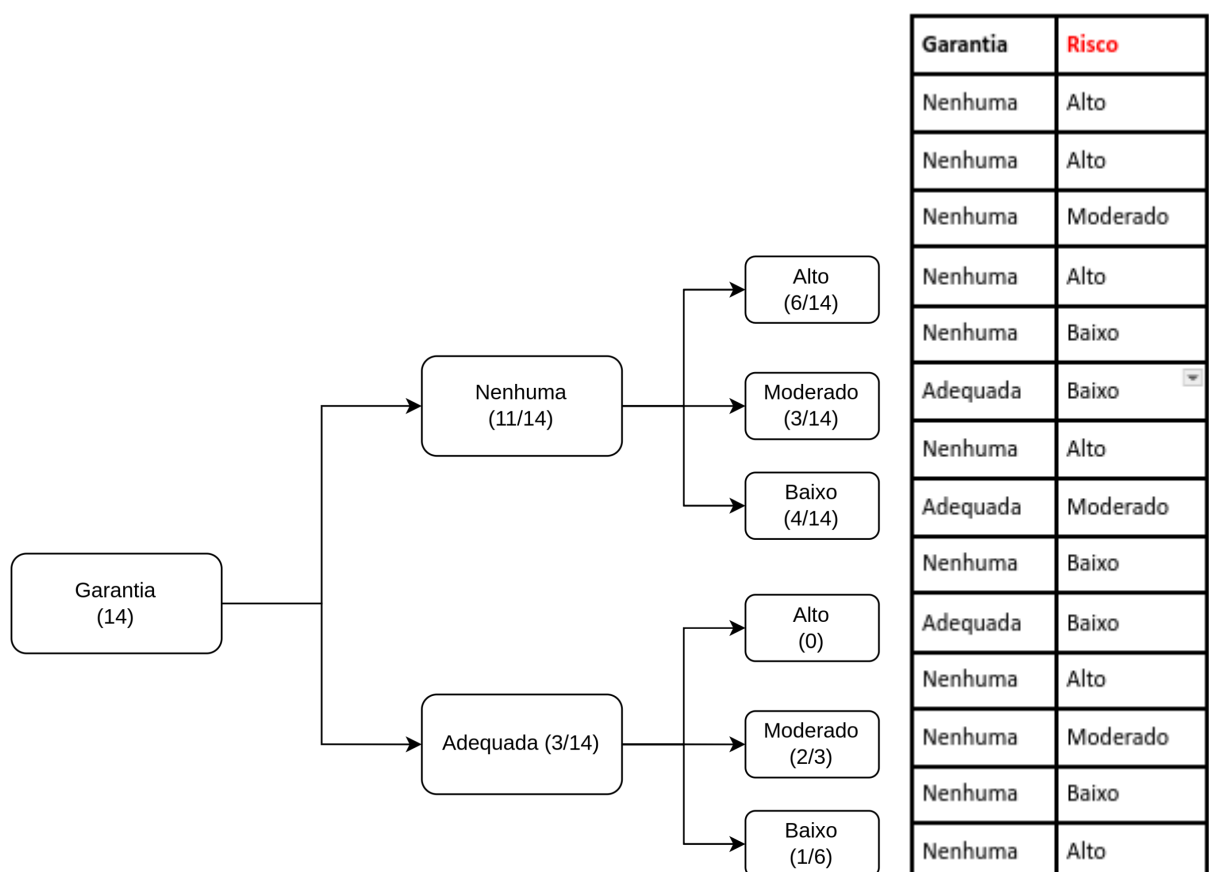
$Entropia(S\ \$0\ a\ 15k) = 0$

$Entropia(s\ \$15\ a\ 30k) = 0.91829$

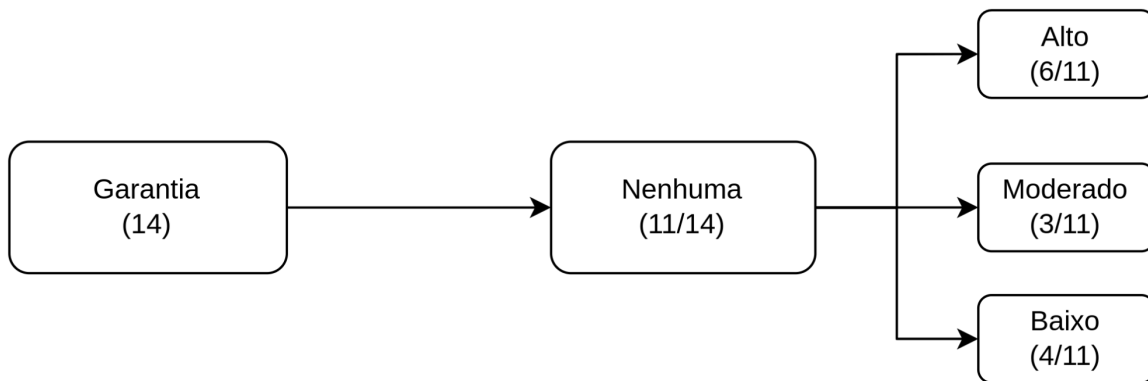
$Entropia(s\ Acima\ de\ 30k) = 0.65002$

$Ganho(S, Renda) = 1.53061 - (5/14) * 0 - (3/14) * 0.91829 - (6/14)$

$* 0.65002 = \underline{1.05525}$



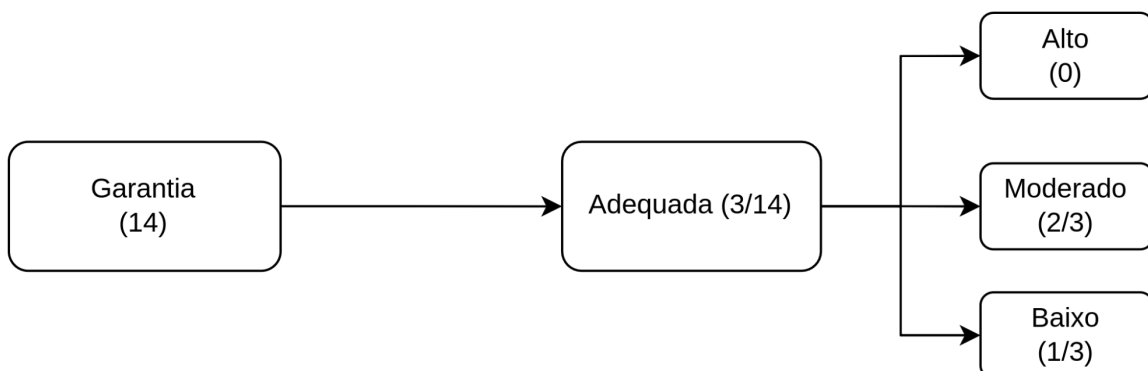
,Garantia: Nenhuma



Garantia = Nenhuma (11/14), tendo $S_{Nenhuma} \leftarrow [6_{alto}, 3_{moderado}, 4_{baixo}]$

Entropia($S_{Nenhuma}$) = $-(6/11)\log_2(6/11) - (3/11)\log_2(3/11) - (4/11)\log_2(4/11) = 1.51890$

Garantia: Adequada



Garantia = Adequada (3/14), tendo $S_{Nenhuma} \leftarrow [0_{alto}, 2_{moderado}, 1_{baixo}]$

