

LISTA 1 - Arthur Gonçalves de Moraes

1)

	D	S	T	W	T	F	S	S
1	●							
2		●						
3			●					
4				●				
5					●			
6						●		
7							●	
8								●
9								●
10								●
11								●
12								●
13								●
14								●
15								●
16								●
17								●
18								●
19								●
20								●
21								●
22								●
23								●
24								●
25								●
26								●
27								●
28								●
29								●
30								
31								

1.1)

Alternativo $\rightarrow I\left(\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1 \Rightarrow$ conclusão

$$\hookrightarrow \text{ganho} = 1 - \left[\frac{1}{2}I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}I\left(\frac{1}{2}, \frac{1}{2}\right) \right] = 0$$

Bar $\rightarrow I(1, 1)$

$$\hookrightarrow \text{ganho} = 1 - \left[\frac{1}{2}I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}I\left(\frac{1}{2}, \frac{1}{2}\right) \right] = 0$$

Sex/Sab

$$\hookrightarrow \text{ganho} = 1 - \left[\frac{7}{12}I\left(\frac{4}{7}, \frac{3}{7}\right) + \frac{5}{12}I\left(\frac{3}{5}, \frac{2}{5}\right) \right] \approx 0,02$$

Fome

$$\hookrightarrow \text{ganho} = 1 - \left[\frac{7}{12}I\left(\frac{2}{3}, \frac{2}{3}\right) + \frac{5}{12}I\left(\frac{4}{5}, \frac{1}{5}\right) \right] \approx 0,195$$

Clientes

$$\hookrightarrow \text{ganho} = 1 - \left[\frac{1}{3}I(1, 0) + \frac{1}{3}I(0, 1) + \frac{1}{2}I\left(\frac{1}{3}, \frac{2}{3}\right) \right] \approx 0,540$$

Prego \rightarrow ganho = $1 - \left[\frac{1}{4}I\left(\frac{1}{3}, \frac{2}{3}\right) + \frac{1}{6}I(1, 0) + \frac{7}{12}I\left(\frac{3}{7}, \frac{4}{7}\right) \right] \approx 0,195$

Chuva \rightarrow ganho = $1 - \left[\frac{7}{12}I\left(\frac{3}{5}, \frac{2}{5}\right) + \frac{7}{12}I\left(\frac{3}{7}, \frac{4}{7}\right) \right] \approx 0,020$

Rcs \rightarrow ganho = $1 - \left[\frac{5}{12}I\left(\frac{2}{3}, \frac{2}{3}\right) + \frac{7}{12}I\left(\frac{3}{7}, \frac{4}{7}\right) \right] \approx 0,020$

Tipo \rightarrow ganho = $1 - \left[\frac{1}{6}I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{3}I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{3}I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{6}I\left(\frac{1}{2}, \frac{1}{2}\right) \right] = 0$

Tempe \rightarrow ganho = $1 - \left[\frac{1}{2}I\left(\frac{2}{3}, \frac{1}{3}\right) + \frac{1}{6}I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{6}I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{6}I(0, 1) \right] \approx 0,207$

Raiz = CLIENTES

data
fecha

(D) (S) (T) (Q) (Q) (S) (S)
(D) (L) (M) (M) (J) (V) (S)

1.2)

Cliente

Algans Nenham chalo
↓ ↓ ↓
sim não fame

sim não

$$\text{chalo} = I\left(\frac{2}{3}, \frac{1}{3}\right) \approx 0,918$$

$$\text{alternative} \rightarrow 0,918 - \left[\frac{1}{2}\left(\frac{1}{3}, \frac{2}{3}\right) + \frac{1}{2}\left(\frac{1}{3}, \frac{2}{3}\right) \right] = 0$$

$$\text{Bar} \rightarrow 0,918 - \left[\frac{1}{2}\left(\frac{1}{3}, \frac{2}{3}\right) + \frac{1}{2}\left(\frac{1}{3}, \frac{2}{3}\right) \right] = 0$$

$$\text{Sex/Sab} \rightarrow 0,918 - \left[\frac{5}{6}\left(\frac{1}{3}, \frac{3}{5}\right) + \frac{1}{6}(0,1) \right] \approx 0,109$$

$$*\text{fome} \rightarrow 0,918 - \left[\frac{2}{3}\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{3}(0,1) \right] \approx 0,251$$

$$*\text{prego} \rightarrow 0,918 - \left[\frac{2}{3}\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{3}(0,1) \right] \approx 0,251$$

$$\text{chuva} \rightarrow 0,918 - \left[\frac{1}{3}\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{3}\left(\frac{1}{4}, \frac{3}{4}\right) \right] \approx 0,044$$

$$*\text{Res} \rightarrow 0,918 - \left[\frac{1}{3}(0,1) + \frac{2}{3}\left(\frac{1}{2}, \frac{1}{2}\right) \right] \approx 0,251$$

