

Lista 1 - IA

Sophia Carrazza Ventorim de Sousa

PUC Minas - 2024

1-a) Dados de Restaurante.csv tabelados:

Exemplo	Alternativo	Bar	Sex/Sáb	Fome	Cliente	Preço	Chuva	Res	Tipo	Tempo	Conclusao
X1	Sim	Não	Não	Sim	Alguns	RRR	Não	Sim	Francês	0-10	Sim
X2	Sim	Não	Não	Sim	Cheio	R	Não	Não	Tailandês	30-60	Não
X3	Não	Sim	Não	Não	Alguns	R	Não	Não	Hambúrguer	0-10	Sim
X4	Sim	Não	Sim	Sim	Cheio	R	Sim	Não	Tailandês	30-60	Sim
X5	Sim	Não	Sim	Não	Cheio	RRR	Não	Sim	Francês	>60	Não
X6	Não	Sim	Não	Sim	Alguns	RR	Sim	Sim	Italiano	0-10	Sim
X7	Não	Sim	Não	Não	Nenhum	R	Sim	Não	Hambúrguer	0-10	Não
X8	Não	Não	Não	Sim	Alguns	RR	Sim	Sim	Tailandês	0-10	Sim
X9	Não	Sim	Sim	Não	Cheio	R	Sim	Não	Hambúrguer	>60	Não
X10	Sim	Sim	Sim	Sim	Cheio	RRR	Não	Sim	Italiano	30-60	Não
X11	Não	Não	Não	Não	Nenhum	R	Não	Não	Tailandês	0-10	Não
X12	Sim	Sim	Sim	Sim	Cheio	R	Não	Não	Hambúrguer	30-60	Sim

Resultados dos ganhos de informação:

Alternativo: 0

Bar: 0

Sex/Sáb: 0,022

Fome: 0,022

Preço: 0,197

Chuva: 0,022

Res: 0,022

Tipo: 0

Tempo: 0,208

Raíz da árvore: Cliente (maior ganho de informação)

Todos os cálculos estão no final do pdf

1-b) Tabela filtrada só com Cliente == Cheio para o segundo nível da árvore:

Exemplo	Alternativo	Bar	Sex/Sáb	Fome	Cliente	Preço	Chuva	Res	Tipo	Tempo	Conclusao
X2	Sim	Não	Não	Sim	Cheio	R	Não	Não	Tailandês	30-60	Não
X4	Sim	Não	Sim	Sim	Cheio	R	Sim	Não	Tailandês	30-60	Sim
X5	Sim	Não	Sim	Não	Cheio	RRR	Não	Sim	Francês	>60	Não
X9	Não	Sim	Sim	Não	Cheio	R	Sim	Não	Hambúrguer	>60	Não
X10	Sim	Sim	Sim	Sim	Cheio	RRR	Não	Sim	Italiano	30-60	Não
X12	Sim	Sim	Sim	Sim	Cheio	R	Não	Não	Hambúrguer	30-60	Sim

Resultados dos ganhos de informação para o segundo nível da árvore:

Alternativo: 0,118

Bar: 0

Sex/Sáb: 0,118

Fome: 0,318

Preço: 0,318

Chuva: 0,44

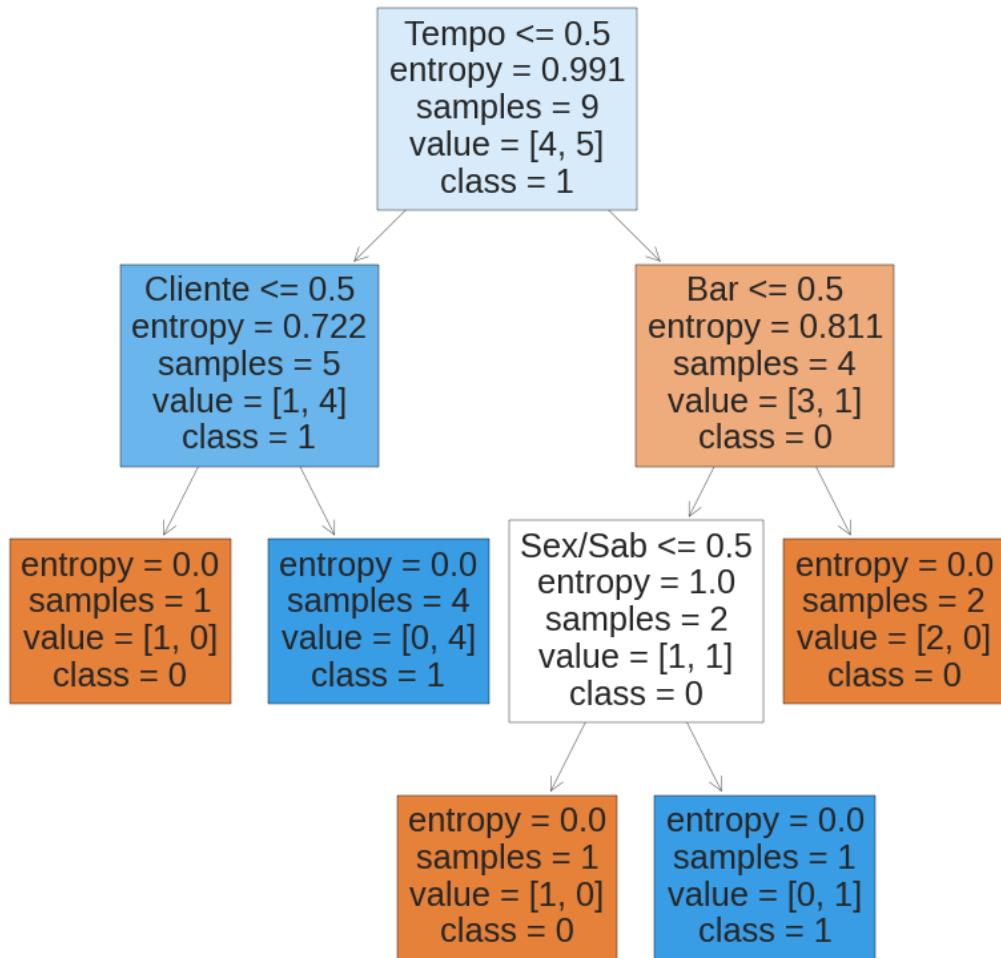
Res: 0,318

Tipo: 0,318

Tempo: 0,208

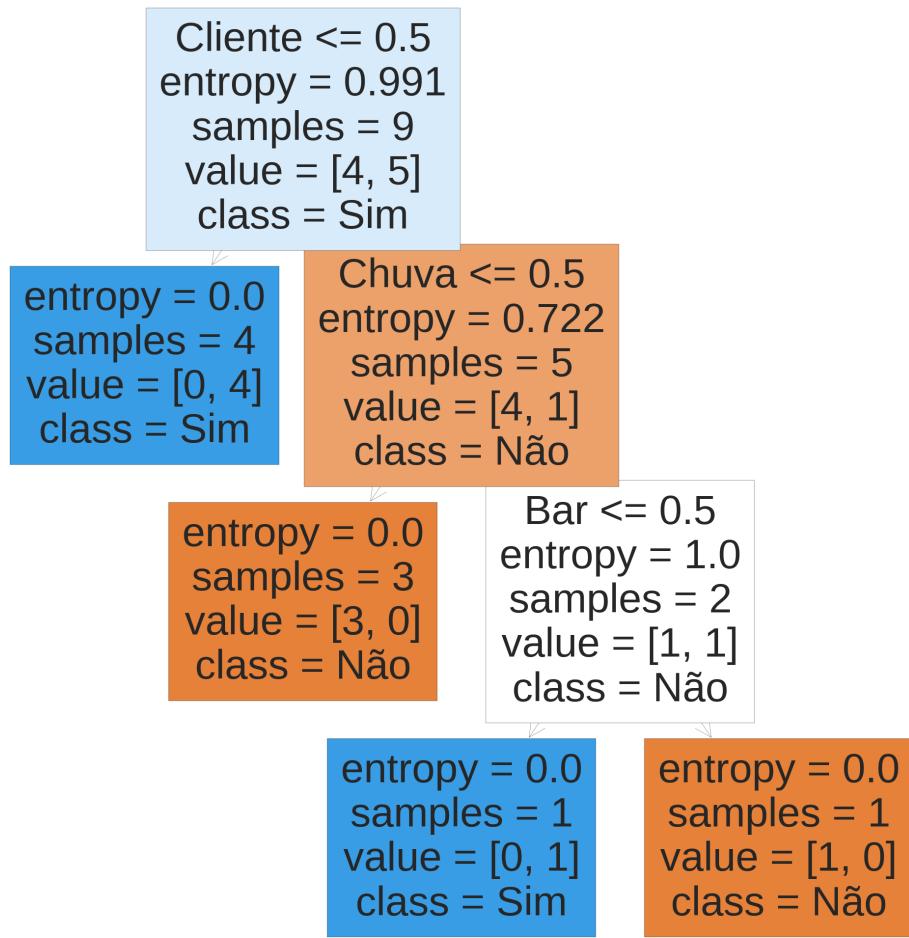
Raíz da árvore: Chuva (maior ganho de informação)

2- a) Árvore de decisão obtida com a codificação atual:



$\text{precisão} \cong 0,33$

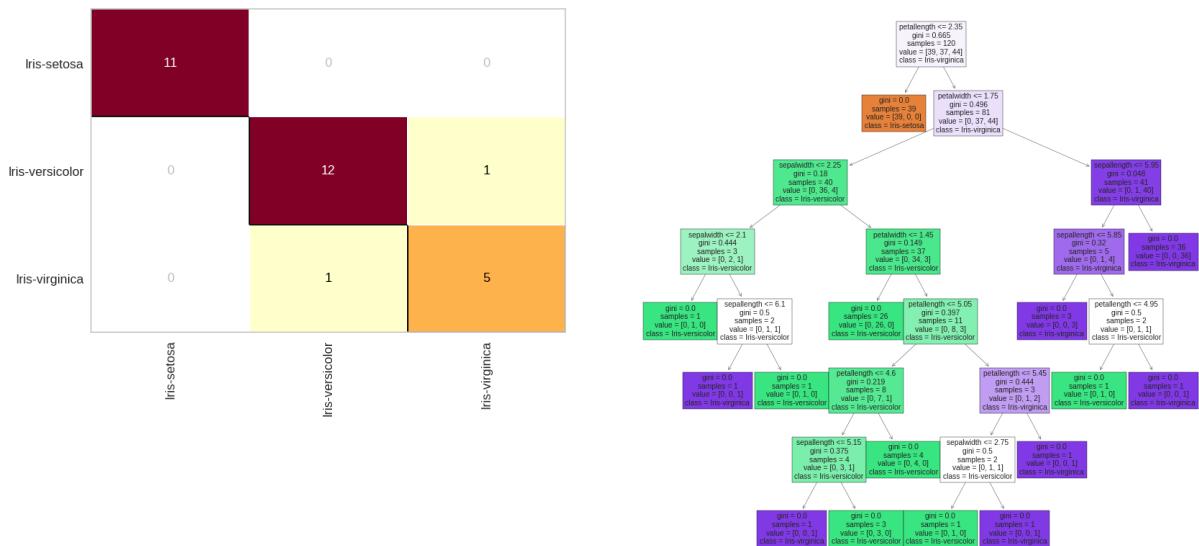
2- b) Árvore de decisão após a binarização dos atributos (preço, cliente e tempo de espera):



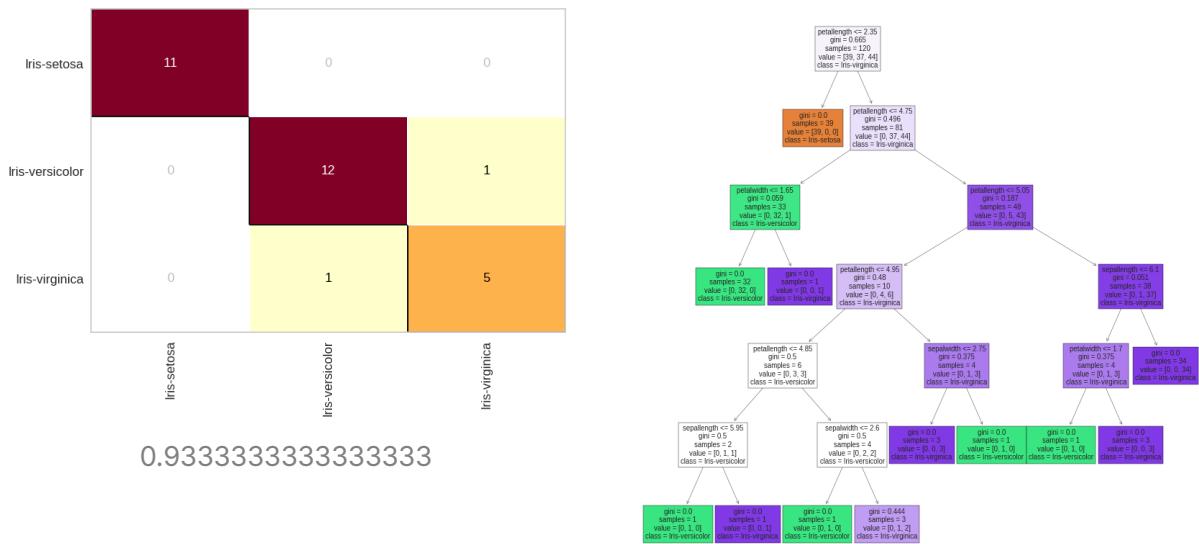
precisão $\cong 0,67$

Observação: Há mudança na ordem das classes e, principalmente, uma mudança significativa de aumento de ganho e precisão da primeira para a segunda árvore.