Entropia($S_{Adequada}$) = $-(2/3)\log_2(2/3) - (1/3)\log_2(1/3) = 0.91829$

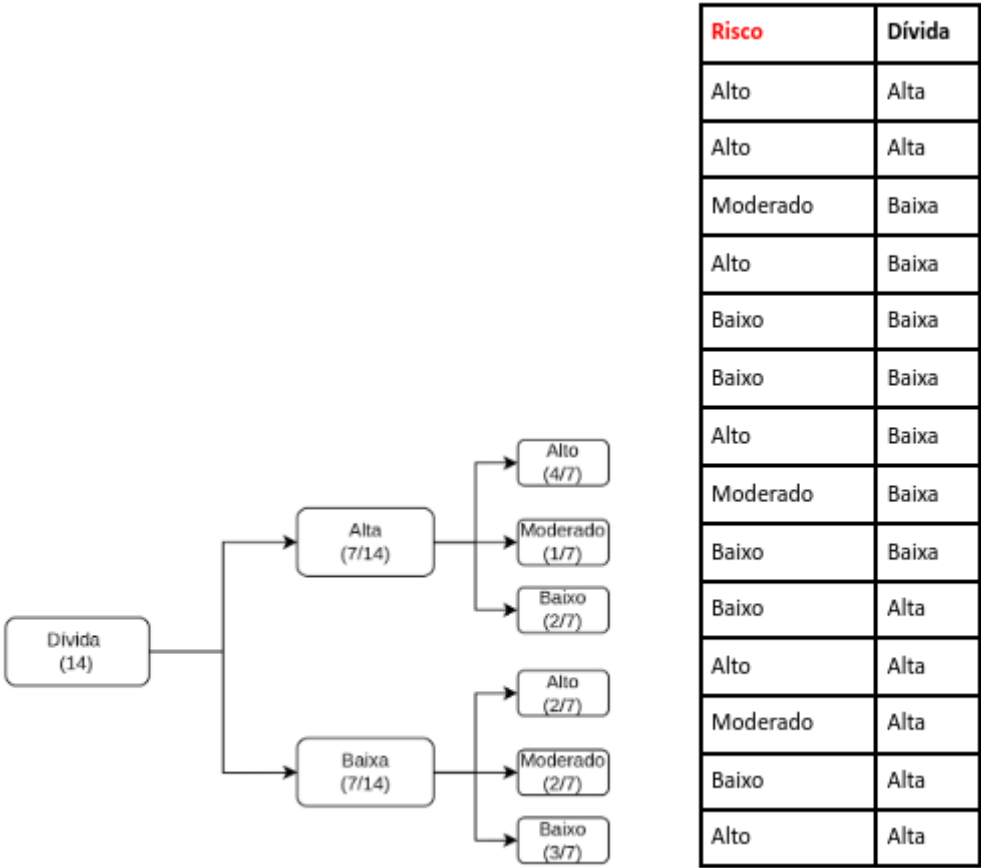
Entropia para o atributo Garantia:

Entropia($S_{Nenhuma}$) = 1.51890

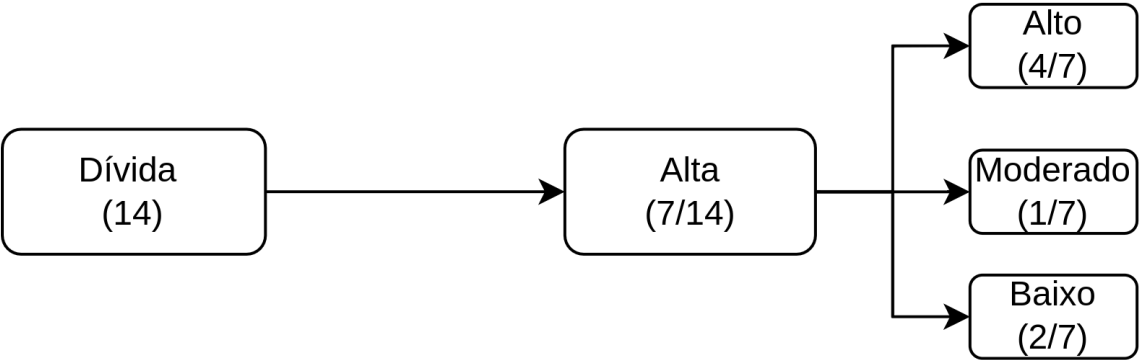
Entropia($S_{Adequada}$) = 0.91829

*Ganho($S, Garantia$) = $1.53061 - (11/14) * 1.51890 - (3/14) * 0.91829 = 0.1404$*

Ganho de atributo com **Dívida**:



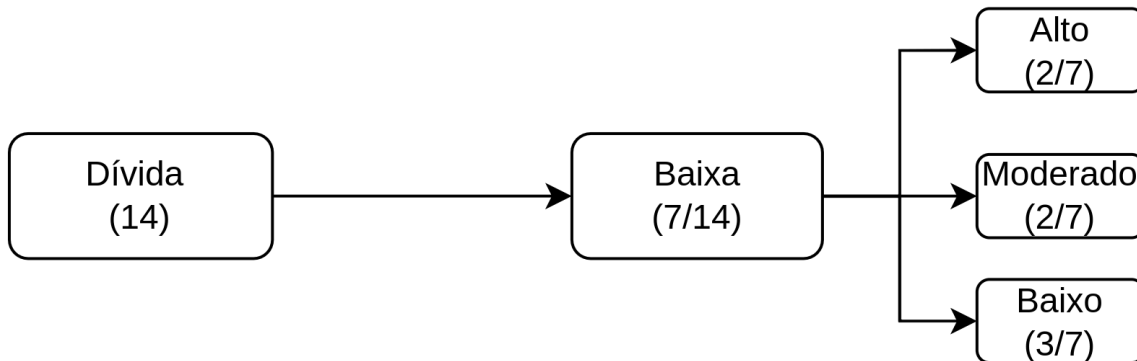
Dívida: **Alta**



Dívida = Alta (7/14), tendo $S_{Alta} \leftarrow [4_{alto}, 1_{moderado}, 2_{baixo}]$

$$Entropia(S Alta) = - (4/7)\log_2 (4/7) - (1/7)\log_2 (1/7) - (2/7)\log_2 (2/7) = 1.37878$$

Dívida: **Baixa**



$Dívida = Baixa (7/14)$, tendo $S_{baixa} \leftarrow [2 \text{ alto}, 2 \text{ moderado}, 3 \text{ baixo}]$

$$Entropia(S_{baixa}) = - (2/7)\log_2 (2/7) - (2/7)\log_2 (2/7) - (3/7)\log_2 (3/7) = 1.55665$$

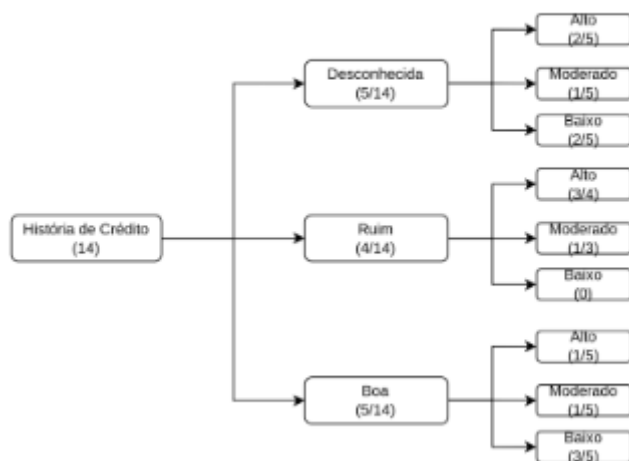
Entropia para o atributo Dívida: E

$$Entropia(S Alta) = 1.37878$$

$$Entropia(S Baixa) = 1.55665$$

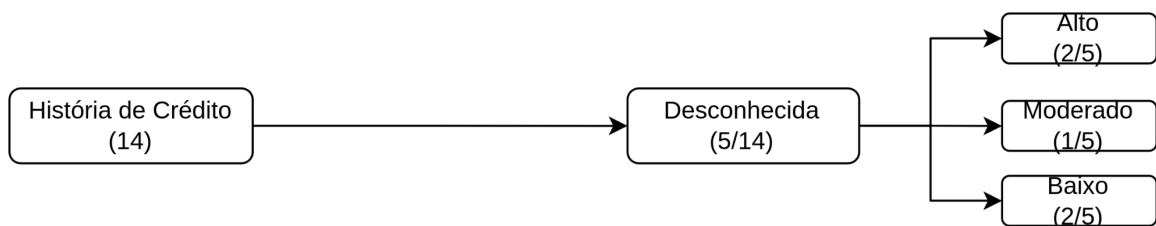
$$Ganho(S, Dívida) = 1.53061 - (7/14) * 1.37878 - (7/14) * 1.55665 = 0.062895$$

