

# Lista 01 – Inteligência Artificial

Nome: Júlio César Gonzaga Ferreira Silva

D S T Q U S S

IA – Inteligência Artificial

01) Lista:

01) Cálculo de ganho

$p_i$  = probabilidade  
 $p_i$  = Ex. positivos  
 $n_i$  = Ex. negativos

$$I(p_1, \dots, p_n) = \sum_{i=1}^n -P(v_i) \log_2(P(v_i))$$

Entropia de classe

$$\sum_{i=1}^n \frac{p_i + n_i}{p + n} I\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

Entropia de atributo

$$\text{ganho(atributo)} = \text{Entropia(classe)} - \text{Entropia(atributo)}$$

Alternativo:

$$\text{ganho(alternativo)} = 1 - \sum_{i=1}^n \frac{p_i + n_i}{p + n} I\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

$$E_0 = 1$$

$$\hookrightarrow -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right)$$

$$1 - \sum_{i=1}^n \frac{p_i + n_i}{p + n} \cdot I(1, 1)$$

$$G(\text{alt}) = 1 - \left(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1\right)$$

SIM NAO

X X

SNSNN/S

SSNSNN

$$G(\text{alt}) = 1 - \frac{2}{2} = \frac{2-2}{2} = 0$$

BAB

$$G(bab) = 1 - \sum_{i=1}^I E_p \cdot E_a$$

$$G(b) = 1 - \left(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1\right) = 0$$

 $E_p = 1$ 

S N

↓ ↓

SSNNNS SNSNSN

I(1)

Sex/SAB

$$G(\text{sex}/\text{SAB}) = 1 - \sum_{i=1}^I E_p \cdot E_a$$

$$E_a = -\left(\frac{2}{5} \cdot \log_2 \frac{2}{5} + \frac{3}{5} \cdot \log_2 \frac{3}{5}\right) \approx 0,97045$$

 $E_a$ S  
SMNNS

N

↓

SNSNSN

$$E_a = -\left(\frac{4}{7} \cdot \log_2 \frac{4}{7} + \frac{3}{7} \cdot \log_2 \frac{3}{7}\right) \approx 0,98522$$

 $\frac{7}{12}$ 

$$G(\text{sex}) = 1 - \left(\frac{5}{12} \cdot 0,97 + \frac{7}{12} \cdot 0,98\right) \approx 0,2073$$

Fome

$$G(\text{fome}) = 1 - \sum_{i=1}^I E_p \cdot E_a$$

S N

↓ ↓

SNSNSS SNWN

$$E_a = -\left(\frac{5}{7} \cdot \log_2 \frac{5}{7} + \frac{2}{7} \cdot \log_2 \frac{2}{7}\right) \approx 0,86312$$

$$E_a = -\left(\frac{1}{5} \cdot \log_2 \frac{1}{5} + \frac{4}{5} \cdot \log_2 \frac{4}{5}\right) \approx 0,72192$$

$$G(\text{fome}) = 1 - \left(\frac{7}{12} \cdot 0,86312 + \frac{5}{12} \cdot 0,72192\right) \approx$$

0,19591



cliente

$$G(\text{cliente}) = 1 - \sum E_g \cdot E_c$$

Atqun C Nen  
SSS NSNNNS NN

$$E_c = -\left(1 \log 1 + \frac{0}{3} \log \frac{0}{3}\right) \approx 0$$

$$E_{E_2} = -\left(\frac{1}{3} \log \frac{1}{3} + \frac{2}{3} \log \frac{2}{3}\right) \approx 0,91829$$

$$E_{E_3} = -\left(\frac{0}{2} \log \frac{0}{2} + 1 \log 1\right) \approx 0$$

$$G(l) = 1 - \left(\frac{1}{2} \cdot 0,91829\right) \approx 0,540855$$

prego

$$G(\text{prego}) = 1 - \sum E_g \cdot E_E$$

R RR RRR  
NSNNSSN SS SMN

$$E_{E_1} = -\left(\frac{3}{7} \log \frac{3}{7} + \frac{4}{7} \log \frac{4}{7}\right) \approx 0,98522$$

$$E_{E_2} = 0$$

$$E_{E_3} \approx 0,91829$$

$$G(p) = 1 - \left(\frac{7}{12} \cdot 0,98522 + \frac{1}{4} \cdot 0,91829\right) \approx$$

$$0,195715$$

Chuva

$$ganho(chuva) = 1 - \sum E_g \cdot E_F$$

S	N
NSSNS	SNSNNS

$$E_{F1} = -\left(\frac{3}{5} \log \frac{3}{5} + \frac{2}{5} \log \frac{2}{5}\right) \approx 0,97045$$