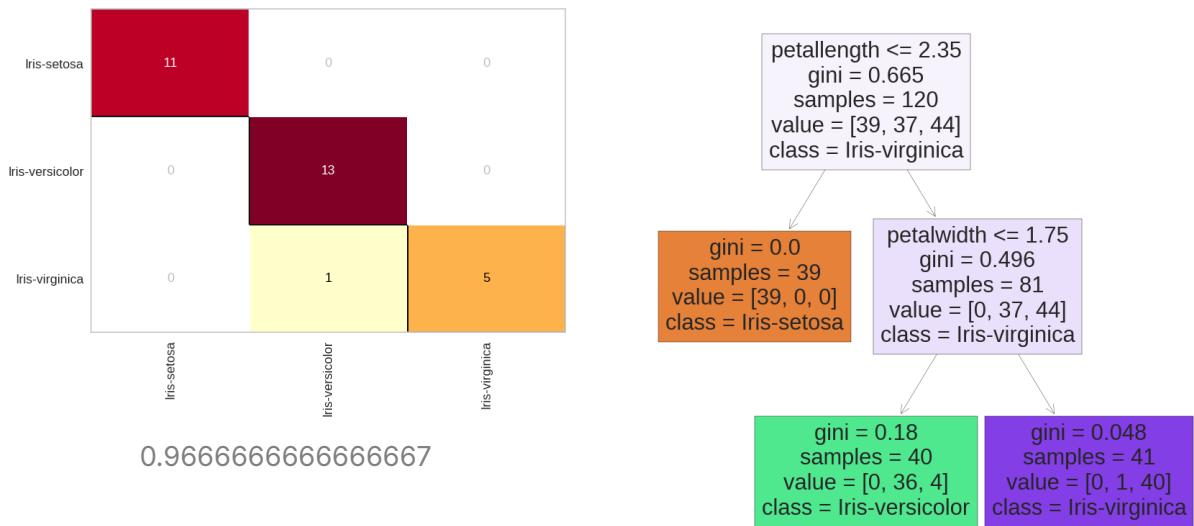
2-c) Antes de ajustar os hiperparâmetros:



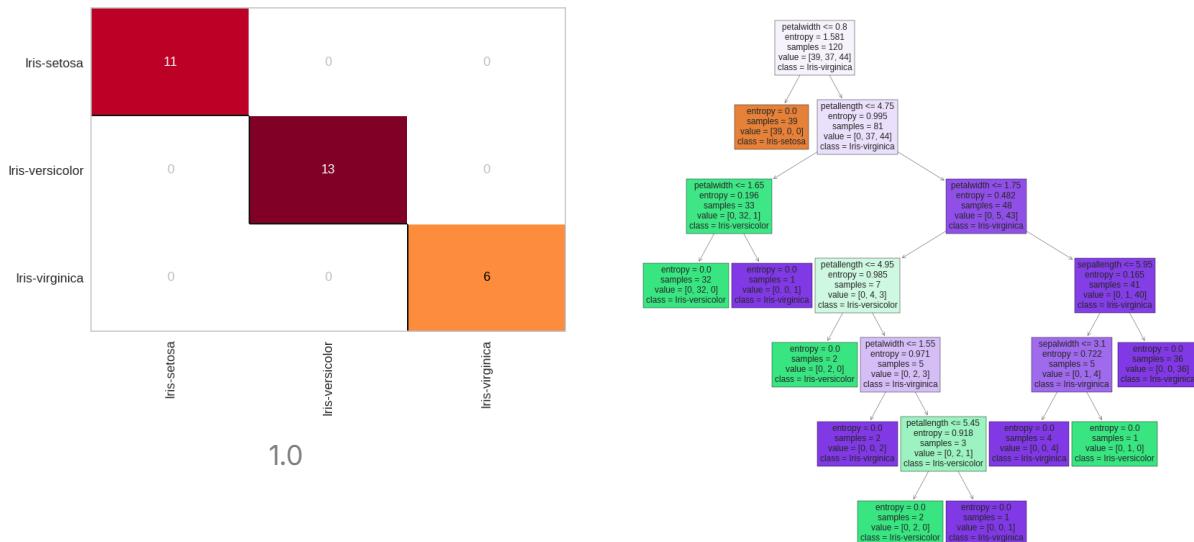
Com `max_depth=None`, `criterion=gini`, e `max_features=0.2`:



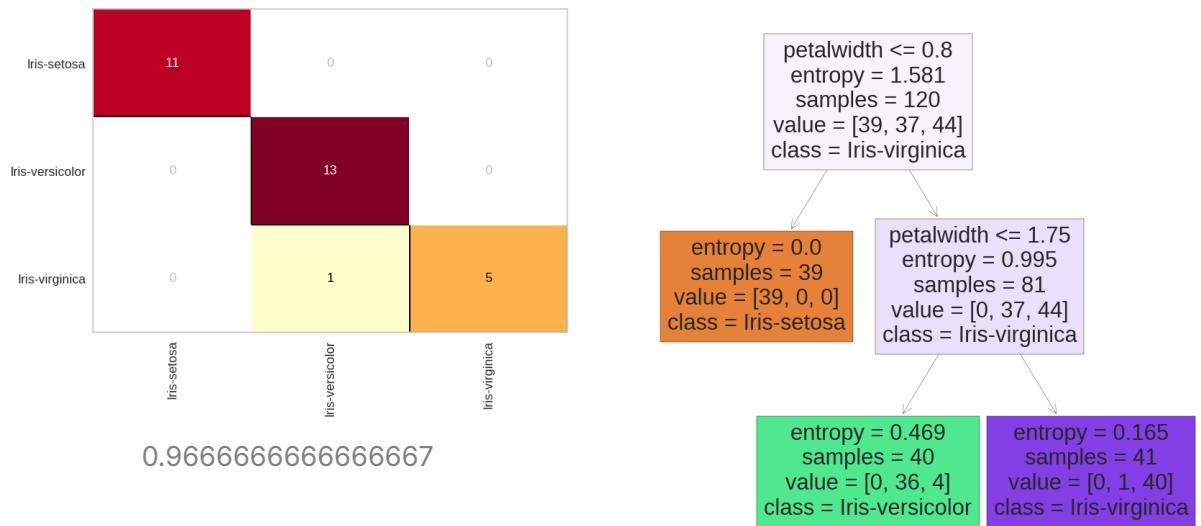
Com `max_depth=2`, `criterion=gini`, e `max_features=0.83`



Com `max_depth=6`, `criterion=entropy`, e `max_features=0.8` (o melhor resultado, com 1.0 de acurácia):



Explorando outros hiperparâmetros → `max_leaf_nodes=4` e `min_samples_split=5`:



3-a) As diferenças entre os algoritmos de árvore ID3 e C4.5 estão no tipo dos atributos, em que o ID3 utiliza somente atributos categóricos, enquanto o C4.5 converte os atributos numéricos em valores discretos para aumentar o ganho de informação. Além disso, ele também maneja dados desconhecidos estimando a probabilidade de diferentes resultados.

Outra diferença reside na inclusão do C4.5 de um processo que reduz o tamanho da árvore e evita overfitting, chamado de poda, o que não é feito pelo ID3. Por fim, há também a adição de razão de ganho, diferente do ganho de informação do ID3, já que o novo ajuda a evitar atributos com muitos valores diferentes.

3-b) O C4.5 gerencia os atributos numéricos examinando os valores de cada um, ordenando e identificando os melhores pontos de divisão que separarão melhor os dados para ter maior ganho de informação. Esses pontos criam intervalos que permitem o tratamento desses atributos como discretos.

Cálculos da questão 1-a) e 1-b):

Lista de IA-DL

coluna da
Conclusão
 n

$\rightarrow \{Sim, Não\}$
aparece 6 vez
vez, 6 vez

Sophia Calmeza
proporção:
 $Sim = \frac{6}{12} = 0,5$ $Não = \frac{6}{12} = 0,5$

$$1. \text{ } E(S) = - \sum_{i=1}^n p_i \log_2(p_i)$$

$p/Sim: -0,5 \cdot \underbrace{\log_2(0,5)}_{-2} = 0,5$

$p/Não: -0,5 \cdot \underbrace{\log_2(0,5)}_{-1} = 0,5$

$\rightarrow \text{SOMA dos entropios} = 0,5 + 0,5 = 1 //$

* A entropia da base é 1 //

Cliente: ALGUNS, CHEIO, NENHUM

$$* E(P/A) = \sum_{i=1}^n \frac{\text{tam. do subconjunto}}{\text{tam. total}} \circ H(\text{subconjunto})$$

$$\text{ALGUNS: 4 instâncias} \rightarrow Sim = 4/4 = 1 / Não = 0/4$$

$$\text{entropia} \rightarrow -1 \cdot \underbrace{\log_2(1)}_{0} = 0 \rightarrow 0 + 0 = 0 //$$

$$\text{CHEIO: 6 instâncias} \rightarrow Sim = 2/6 / Não = 4/6$$

$$\text{entropia} \rightarrow -\frac{2}{6} \log_2\left(\frac{2}{6}\right) / -\frac{4}{6} \log_2\left(\frac{4}{6}\right)$$