Ganho de atributo com **História de crédito**:



Risco	História de Crédito
Alto	Ruim
Alto	Desconhecida
Moderado	Desconhecida
Alto	Desconhecida
Baixo	Desconhecida
Baixo	Desconhecida
Alto	Ruim
Moderado	Ruim
Baixo	Boa
Baixo	Boa
Alto	Boa
Moderado	Boa
Baixo	Boa
Alto	Ruim

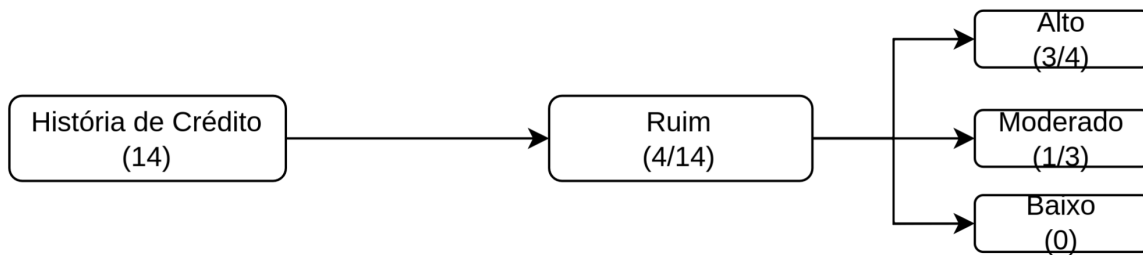
História: *Desconhecida*



História = Desconhecida (5/14), tendo S Desconhecida <- [2 alto , 1 moderado , 2 baixo]

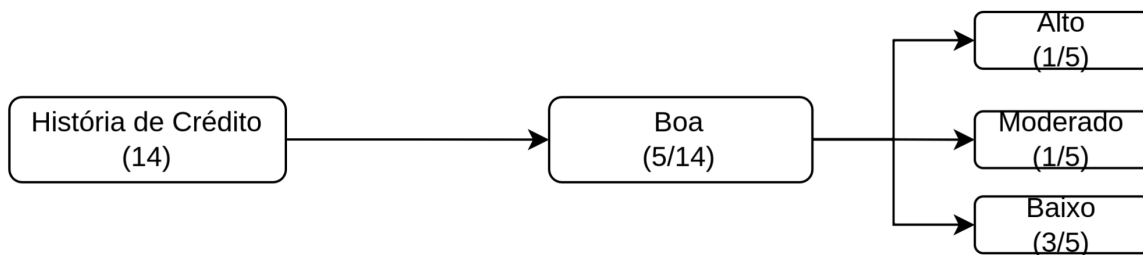
$$Entropia(S Desconhecida) = - (2/5)\log_2 (2/5) - (1/5)\log_2 (1/5) - (2/5)\log_2 (2/5) = 1.52192$$

História: Ruim



História = Ruim (4/14), tendo $S_{Ruim} \leftarrow [3 \text{ alto}, 1 \text{ moderado}, 0 \text{ baixo}]$ $Entropia(S_{Ruim}) = - (3/14)\log_2 (3/14) - (1/14)\log_2 (1/14) = 1.00454$

História: Boa



História = Boa (5/14), tendo $S_{Boa} \leftarrow [1 \text{ alto}, 1 \text{ moderado}, 3 \text{ baixo}]$ $Entropia(S_{Boa}) = - (1/5)\log_2 (1/5) - (1/5)\log_2 (1/5) - (3/5)\log_2 (3/5) = 1.37095$

Entropia para o atributo História de Crédito:

$$Entropia(S Desconhecida) = 1.52192$$

$$Entropia(S Ruim) = 1.00454$$

$$Entropia(S Boa) = 1.37095$$

$$Ganho(S, História) = 1.53061 - (5/14) * 1.52192 - (4/14) * 1.00454 - (5/14) * 1.37095 = -1.01363$$

Temos que:

$$\text{Ganho}(S, \text{Garantia}) = 0.14041$$

$$\text{Ganho}(S, \text{Renda}) = 1.05525$$

$$\text{Ganho}(S, \text{Dívida}) = 0.062895$$

$$\text{Ganho}(S, \text{História}) = -1.01363$$

O atributo a ser escolhido deve ser *Renda*. Renda: \$0 a 15k

	História de Crédito	Dívida	Garantia	Renda	Risco
E1	Ruim	Alta	Nenhuma	\$0 a \$15k	Alto
E2	Desconhecida	Alta	Nenhuma	\$0 a \$15k	Alto
E4	Desconhecida	Baixa	Nenhuma	\$0 a \$15k	Alto
E7	Ruim	Baixa	Nenhuma	\$0 a \$15k	Alto
E11	Boa	Alta	Nenhuma	\$0 a \$15k	Alto

Todos apontam para risco Alto.

Renda: \$15k a \$35k

	História de Crédito	Dívida	Garantia	Renda	Risco
E3	Desconhecida	Baixa	Nenhuma	\$15 a \$35k	Moderado
E12	Boa	Alta	Nenhuma	\$15 a \$35k	Moderado
E14	Ruim	Alta	Nenhuma	\$15 a \$35k	Alto

$$\text{Entropia}(15k - 35k) = 0 - (2/3)\log_2(2/3) - (1/3)\log_2(1/3) = 0.91829$$

$$Entropia(Garantia Nenhuma) = - (2/3)\log_2 (2/3) - (1/3)\log_2 (1/3) = 0.91829$$

$$Ganho(Garantia) = 0.91829 - (3/3) * 0.91829 = 0$$

$$Entropia(Dívida Baixa) = - (1/3)\log_2 (1/3) = 0$$

$$Entropia(Dívida Alta) = - (1/3)\log_2 (1/3) - (1/3)\log_2 (1/3) = 1.05664$$

$$Ganho(Dívida) = 0.91829 - (1/3) * 0 - (2/3) * 1.05664 = 0.21386$$

$$Entropia(História Desconhecida) = - (1/3)\log_2 (1/3) = 0$$

$$Entropia(História Ruim) = - (1/3)\log_2 (1/3) = 0$$

$$Entropia(História Boa) = - (1/3)\log_2 (1/3) = 0$$

$$Ganho(História) = 0.91829 - (1/3) * 0 - (1/3) * 0 - (1/3) * 0 = 0.91829$$

O atributo a ser escolhido deve ser Dívida. Dívida:

	História de Crédito	Garantia	Dívida	Risco
E3	Desconhecida	Nenhuma	Baixa	Moderado
E12	Boa	Nenhuma	Alta	Moderado
E14	Ruim	Nenhuma	Alta	Alto

Para a dívida Baixa, o risco é Moderado.