$$E_{F2} = -\left(\frac{3}{7} \log \frac{3}{7} + \frac{4}{7} \log \frac{4}{7}\right) \approx 0,98522$$

$$G(chuva) \approx 0,02073$$

Res

$$ganho(res) = 1 - \sum E_g \cdot E_F$$

S	N
SNSSN	NSSNNS

$$E_{F1} \approx 0,97045$$

$$G(res) = 0,02073$$

$$E_{F2} \approx 0,98522$$

Tiro

$$ganho(tiro) = 1 - \sum E_g \cdot E_F$$

F	T	H	A
SN	NSSN	SNBSN	
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

$$E_F = 0$$

$$G(t) = 0$$



Tempe

$$h_{\text{pinho}}(\text{Tempe}) = 1 - \sum_i E_{g_i} \cdot F_i$$

$$EE = 0,91829 \quad g(t) = 1 - \left( \frac{1}{6} + \frac{1}{6} + \frac{1}{2} \right) \cdot 0,91829 = 0,91829$$

$$g(t) = 1 - \left( \frac{1}{3} + \frac{1}{2} \cdot 0,91829 \right) = 0,2075$$

Conclua-se assim que o raio da árvore é o atributo Clientes.

b) Nova entropia de classe: 0,91829

NSNNS =  $\frac{3}{5}$   
N = 1

Alternativo:

$$h_{\text{pinho}}(\text{alternativo}): 0,91829 - \sum_i E_{g_i} \cdot EE_i$$

$$EE = - \left( \frac{2}{5} \log_2 \frac{2}{5} + \frac{3}{5} \log_2 \frac{3}{5} \right) = 0,97045$$

$$EE = 0$$

$$g(\text{alt}) = 0,91829 - \left( \frac{1}{2} \cdot 0,97045 \right)$$

$$g(\text{alt}) = 0,4581$$

Bar

$$f(\text{bar}) = 0,91829 - \sum E_g \cdot E_F$$

NS, N

NN, S

$$f(\text{bar}) = 0,91829 - \left( \frac{1}{2} \cdot 0,91829 + \frac{1}{2} \cdot 0,91829 \right)$$

$$E_F \approx 0,91829$$

$$E_F \approx 0,91829$$

$$f(b) = 0$$

sex / salo

$$f(\text{sex / salo}) = 0,91829 - \sum E_g \cdot E_F$$

S, N, N, N, S

$$f(\text{sex}) = 0,91829 - \left( \frac{5}{6} \cdot 0,97045 \right)$$

N

$$E_F \approx 0,97045$$

$$f(\text{sex}) = 0,109581$$

Fome

$$f(\text{fome}) = 0,91829 - \sum E_g \cdot E_F$$

NS, NS =  $\frac{1}{2}$ 

NN, S

$$f(\text{fome}) = 0,91829 - \left( \frac{1}{3} \cdot 1 \right) = 0,2516$$

Eucos

$$f(\text{precos}) = 0,91829 - \sum E_g \cdot E_F$$

NS, NS

$$f(\text{precos}) = 0,2516$$

NN

RES

$$f(\text{res}) = 0,91829 - \sum E_g \cdot E_F$$

NN

NS, NS

$$f(\text{res}) = 0,2516$$



Chuva

$$I_f(\text{Chuva}) = 0,91829 - \sum I_{Eg} \cdot EE$$

SN

NNNS

$$EE_1 = 1$$

$$EE_2 = -\left(\frac{1}{4} \log \frac{1}{4}\right) + \frac{3}{4} \log \frac{3}{4} \approx 0,81127$$

$$\frac{21}{23} \quad \frac{4}{6}$$

$$I_f(c) = 0,91829 - \left(\frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 0,81127\right)$$

$$I_f(c) \approx 0,04411$$

Tempo

$$I_f(\text{Tempo}) = 0,91829 - \sum I_{Eg} \cdot EE$$

NS

NS

$$I_f(\text{Tempo}) = 0,91829 - \left(\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 1\right)$$