$$= \frac{1}{3} \log_2(3) = 0,528$$

$$= 0,528 + 0,38 = 0,918$$

$$\text{NENHUM: 2 instâncias} \rightarrow Sim = 0/2 / Não = 2/2$$

$$\text{entropia} \rightarrow 0$$

$$\rightarrow \frac{4}{12} \cdot 0 + \frac{6}{12} \cdot 0,918 + \frac{2}{12} \cdot 0 \Rightarrow 0,459 //$$

$$\text{ganho} = 1 - 0,459 = 0,541 //$$

Altuntróvo:

$$\text{SIM: 6 instâncias} \rightarrow \text{SIM: } \frac{3}{6} \quad \text{NÃO: } \frac{3}{6}$$

$$-\frac{3}{6}(0,5) \cdot \log_2\left(\frac{3}{6}\right) = 1$$

$$= 1 \quad \underbrace{1+1}_{1+1 \in \mathbb{P}} \rightarrow 1$$

$$\text{NÃO: 6 instâncias} \rightarrow \text{SIM: } \frac{3}{6} \quad \text{NÃO: } \frac{3}{6}$$

$$\text{entropia} \rightarrow 1$$

$$\text{entropia condicional: } \frac{6}{12} \cdot 1 + \frac{6}{12} \cdot 1 = 1$$

$$\text{ganho} = 1 - 1 = 0$$

Bar: SIM: 6 instâncias $\rightarrow \frac{3}{6}$ NÃO $\frac{3}{6}$

$$\text{NÃO: 6 instâncias} \rightarrow \text{SIM: } 1 \quad (\text{mesma coisa da anterior})$$

$$\text{ganho} = 0$$

Sex/Sib: SIM: 5 instâncias $\rightarrow \text{SIM: } \frac{2}{5}$ NÃO $\frac{3}{5}$

$$-\frac{2}{5} \cdot \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \cdot \log_2\left(\frac{3}{5}\right)$$

$$= 0,528 \quad \underbrace{0,97}_{0,97} = 0,492$$

$$\text{NÃO: 7 instâncias} \rightarrow \text{SIM: } \frac{4}{7} \quad \text{NÃO: } \frac{3}{7}$$

$$-\frac{4}{7} \cdot \log_2\left(\frac{4}{7}\right) - \frac{3}{7} \cdot \log_2\left(\frac{3}{7}\right)$$

$$= 0,461 \quad \underbrace{0,984}_{0,984} = 0,523$$

$$\frac{5}{12} \cdot 0,97 + \frac{7}{12} \cdot 0,984 = 0,978$$

tilibra $\text{ganho} = 1 - 0,978 = 0,022$

Fome: (número de sim/sab)

$$\text{Sim: } S_{\text{int}} \rightarrow \text{Sim} = \frac{2}{5} \quad \text{N}\tilde{\text{A}}\text{o} = \frac{3}{5}$$

$$\text{entropia} = 0,97 \quad (\text{a}^{\circ}\text{jig er colado antro})$$

$$\text{N}\tilde{\text{A}}\text{o: } T_{\text{int}} \rightarrow \text{Sim} = \frac{4}{7} \quad \text{N}\tilde{\text{A}}\text{o} = \frac{3}{7}$$

$$\text{entropia} = 0,984$$

$$\text{entropia condicional} = 0,078$$

$$\text{ganho} = 0,022 //$$

PREC: R: Tint $\rightarrow \text{Sim} = \frac{3}{7} \quad \text{N}\tilde{\text{A}}\text{o} = \frac{4}{7}$

$$\text{entropia} = 0,984$$

$$RR: 2_{\text{int}} \rightarrow \text{Sim} = \frac{2}{12} \quad \text{N}\tilde{\text{A}}\text{o} = 0$$

$$\text{entropia} = 0$$

$$RRR: 3_{\text{int}} \rightarrow \text{Sim} = \frac{1}{3} \quad \text{N}\tilde{\text{A}}\text{o} = \frac{2}{3} \Rightarrow -\frac{2}{3} \cdot \log_2 \left(\frac{2}{3}\right) = 0,389$$

$$-\frac{1}{3} \cdot \log_2 \left(\frac{1}{3}\right) = 0,528 \quad \text{entropia} = 0,917$$

$$\text{entro. card} = \frac{7}{12} \cdot 0,984 + \frac{2}{12} \cdot 0 + \frac{3}{12} \cdot 0,917 = 0,893 //$$

$$\text{ganho} = 1 - 0,893 = 0,107$$

CHUVA: sim: 5 $\rightarrow \text{Sim} = \frac{3}{5} \quad \text{N}\tilde{\text{A}}\text{o} = \frac{2}{5}$