Dívida Alta:

	História de Crédito	Dívida	Garantia	Risco
E12	Boa	Alta	Nenhuma	Moderado
E14	Ruim	Alta	Nenhuma	Alto

$$Entropia(Alta) = - (1/2)\log_2 (1/2) - (1/2)\log_2 (1/2) = 1$$

$$Entropia(Garantia) = - (1/2)\log_2 (1/2) - (1/2)\log_2 (1/2) = 1$$

$$Ganho(Garantia) = 0$$

$$Entropia(História Boa) = - (1/1)\log_2 (1/1) = 0$$

$$Entropia(História Ruim) = - (1/1)\log_2 (1/1) = 0$$

$$Ganho(História) = 1$$

O atributo a ser escolhido deve ser História, e, como se vê pela tabela, história boa gera um risco Moderado e história ruim gera um risco Alto.

Renda: Acima de \$35k

	História de Crédito	Dívida	Garantia	Renda	Risco
E5	Desconhecida	Baixa	Nenhuma	Acima de \$35k	Baixo
E6	Desconhecida	Baixa	Adequada	Acima de \$35k	Baixo
E8	Ruim	Baixa	Adequada	Acima de \$35k	Moderado
E9	Boa	Baixa	Nenhuma	Acima de \$35k	Baixo
E10	Boa	Alta	Adequada	Acima de \$35k	Baixo
E13	Boa	Alta	Nenhuma	Acima de \$35k	Baixo

$$Entropia(35k -) = - (5/6)\log_2(5/6) - (1/6)\log_2(1/6) = 0.65002$$

$$Entropia(Garantia Nenhuma) = 0$$

$$Entropia(Garantia Adequada) = - (2/3)\log_2 (2/3) - (1/3)\log_2 (1/3) = 0.91829$$

$$Ganho(Garantia) = 0.65002 - (3/3) * 0 - (3/3) * 0.91829 = -0.26827$$

$$Entropia(Dívida Baixa) = - (3/4)\log_2 (3/4) - (1/4)\log_2 (1/4) = 0.81127$$

$$Entropia(Dívida Alta) = 0$$

$$Ganho(Dívida) = 0.65002 - (4/6) * 0.81127 - (2/6) * 0 = 0.10917$$

$$Entropia(História Desconhecida) = 0$$

$$Entropia(História Ruim) = 0$$

$$Entropia(História Boa) = 0$$

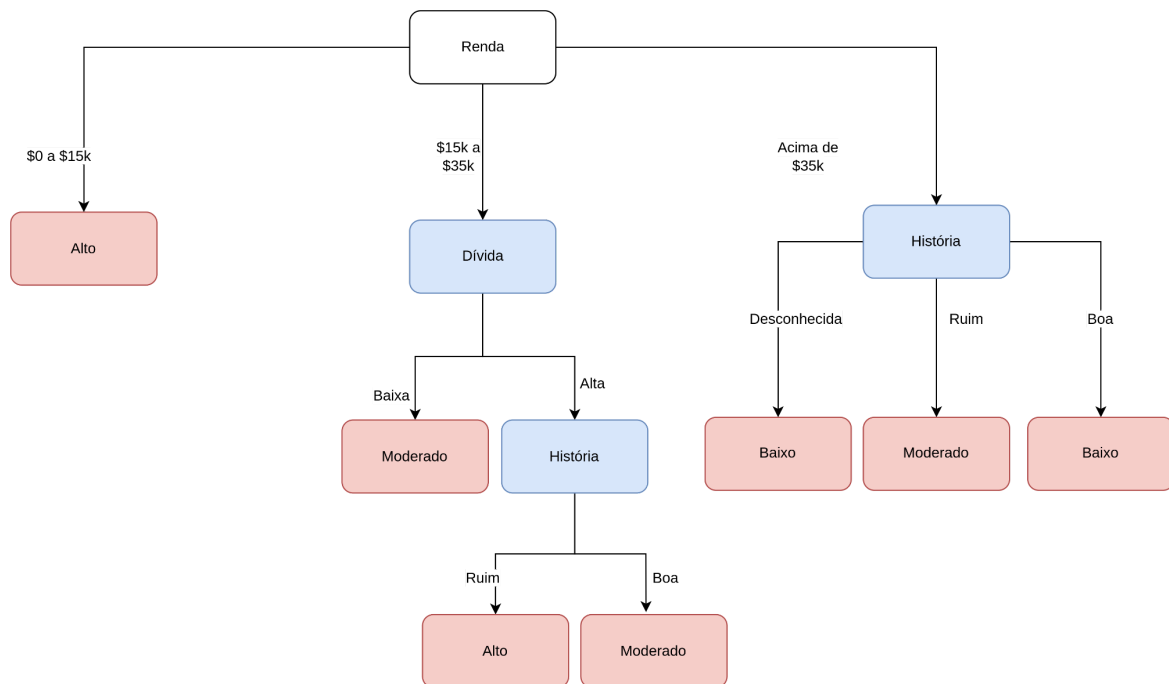
$$Ganho(História) = 0.65002$$

O atributo a ser escolhido deve ser História.

	História de Crédito	Dívida	Garantia	Risco
E5	Desconhecida	Baixa	Nenhuma	Baixo
E6	Desconhecida	Baixa	Adequada	Baixo
E8	Ruim	Baixa	Adequada	Moderado
E9	Boa	Baixa	Nenhuma	Baixo
E10	Boa	Alta	Adequada	Baixo
E13	Boa	Alta	Nenhuma	Baixo

Vê-se pela tabela que História Desconhecida gera Risco Baixo, História Ruim gera risco Moderado e História Boa gera risco Baixo.

Portanto, esta deve ser a Árvore de Decisão para o algoritmo ID3:



Árvore gerada rodando o algoritmo ID3 na aplicação WEKA:

Classifier output

=== Run information ===

Scheme: weka.classifiers.trees.Id3
Relation: tabela (2)
Instances: 14
Attributes: 5
História de Crédito
Dívida
Garantia
Renda
Risco
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Id3

Renda = \$0 a \$15k: Alto
Renda = \$15 a \$35k
| História de Crédito = Ruim: Alto
| História de Crédito = Desconhecida: Moderado
| História de Crédito = Boa: Moderado
Renda = Acima de \$35k
| História de Crédito = Ruim: Moderado
| História de Crédito = Desconhecida: Baixo
| História de Crédito = Boa: Baixo

Algoritmo C4.5

Para começar, é calculado primeiramente a entropia de **Risco**.

$$\begin{aligned} \text{Entropia}(\text{Risco}) &= - (6/14) * \log_2 (6/14) - (3/14) * \log_2 (3/14) - (5/14) * \log_2 (5/14) \\ &= 1.5794 \end{aligned}$$

Agora, precisamos calcular as taxas de ganhos:

$$\text{GainRatio}(A) = \text{Gain}(A) / \text{SplitInfo}(A)$$

$$\text{SplitInfo}(A) = - \sum |D_j|/|D| \times \log_2 |D_j|/|D|$$

Calculando o atributo de dívida:

$$\text{Gain}(\text{Decision}, \text{Dívida}) =$$

$$\begin{aligned} &\text{Entropia}(\text{Risco}) - \sum [p(\text{Risco} | \text{Dívida}) * \text{Entropia}(\text{Risco}, \text{Dívida})] \\ &\text{Gain}(\text{Risco}, \text{Dívida}) = \\ &\text{Entropia}(\text{Risco}) - [p(\text{Risco} | \text{Dívida=Baixa}) * \text{Entropia}(\text{Risco} | \text{Dívida=Baixa})] + \\ &[p(\text{Risco} | \text{Dívida=Alta}) * \text{Entropia}(\text{Risco} | \text{Dívida=Alta})] \end{aligned}$$

Entropia(Risco | Dívida=Baixa)

$$\begin{aligned} &- p(\text{Baixo}) * \log_2(p(\text{Baixo})) - p(\text{Moderado}) * \log_2(p(\text{Moderado})) - p(\text{Alto}) * \log_2(p(\text{Alto})) \\ &= -3/7 * \log_2(3/7) - 2/7 * \log_2(2/7) - 2/7 * \log_2(2/7) \\ &= 1.55 \end{aligned}$$