N

N

$$I_f(t) \approx 0,25162$$

Tempo

$$I_f(\text{Tempo}) = 0,91829 - \sum I_{Eg} \cdot EE$$

NS

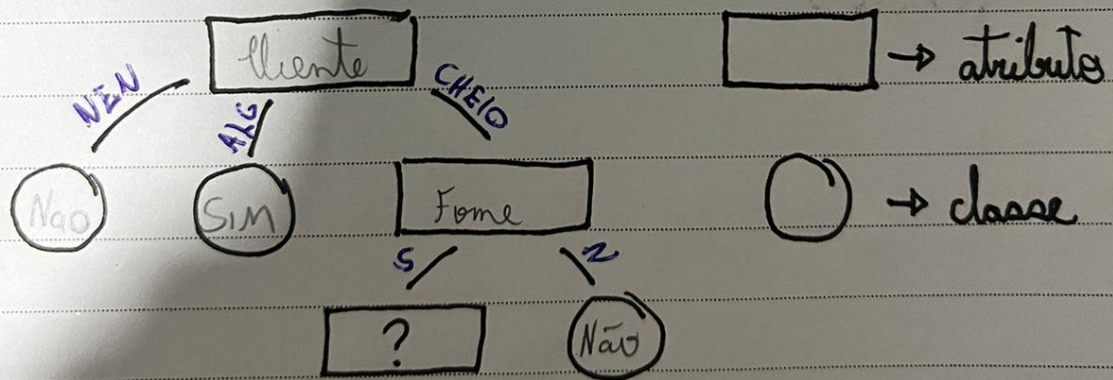
NN

$$I_f(t) = 0,25162$$

O atributo escolhido como segundo nível é a fome.

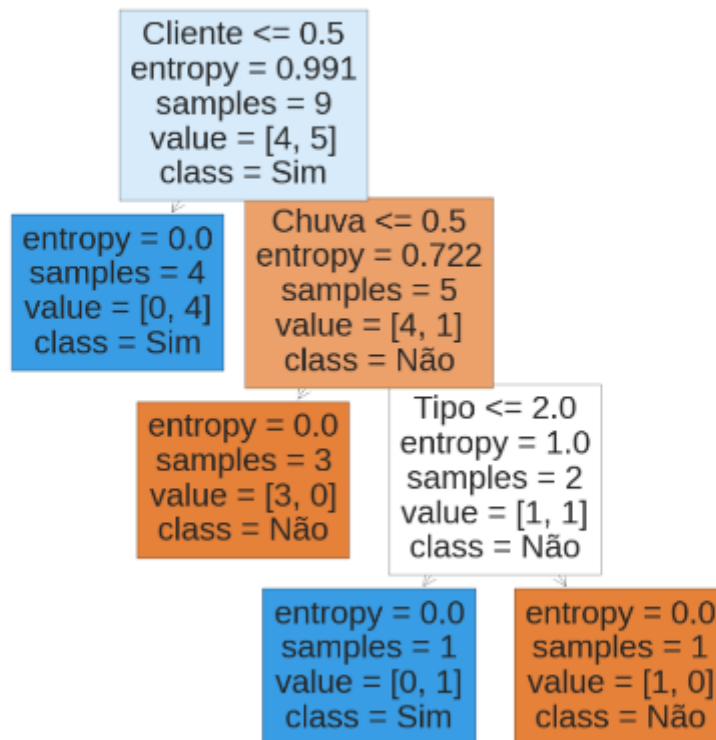
D S T Q Q S S

Árvore gerada



2-a)





b) Não houve melhorias significativas. Algumas regras de decisão foram alteradas, como a raiz da árvore, que mudou de "Clientes" para "Alternativo", mas as entropias permaneceram inalteradas.

c) Os testes experimentais variando os hiperparâmetros. Nos dois primeiros, apliquei o "critério gini" em conjunto com "max\_features" definidos em 0,8 e 0,4. Observou-se que, quando "max\_features" foi 0,4, o modelo teve um desempenho superior em termos de generalização. Nos outros dois experimentos, usei "entropia" com "max\_features" ajustados para 0,8 e 0,4. O primeiro teste com "entropia" mostrou resultados aceitáveis, com um conjunto de regras razoavelmente bom. Em contrapartida, o segundo teste apresentou desempenho insatisfatório, mostrando pouca capacidade de generalizar.

**3 - a)** O algoritmo ID3 é baseado no algoritmo CLS (Concept Learning System) e é um algoritmo de aprendizado supervisionado, construindo a árvore a partir de um conjunto fixo de dados. O ID3 constrói a árvore baseado nas informações de ganho obtidas das instâncias de treinamento que então são usadas para classificar as instâncias de teste. O ID3 geralmente utiliza atributos nominais sem valores faltantes para a classificação.

**3 - b)** O C4.5 surge como uma extensão do ID3 com a proposta de superar suas limitações. O C4.5 usa o "ganho de informação" para medir a taxa de ganho e consegue lidar com dados contínuos e dados faltantes. Assim como no ID3, os dados

são classificados em cada nó da árvore. O algoritmo usa a taxa de ganho para avaliar o atributo de divisão, sendo que o atributo com o maior ganho de informação é escolhido para fazer a decisão. O C4.5 transforma os dados numéricos em intervalos. Para calcular o ganho desses atributos, primeiro os dados são ordenados de maneira crescente, removendo os iguais e em seguida considera-se os pontos de divisão possíveis para criação de intervalos, cada intervalo é tratado como um atributo categórico.