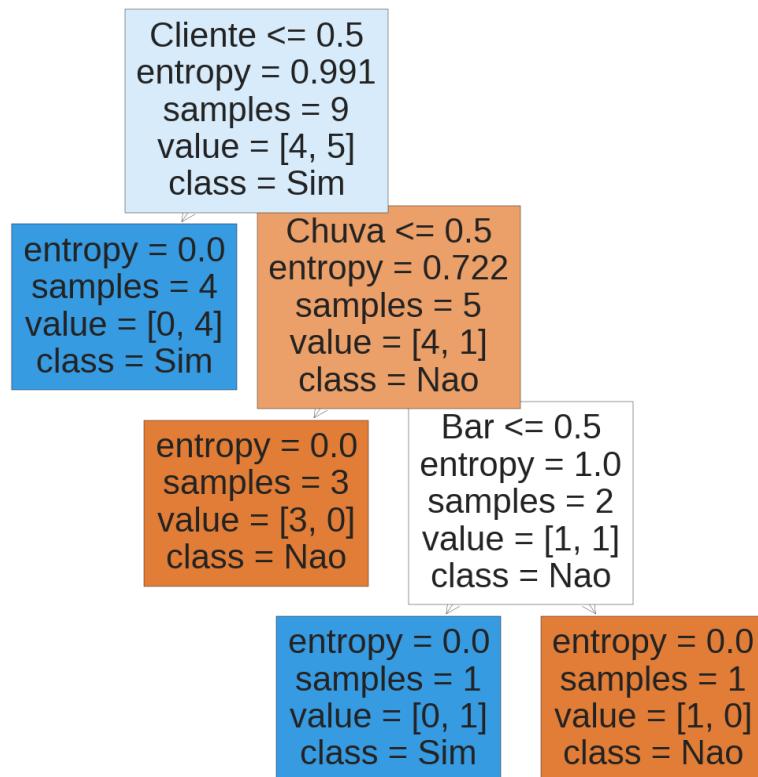
$$*\text{liso} \rightarrow 0,918 - \left[\frac{1}{3}\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{6}(0,1) + \frac{1}{3}\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{6}(0,1) \right] \approx 0,251$$

$$*\text{tempo} \rightarrow 0,918 - \left[\frac{1}{3}\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{3}\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{3}(0,1) \right] = 0,251$$

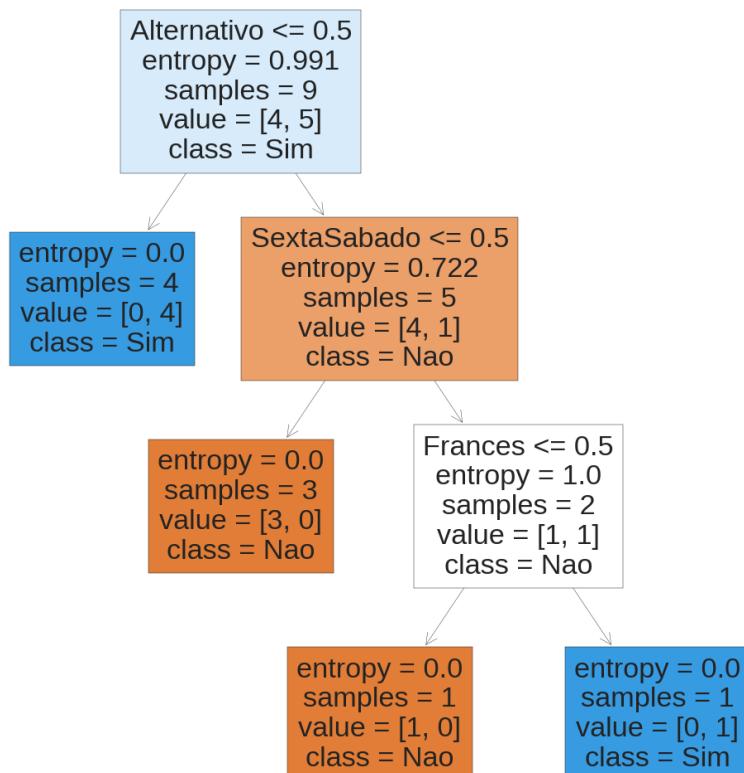
2.1)

2)

2.1)



2.2)



2.3) {'criterion': 'gini', 'max_depth': 6, 'max_features': 'sqrt'}

3)

3.1)

Diferença 1: ID3 não recebe dados numéricos, já o C4.5 recebe.

Diferença 2: ID3 não recebe dados ausentes. C4.5 recebe e estima o resultado com base em outros valores.

Diferença 3: ID3 não tem poda. C4.5 tem

Diferença 4: o ID3 tem dificuldade de lidar com dados nominais com muitas opções de resposta. O C4.5 resolve este com uma nova equação de ganho (razão de ganho)

3.2) O algoritmo C4.5 discretiza dados numéricos