Entropia(Risco | Dívida=Alta)

$$\begin{aligned} &= - 2/7 * \log_2(2/7) - 1/7 * \log_2(1/7) - 4/7 * \log_2(4/7) \\ &= 1.37 \end{aligned}$$

$$\text{Gain}(\text{Risco}, \text{Dívida})$$

$$= 1.53 - (7/14) * 1.55 - (7/14) * 1.37 = 0.07$$

SplitInfo(Risco, Dívida)

$$= - (7/14) * \log(7/14) - (7/14) * \log(7/14) = 1$$

GainRatio(Risco, Dívida)

$$\text{Gain}(\text{Risco}, \text{Dívida}) / \text{SplitInfo}(\text{Risco}, \text{Dívida}) = 0.07 / 1 = 0.07$$

Calculando o atributo de garantia:

Gain(Decision, Garantia) =

$$\begin{aligned} & \text{Entropia}(\text{Risco}) - \sum [p(\text{Risco} | \text{Garantia}) * \text{Entropia}(\text{Risco}, \text{Garantia})] \\ \text{Gain}(\text{Risco}, \text{Garantia}) &= \text{Entropia}(\text{Risco}) - [p(\text{Risco} | \text{Garantia}=\text{Nenhuma}) * \\ & \text{Entropia}(\text{Risco} | \text{Garantia}=\text{Nenhuma})] + [p(\text{Risco} | \text{Garantia}=\text{Adequada}) * \\ & \text{Entropia}(\text{Risco} | \text{Garantia}=\text{Adequada})] \end{aligned}$$

Entropia(Risco | Garantia=Nenhuma)

$$\begin{aligned} &= - p(\text{Baixo}) * \log p(\text{Baixo}) - p(\text{Moderado}) * \log p(\text{Moderado}) - p(\text{Alto}) * \log p(\text{Alto}) \\ &= - (3/11) * \log(3/11) - (2/11) * \log(2/11) - 6/11 * \log(6/11) \\ &= 1.43 \end{aligned}$$

Entropia(Risco | Garantia=Adequada)

$$\begin{aligned} &= - (2/3) * \log(2/3) - (1/3) * \log(1/3) \\ &= 0.91 \end{aligned}$$

Gain(Risco, Garantia)

$$1.53 - (11/14) * 1.43 - 3/14 * 0.91 = 0.14$$

SplitInfo(Risco, Garantia)

$$- (11/14) * \log(11/14) - (3/14) * \log(3/14) = 0.74$$

GainRatio(Risco, Garantia)

$$Gain(Risco, Garantia) / SplitInfo(Risco, Garantia) = 0.14 / 0.74 = 0.07$$

Calculando o atributo de História de Crédito:

$$Gain(Decision, História) =$$

$$\begin{aligned} & Entropia(Risco) - \sum [p(Risco | História) * Entropia(Risco, História)] \text{ Gain}(Risco, História) \\ = & Entropia(Risco) - [p(Risco | História=Ruim) * Entropia(Risco | História=Ruim)] + \\ & [p(Risco | História=Desconhecida) * Entropia(Risco | História=Desconhecida)] + \\ & [p(Risco | História=Boa) * Entropia(Risco | História=Boa)] \end{aligned}$$

Entropia(Risco | História=Desconhecida)

$$\begin{aligned} = & -p(Baixo) * \log p(Baixo) - p(Moderado) * \log p(Moderado) - p(Alto) * \log p(Alto) \\ = & - (2/5) \log(2/5) - (1/5) \log(1/5) - (2/5) \log(2/5) \\ = & 1.52 \end{aligned}$$

Entropia(Risco | História=Ruim)

$$\begin{aligned} = & - (3/4) \log(3/4) - (1/4) \log(1/4) \\ = & 1 \end{aligned}$$

Entropia(Risco | História=Boa)

$$\begin{aligned} = & - (1/5) \log(1/5) - (1/5) \log(1/5) - (3/5) \log(3/5) \\ = & 1.37 \end{aligned}$$

$$\begin{aligned} Gain(Risco, História) = & 1.5 - (5/14) * 1.5 - (4/14) * 1 - \\ & (5/4) * 1.3 = 1.07 \end{aligned}$$

SplitInfo(Risco, História)

$$- (4/14) * \log(4/14) - (5/14) * \log(5/14) - (5/14) * \log(5/14) = 1.57$$

GainRatio(Risco, História)

$$Gain(Risco, História) / SplitInfo(Risco, História) = 1.07 / 1.57 = 0.68$$

Calculando o atributo de Renda:

Gain(Decision, Renda) =

$$\begin{aligned} & \text{Entropia(Risco)} - \sum [p(\text{Risco} | \text{Renda}) * \text{Entropia(Risco, Renda)}] \\ \text{Gain(Risco, Renda)} = & \text{Entropia(Risco)} - [p(\text{Risco} | \text{Renda}=0 \text{ a } 15\text{k}) * \text{Entropia(Risco} | \text{História=Renda}=0 \text{ a } 15\text{k})] + \\ & [p(\text{Risco} | \text{Renda}=15 \text{ a } 35\text{k}) * \text{Entropia(Risco} | \text{Renda}=15 \text{ a } 35\text{k})] + [p(\text{Risco} | \text{Renda}= >35\text{k}) * \\ & \text{Entropia(Risco} | \text{Renda}= >35\text{k})] \end{aligned}$$

Entropia(Risco | Renda=0 a 15k)

$$- (5/5) \log(5/5) = 0$$

Entropia(Risco | Renda=15 a 35k)

$$- (1/3) \log(1/3) - (2/3) \log(2/3) = 0.91$$

Entropia(Risco | Renda= >35k)

$$- (1/6) \log(1/6) - (5/6) \log(5/6) = 0.65$$

$$\begin{aligned} \text{Gain(Risco, Renda)} = & 1.53 - (5/14) * 0 - (3/14) * 0.91 - (6/14) \\ & * 0.65 = 1.05 \end{aligned}$$

SplitInfo(Risco, História)

$$- (4/14) * \log(4/14) - (4/14) * \log(4/14) - (6/14) * \log(6/14) = 1.55$$

GainRatio(Risco, História)

$$\text{Gain(Risco, História)} / \text{SplitInfo(Risco, História)} = 1.05 / 1.55 = 0.67$$

Atributo	Ganho	Taxa de Ganho
História de Crédito	-1	-0.63
Dívida	0.07	0.07
Garantia	0.14	0.189189189
Renda	1.05	1.55

Baseando-se na taxa de ganho, o atributo **Renda** será o nó da nossa árvore.

Tendo isto em conta, separando a tabela de acordo com a renda, temos: Renda = \$0 a \$15k

	História de Crédito	Dívida	Garantia	Renda	Risco
E1	Ruim	Alta	Nenhuma	\$0 a \$15k	Alto
E4	Desconhecida	Baixa	Nenhuma	\$0 a \$15k	Alto
E7	Ruim	Baixa	Nenhuma	\$0 a \$15k	Alto
E11	Boa	Alta	Nenhuma	\$0 a \$15k	Alto

Independentemente da história de crédito ou dívida, não existirá garantia e o risco será sempre alto.

Renda = \$15k a \$35k

	História de Crédito	Dívida	Garantia	Renda	Risco
E2	Desconhecida	Alta	Nenhuma	\$15k a \$35k	Alto
E3	Desconhecida	Baixa	Nenhuma	\$15k a \$35k	Moderado
E12	Boa	Alta	Nenhuma	\$15k a \$35k	Moderado
E14	Ruim	Alta	Nenhuma	\$15k a \$35k	Alto

Mesmo com um aumento considerável da renda em comparação ao anterior, ainda existem altos (e moderados) riscos, sem nenhuma garantia e dívidas altas.