$$0,97$$

$$\text{N}\tilde{\text{A}}\text{o: } 7 \rightarrow \text{Sim} = \frac{3}{7} \quad \text{N}\tilde{\text{A}}\text{o} = \frac{4}{7}$$

$$0,984$$

$$\text{entropia condicional} = 0,978$$

$$\text{ganho} = 0,022 //$$

Res: $SIM = 5 \rightarrow NIM = 3/5$, $\overbrace{N\tilde{AO} = 2/5}^{0,97}$

$N\tilde{AO} = 7 \rightarrow NIM = 3/7$, $\overbrace{N\tilde{AO} = 4/7}^{0,984}$, $\left. \begin{array}{l} ent = 0,978 \\ cond = 1 - 0,978 \end{array} \right\} (1 - 0,978)$

ganho = 0,022

Tipo: Francês = 2 $\rightarrow SIM = 1/2$, $\overbrace{N\tilde{AO} = 1/2}^{entropia = 2 = 1}$

Tailandês = 4 $\rightarrow SIM = 2/4$, $\overbrace{N\tilde{AO} = 2/4}^{entropia = 2 = 1}$

Italiano = 2 $\rightarrow SIM = 1/2$, $\overbrace{N\tilde{AO} = 1/2}^{entropia = 2 = 1}$

Hambúrguer = 4 $\rightarrow SIM = 2/4$, $\overbrace{N\tilde{AO} = 2/4}^{entropia = 2}$

entropia condicional = $\frac{2}{12} \cdot 1 + \frac{4}{12} \cdot 1 + \frac{2}{12} \cdot 2 + \frac{4}{12} \cdot 2 = 1$

ganho = $1 - 1 + 0$

Tempo: 0 - 10 = 6 $\rightarrow SIM = 4/6 / N\tilde{AO} = 2/6 \Rightarrow entropia = 0,918$

30 - 60 = 4 $\rightarrow SIM = 2/4 / N\tilde{AO} = 2/4 \Rightarrow entropia = 1$

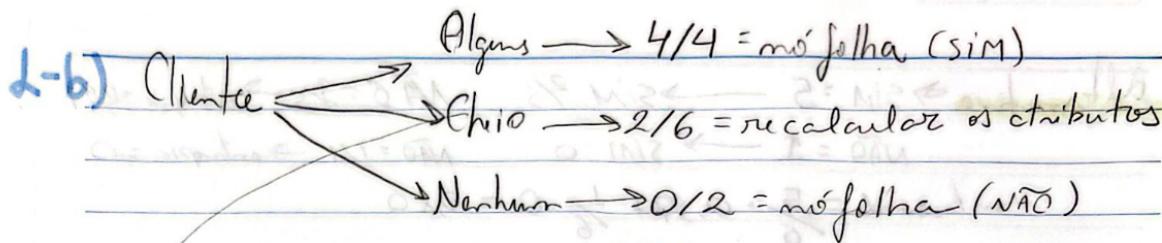
> 60 = 2 $\rightarrow SIM = 0 / N\tilde{AO} = 2/2 \Rightarrow entropia = 0$

entropia condicional = $\frac{6}{12} \cdot 0,918 + \frac{4}{12} \cdot 1 + \frac{2}{12} \cdot 0 = 0,792$

ganho = $1 - 0,846 = 0,154$

tilibra Raiz da árvore = Clínica (maior ganho = 0,541)

$$E(S) = \begin{cases} 2/6 = \text{sim} & \rightarrow \text{entropia} = 0,528 \\ 4/6 = \text{não} & \rightarrow \text{entropia} = 0,38 \end{cases} \quad \boxed{0,518}$$



Subconjunto Cliente == Chão \rightarrow 6 elementos (valores)

$$\text{Preço} \Rightarrow R=4 \rightarrow \text{SIM} = 2/4 \quad \text{NÃO} = 2/4$$

$\text{entropia} = -\frac{1}{2} \log_2 \frac{1}{2} = 1$

$$RRR = 2 \rightarrow \text{SIM} = 0 \quad \text{NÃO} = 2/2 \rightarrow \text{entropia} = 0$$

$$\text{entr. cond} = \frac{4}{6} \cdot 1 + \frac{2}{6} \cdot 0 = 0,6$$

$$\text{ganho} = 0,518 - 0,6 = \boxed{0,318}$$

$$\text{Tempo} \Rightarrow 30 \cdot 60 = 4 \rightarrow \text{SIM} = 2/4 \quad \text{NÃO} = 2/4 \Rightarrow \text{entropia} = 1$$

$$> 60 = 2 \rightarrow \text{SIM} = 0 \quad \text{NÃO} = 2/2 \rightarrow \text{entropia} = 0$$