Renda = Acima de \$35k

	História de Crédito	Dívida	Garantia	Renda	Risco
E5	Desconhecida	Baixa	Nenhuma	Acima de \$35k	Baixo
E6	Desconhecida	Baixa	Adequada	Acima de \$35k	Baixo
E8	Ruim	Baixa	Adequada	Acima de \$35k	Moderado
E9	Boa	Baixa	Nenhuma	Acima de \$35k	Baixo
E10	Boa	Alta	Adequada	Acima de \$35k	Baixo
E13	Boa	Alta	Nenhuma	Acima de \$35k	Baixo

O risco tende sempre a apontar para “Baixo”, a menos que a história de crédito seja ruim.

Árvore gerada rodando o algoritmo C4.5 na aplicação WEKA:

```
=== Classifier model (full training set) ===  
  
J48 pruned tree  
-----  
  
Renda = $0 a $15k: Alto (5.0)  
Renda = $15 a $35k: Moderado (3.0/1.0)  
Renda = Acima de $35k: Baixo (6.0/1.0)  
  
Number of Leaves :    3  
Size of the tree :    4
```

Algoritmo CART

	História de Crédito	Dívida	Garantia	Renda	Risco
E1	Ruim	Alta	Nenhuma	\$0 a \$15k	Alto
E2	Desconhecida	Alta	Nenhuma	\$0 a \$15k	Alto
E3	Desconhecida	Baixa	Nenhuma	\$15 a \$35k	Moderado
E4	Desconhecida	Baixa	Nenhuma	\$0 a \$15k	Alto
E5	Desconhecida	Baixa	Nenhuma	Acima de \$35k	Baixo
E6	Desconhecida	Baixa	Adequada	Acima de \$35k	Baixo
E7	Ruim	Baixa	Nenhuma	\$0 a \$15k	Alto
E8	Ruim	Baixa	Adequada	Acima de \$35k	Moderado
E9	Boa	Baixa	Nenhuma	Acima de \$35k	Baixo
E10	Boa	Alta	Adequada	Acima de \$35k	Baixo
E11	Boa	Alta	Nenhuma	\$0 a \$15k	Alto
E12	Boa	Alta	Nenhuma	\$15 a \$35k	Moderado
E13	Boa	Alta	Nenhuma	Acima de \$35k	Baixo
E14	Ruim	Alta	Nenhuma	\$15 a \$35k	Alto

Passo 1: Calcular o Índice Gini para Risco:

$$Gini(D) = 1 - (5/14)^2 - (3/14)^2 - (6/14)^2 = 0.64285$$

Passo 2: Calcular o Índice Gini para os demais atributos:

$Gini_k(D) = \sum |D_i|/|D| * Gini(D_i)$, para $i = 1$ até n , onde D_i é uma partição de D

Então, para o atributo História de Crédito:

$Gini_{História}(D) = (5/14) * Gini(D_{História=Desconhecida}) + (4/14) * Gini(D_{História=Ruim}) + (5/14) * Gini(D_{História=Boa})$

$Gini_{História}(D) = (5/14) * [1 - (2/5)^2 - (1/5)^2 - (2/5)^2] + (4/14) * [1 - (0/3)^2 - (1/3)^2 - (2/3)^2] + (5/14) * [1 - (3/5)^2 - (1/5)^2 - (1/5)^2] = 0.5555$

Para o atributo Dívida:

$Gini_{Dívida}(D) = (7/14) * Gini(D_{Dívida=baixa}) + (7/14) * Gini(D_{Dívida=Alta})$

$Gini_{Dívida}(D) = (7/14) * [1 - (3/7)^2 - (2/7)^2 - (2/7)^2] + (7/14) * [1 - (2/7)^2 - (1/7)^2 - (4/7)^2] = 0.61224$

$Gini_{Garantia}(D) = (11/14) * Gini(D_{Garantia=nenhuma}) + (3/14) * Gini(D_{Garantia=adequada})$

$Gini_{Garantia}(D) = (11/14) * [1 - (4/11)^2 - (3/11)^2 - (6/11)^2] + (3/14) * [1 - (1/3)^2 - (2/3)^2 - (0/3)^2] = 0.48484$

Para o atributo Renda:

$Gini_{Renda}(D) = (5/14) * Gini(D_{Renda=\$0-15k}) + (3/14) * Gini(D_{Renda=\$15k-35k}) + (6/14) * Gini(D_{Renda=\$35k-})$
 $Gini_{Renda}(D) = (5/14) * [1 - (5/5)^2] + (3/14) * [1 - (2/3)^2 - (1/3)^2] + (6/14) * [1 - (5/6)^2 - (1/6)^2] = 0.214528$

Passo 3: Escolher como nó raiz o atributo que apresentou o menor Índice Gini: Renda

No caso da classe \$0 a \$15k, todas as opções apontam para Risco Alto. Renda: \$15k a \$35k

	História de Crédito	Dívida	Garantia	Renda	Risco
E3	Desconhecida	Baixa	Nenhuma	\$15 a \$35k	Moderado
E12	Boa	Alta	Nenhuma	\$15 a \$35k	Moderado
E14	Ruim	Alta	Nenhuma	\$15 a \$35k	Alto

$$Gini\ História\ (D) = (1/3) * Gini(D\ História=Desconhecida) + (1/3) * Gini(D\ História=Ruim) + (1/3) * Gini(D\ História=Boa)$$

$$Gini\ História\ (D) = (1/3) * [1 - (1)^2] + (1/3) * (1 - 1^2) + (1/3) * (1 - 1^2) = 0$$

$$Gini\ Dívida\ (D) = (1/3) * Gini(D\ Dívida=baixa) + (2/3) * Gini(D\ Dívida=Alta) \quad Gini\ Dívida\ (D) = (1/3) * [1 - 1^2] + (2/3) * [1 - (1/3)^2 - (1/3)^2] = 0.51851$$

$$Gini\ Garantia\ (D) = (3/3) * Gini(D\ Garantia=nenhuma)$$

$$Gini\ Garantia\ (D) = 1 * [1 - (2/3)^2 - (1/3)^2] = 0.44444$$

O próximo nó deve ser, portanto, *História*, cujos filhos estão trivialmente dados na tabela.

Renda: Acima de \$35k

	História de Crédito	Dívida	Garantia	Renda	Risco
E5	Desconhecida	Baixa	Nenhuma	Acima de \$35k	Baixo
E6	Desconhecida	Baixa	Adequada	Acima de \$35k	Baixo
E8	Ruim	Baixa	Adequada	Acima de \$35k	Moderado
E9	Boa	Baixa	Nenhuma	Acima de \$35k	Baixo
E10	Boa	Alta	Adequada	Acima de \$35k	Baixo
E13	Boa	Alta	Nenhuma	Acima de \$35k	Baixo

$$Gini\ História\ (D) = (2/6) * Gini(D\ História=Desconhecida) + (1/6) *$$

$$Gini(D\ História=Ruim) + (3/6) * Gini(D\ História=Boa)$$

$$Gini\ História\ (D) = (2/6) * [1 - (1)^2] + (1/6) * (1 - 1^2) + (3/3) * (1 - 1^2) = 0$$

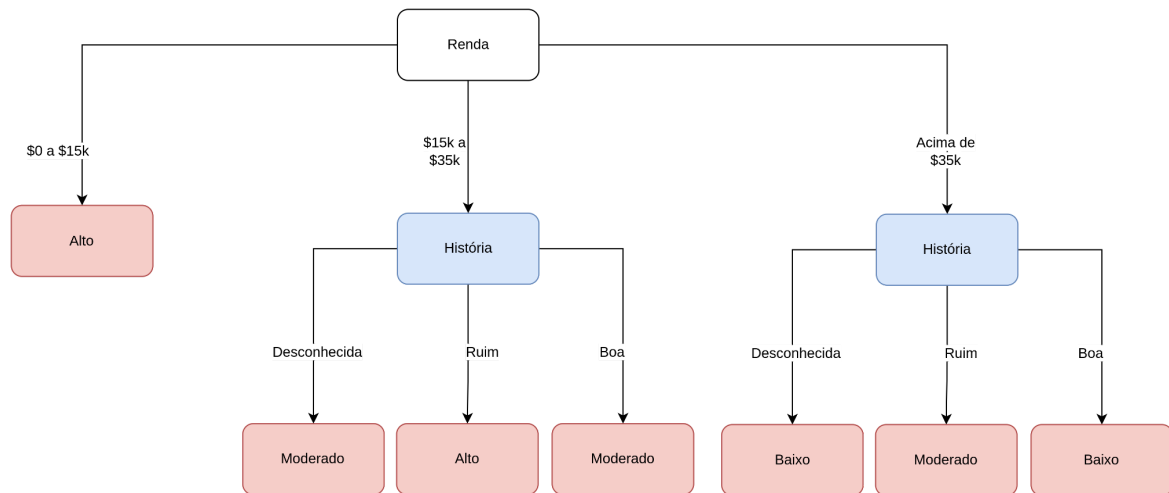
$$Gini\ Dívida\ (D) = (4/6) * Gini(D\ Dívida=baixa) + (2/6) * Gini(D\ Dívida=Alta)$$

$$Gini\ Dívida\ (D) = (4/6) * [1 - (3/4)^2 - (1/4)^2] + (2/6) * [1 - 1^2] = 0.25$$

$$Gini\ Garantia\ (D) = (3/6) * Gini(D\ Garantia=nenhuma) + (3/6) * Gini(D\ Garantia=Adequada)$$

$$Gini\ Garantia\ (D) = 1 * [1 - 1^2] + (3/6) * [1 - (2/3)^2 - (1/3)^2] = 0.22222$$

O próximo nó deve ser História, cujos filhos estão trivialmente dados na tabela. Portanto, esta deve ser a Árvore de Decisão formada pelo algoritmo Cart:



```
Renda=(Acima de $35k): Baixo(5.0/1.0)  
Renda!=(Acima de $35k): Alto(6.0/2.0)
```

Number of Leaf Nodes: 2

Size of the Tree: 3

```
=== Classifier model (full training set) ===
```

```
Prism rules
```

```
-----
```

```
If come = pouco then magro  
If vegetariano = sim then magro  
If come = muito then gordo  
If come = médio  
    and vegetariano = não then gordo
```

mo Prism:

Regras de classificação utilizando o algoritmo OneR:

```
=== Classifier model (full training set) ===
```

```
come:
```

```
    pouco    -> magro
```

```
    médio    -> gordo
```

```
    muito    -> gordo
```

```
(7/8 instances correct)
```

Árvore de decisão utilizando o algoritmo ID3:

```
Id3
```

```
come = pouco: magro
```

```
come = médio
```

```
| vegetariano = sim: magro
```

```
| vegetariano = não: gordo
```

```
come = muito: gordo
```

A partir da árvore, podem-se extrair as seguintes regras:

1. IF COME=POUCO THEN CLASSE=MAGRO
2. IF COME=MÉDIO AND VEGETARIANO=SIM THEN CLASSE=MAGRO
3. IF COME=MÉDIO AND VEGETARIANO=NÃO THEN
CLASSE=GORDO
4. IF COME=MUITO THEN CLASSE=GORDO

As regras 1, 3 e 4 também se encontram na classificação do algoritmo Prism. Árvore de decisão utilizando o algoritmo C4.5:

J48 pruned tree

vegetariano = sim: magro (3.0)

vegetariano = não: gordo (5.0/1.0)

Number of Leaves : 2

Size of the tree : 3

A partir da árvore, podem-se extrair as seguintes regras:

- SE VEGETARIANO=SIM THEN CLASSE=MAGRO
- SE VEGETARIANO=NÃO THEN CLASSE=GORDO

Questão 3:

Encadeamento para frente:

```
class InferenceEngineF:
    def __init__(self):
        self.facts = {}

    def add_fact(self, fact, value):
        self.facts[fact] = value

    def evaluate_risk(self):
        income = self.facts.get('income', 0)
        credit_history = self.facts.get('credit_history', 'Unknown')
        debt = self.facts.get('debt', '2')
        warranty = self.facts.get('warranty', 'Proper')
        if 0 <= income <= 15000:
            self.facts['risk'] = 'ALTO'
        elif 15000 < income <= 35000:
            if credit_history == 1:
                self.facts['risk'] = 'MODERADO'
            elif credit_history == 2:
                self.facts['risk'] = 'ALTO'
            elif credit_history == 3:
                if debt == 1:
                    self.facts['risk'] = 'ALTO'
                elif debt == 2:
                    self.facts['risk'] = 'MODERADO'
        elif income > 35000:
            if warranty == 1 or warranty == 2:
                self.facts['risk'] = 'BAIXO'
            else:
                self.facts['risk'] = 'DESCONHECIDO'
        else:
            self.facts['risk'] = 'DESCONHECIDO'

engine = InferenceEngineF()

engine.add_fact('income', int(input('Insira a Renda: ')))
engine.add_fact('credit_history', int(input('\nQual a Histórico de Crédito: \n1: Bom\n2: Ruim\n3: Desconhecido\n\n')))
engine.add_fact('debt', int(input('\nQual a Divida: \n1: Alta\n2: Baixa\n\n')))
engine.add_fact('warranty', int(input('\nQual a Garantia: \n1: Nenhuma\n2: Adequada\n\n')))
engine.evaluate_risk()

print('Resultado da avaliação de risco:', engine.facts.get('risk'))
```

Resultado:

```
Windows PowerShell
PS C:\Users\paulo\Área de Trabalho> python3 .\questao3.py
Insira a Renda: 32500

Qual a Histórico de Crédito:
1: Bom
2: Ruim
3: Desconhecido

3

Qual a Dívida:
1: Alta
2: Baixa

2

Qual a Garantia:
1: Nenhuma
2: Adequada

2
Resultado da avaliação de risco: MODERADO
```

Encadeamento para trás:

```
def inference_engine(income, credit_history, debt, warranty, risk):
    if 0 <= income <= 15000:
        if risk == 1:
            return True
        else:
            return False
    elif 15000 < income <= 35000:
        if credit_history == 1:
            if risk == 3:
                return True
            else:
                return False
        elif credit_history == 2:
            if risk == 1:
                return True
            else:
                return False
        elif credit_history == 3:
            if debt == 1:
                if risk == 1:
                    return True
                else:
                    return False
            elif debt == 2:
                if risk == 3:
                    return True
                else:
                    return False
    elif income > 35000:
        if warranty == 1 or warranty == 2:
            if risk == 2:
                return True
            else:
                return False
    elif income == None:
        if risk == 1:
            if credit_history == 3:
                if debt == 1:
                    if warranty == None:
                        return print('Renda entre 0 a 35k')

    print('Dados inválidos')
    return False
```

```

def main():
    risk = int(input('Informe o risco: \n1 - Alto\n2 - Baixo\n3 - Moderado\n4 - Desconhecido\n\n'))
    income = int(input('Insira a Renda: '))
    credit_history = int(input('\nQual o Histórico de Crédito: \n1 - Bom\n2 - Ruim\n3 - Desconhecido\n\n'))
    debt = int(input('\nQual a Dívida: \n1 - Alta\n2 - Baixa\n3 - Desconhecida\n\n'))
    warranty = int(input('\nQual a Garantia: \n1: Nenhuma\n2: Adequada\n3 - Desconhecida\n\n'))

    if risk not in [1, 2, 3]:
        risk = None
    if credit_history not in [1, 2, 3]:
        credit_history = None
    if debt not in [1, 2]:
        debt = None
    if warranty not in [1, 2]:
        warranty = None

    if inference_engine(income, credit_history, debt, warranty, risk):
        print('Dados consistentes')
    else:
        print('Dados inconsistentes')

main()