$$\text{entr. condicional} = \frac{4}{6} \cdot 1 + \frac{2}{6} \cdot 0 \Rightarrow 0,6$$

$$\text{ganho} = \boxed{0,318}$$

$$\text{Rua} \Rightarrow \text{SIM} = 2 \rightarrow \text{SIM} = 0 \quad \text{NÃO} = 2/2 \rightarrow \text{entropia} = 0$$

$$\text{NÃO} = 4 \rightarrow 2/4 \text{ SIM} \quad 2/4 \text{ NÃO} \rightarrow \text{entropia} = 1$$

$$\text{entropia condicional} = \frac{2}{6} \cdot 0 + \frac{4}{6} \cdot 1 = 0,6 \Rightarrow \text{ganho} = \boxed{0,318}$$

$$\text{Chuva} \Rightarrow \text{SIM} = 2 \rightarrow \text{SIM} = 1/2 \quad \text{NÃO} = 1/2 \rightarrow \text{entropia} = 1$$

$$\text{NÃO} = 4 \rightarrow \text{SIM} = 1/4 \quad \text{NÃO} = 3/4$$

$$-\frac{1}{4} \cdot \log_2 \left(\frac{1}{4} \right) = \frac{1}{2} \quad -\frac{3}{4} \cdot \log_2 \left(\frac{3}{4} \right) = 0,311 \rightarrow 0,811$$

$$\text{entropia condicional} = \frac{2}{6} \cdot 1 + \frac{4}{6} \cdot 0,811 = 0,874$$



ganho = 0,947



$$\left(-\frac{1}{3} \cdot \log_2 \left(\frac{1}{3}\right) \right) + \left(-\frac{2}{3} \cdot \log_2 \left(\frac{2}{3}\right) \right) = 0,918$$

Alternativo $\Rightarrow SIM = 5 \rightarrow SIM = 2/5 \quad \bar{NAO} = 3/5 \rightarrow entropia = 0,97$
 $\bar{NAO} = 1 \rightarrow SIM = 0 \quad \bar{NAO} = 1/1 \rightarrow entropia = 0$
 $entropia_{cond} = \frac{5}{6} \cdot 0,97 + \frac{1}{6} \cdot 0 = 0,80$

Ganho $= 0,918 - 0,8 \Rightarrow 0,118 //$

Par $\Rightarrow SIM \Rightarrow 3 \rightarrow SIM = 1/3 \quad \bar{NAO} = 2/3 \rightarrow 0,918$
 $\bar{NAO} \Rightarrow 3 \rightarrow SIM = 1/3 \quad \bar{NAO} = 2/3 \rightarrow 0,918$
 $\frac{1}{6} \cdot 0,918 + \frac{1}{6} \cdot 0,918 = 0,918$
 $ganho = 0 // (0,918 - 0,918 = 0)$

Sexta/Sab $\Rightarrow SIM = 5 \rightarrow SIM = 2/5 \quad \bar{NAO} = 3/5 \rightarrow entropia = 0,97$
 $\bar{NAO} = 2 \rightarrow SIM = 0/1 \quad \bar{NAO} = 1/1 \rightarrow entropia = 0$
 $entropia_{cond} = 0,80 \rightarrow ganho = 0,118 //$

Fome $\Rightarrow SIM = 4 \rightarrow SIM = 2/4 \quad \bar{NAO} = 2/4 \rightarrow entropia = 1$
 $\bar{NAO} = 2 \rightarrow SIM = 0/2 \quad \bar{NAO} = 2/2 \rightarrow entropia = 0$
 $entropia_{cond} = \frac{4}{6} \cdot 1 + \frac{2}{6} \cdot 0 = 0,6$
 $ganho = 0,318 //$

Tipo $\Rightarrow Tailandês = 2 \rightarrow SIM = 1/2 \quad \bar{NAO} = 1/2 \rightarrow entropia = 1$
 Hambúrguer = 2 $\rightarrow SIM = 1/2 \quad \bar{NAO} = 1/2 \rightarrow entropia = 1$
 Francês = 1 $\rightarrow SIM = 0/1 \quad \bar{NAO} = 1/1 \rightarrow entropia = 0$
 Italiano = 1 $\rightarrow SIM = 0/1 \quad \bar{NAO} = 1/1 \rightarrow entropia = 0$
 $entropia_{cond} = \frac{2}{6} \cdot 1 + \frac{2}{6} \cdot 1 + \frac{1}{6} \cdot 0 + \frac{1}{6} \cdot 0 = 0,6$
 $ganho = 0,918 - 0,6 = 0,318 //$

Raiz da árvore no 2º nível = Chuva (0,44 de ganho)

tilibra