```

Resultado:

```

Windows PowerShell
PS C:\Users\paulo\Área de Trabalho> python3 .\questao3b.py
Informe o risco:
1 - Alto
2 - Baixo
3 - Moderado
4 - Desconhecido

1
Insira a Renda: 1500

Qual o Histórico de Crédito:
1 - Bom
2 - Ruim
3 - Desconhecido

3

Qual a Dívida:
1 - Alta
2 - Baixa
3 - Desconhecida

1

Qual a Garantia:
1 - Nenhuma
2 - Adequada
3 - Desconhecida

3
Dados consistentes

```


Questão 4:

```
class InferenceEngineF:
    def __init__(self):
        self.facts = {}

    def add_fact(self, fact, value):
        self.facts[fact] = value

    def evaluate_risk(self):
        renda = self.facts.get('renda', 0)
        historico_credito = self.facts.get('historico_credito', 'Desconhecida')
        divida = self.facts.get('divida', 'Baixa')

        if 0 <= renda <= 15000:
            self.facts['risco'] = 'ALTO'
        elif 15000 < renda <= 35000:
            if historico_credito == 1:
                self.facts['risco'] = 'MODERADO'
            elif historico_credito == 2:
                self.facts['risco'] = 'ALTO'
            elif historico_credito == 3:
                if divida == 1:
                    self.facts['risco'] = 'ALTO'
                elif divida == 2:
                    self.facts['risco'] = 'MODERADO'
        elif renda > 35000:
            self.facts['risco'] = 'BAIXO'
        else:
            self.facts['risco'] = 'DESCONHECIDO'
```

```
class InferenceEngineI:
    def __init__(self):
        self.facts = {}

    def add_fact(self, fact, value):
        self.facts[fact] = value

    def infer_from_risk(self, risco):
        if risco == 1:
            self.facts['renda'] = 'Renda entre 0 a 35k'
            self.facts['historico_credito'] = 'Ruim ou Desconhecida'
            self.facts['divida'] = 'Alta ou Desconhecida'
        elif risco == 2:
            self.facts['renda'] = 'Renda entre 15k a 35k'
            self.facts['historico_credito'] = 'Boa ou Desconhecido'
            self.facts['divida'] = 'Baixa ou Desconhecida'
        elif risco == 3:
            self.facts['renda'] = 'Renda maior que 35k'
        else:
            self.facts['renda'] = None
            self.facts['historico_credito'] = None
            self.facts['divida'] = None
```

```

inferencia = int(input('Selecione uma opção: \n1: Encadeamento para Frente\n2: Encadeamento para Trás\n\n'))

if inferencia == 1:
    engine = InferenceEngineF()
    engine.add_fact('renda', int(input('Qual a Renda: ')))
    engine.add_fact('historico_credito', int(input('\nQual a Histórico de Crédito: \n1: Bom\n2: Ruim\n3: Desconhecido\n\n')))
    engine.add_fact('divida', int(input('\nQual a Dívida: \n1: Alta\n2: Baixa\n\n')))

    engine.evaluate_risk()

    print("\nResultado da avaliação de risco:")
    print(f"Risco: {engine.facts.get('risco')}")

elif inferencia == 2:
    engine = InferenceEngineI()
    engine.infer_from_risk(int(input('\nRisco: \n1: Alto\n2: Moderado\n3: Baixo\n\n')))

    print("\nInferência a partir do Risco:")
    print(f"Renda: {engine.facts.get('renda')}")
    print(f"História de Crédito: {engine.facts.get('historico_credito')}")
    print(f"Dívida: {engine.facts.get('divida')}")

```

Resultados:

```

Windows PowerShell
PS C:\Users\paulo\Área de Trabalho> python3 .\questao4.py
Selecione uma opção:
1: Encadeamento para Frente
2: Encadeamento para Trás

1
Qual a Renda: 35000

Qual a Histórico de Crédito:
1: Bom
2: Ruim
3: Desconhecido

1

Qual a Dívida:
1: Alta
2: Baixa

2

Resultado da avaliação de risco:
Risco: MODERADO

```

```

Windows PowerShell
PS C:\Users\paulo\Área de Trabalho> python3 .\questao4.py
Selecione uma opção:
1: Encadeamento para Frente
2: Encadeamento para Trás

2

Risco:
1: Alto
2: Moderado
3: Baixo

1

Inferência a partir do Risco:
Renda: Renda entre 0 a 35k
História de Crédito: Ruim ou Desconhecida
Dívida: Alta ou Desconhecida

```

Questão 5:

Para essa questão, foi usada a ferramenta Expert Sinta com as seguintes regras:

The image shows a software interface for an expert system. At the top, a window titled 'Resultados' is partially visible. Below it, a dialog box titled 'SEM NOME' contains the question 'Qual a renda?' with the instruction '(Marque somente uma alternativa)'. The dialog has two columns: 'Opção:' and 'Grau de Confiança %:'. Three options are listed: '0 a 15k', '15 a 35k' (which is selected with a checkmark), and 'mais de 35k'. The confidence level for the selected option is '100'. At the bottom of the dialog are 'OK' and 'Por que?' buttons. Below the dialog, a list of rules is displayed:

REGRA	SE	ENTÃO
5	Renda = 15 a 35k	
	Historico_de_credito = Desconhecido	
	Dívida = Baixa	
		Risco = Moderado CNF 100%
6	Renda = mais de 35k	
	Garantia = Adequada	
	Renda = mais de 35k	
	Garantia = Nenhuma	
		Risco = Baixo CNF 100%

At the bottom of the interface, there are tabs: 'Resultados', 'Histórico', 'Todos os valores', and 'O sistema'.

Exemplo de uso 1:

SEM NOME

Qual o valor de Historico_credito ?
(Marque somente uma alternativa)

Opção: _____ Grau de Confiança %:

☒ Bom 100

☐ Ruim

☐ Desconhecido

OK Por que?

SEM NOME

Qual o valor de Divida ?
(Marque somente uma alternativa)

Opção: _____ Grau de Confiança %:

☐ Alta

☒ Baixa 100

OK Por que?

Resultados

Risco

Valor	CNF (%)
Moderado	100

Fechar Ajuda

Resultados Histórico Todos os valores O sistema

Exemplo de uso 2:

SEM NOME

Qual a renda?
(Marque somente uma alternativa)

Opção: _____ Grau de Confiança %:

<input checked="" type="checkbox"/> 0 a 15k	100
<input type="checkbox"/> 15 a 35k	
<input type="checkbox"/> mais de 35k	

OK Por que?

SEM NOME

Qual o valor de Historico_credito ?
(Marque somente uma alternativa)

Opção: _____ Grau de Confiança %:

<input type="checkbox"/> Bom	
<input checked="" type="checkbox"/> Ruim	100
<input type="checkbox"/> Desconhecido	

OK Por que?

SEM NOME

Qual o valor de Divida ?
(Marque somente uma alternativa)

Opção: _____ Grau de Confiança %:

<input checked="" type="checkbox"/> Alta	100
<input type="checkbox"/> Baixa	

OK Por que?

