

Lista 1 IA

Henrique Oliveira da Cunha Franco

August 2024

1 Q1

Orientação: Considerando-se a base de dados sobre “**Esperar ou não pelo restaurante**” (verificar base de dados ‘**Restaurante.csv**!’ disponibilizada no CANVAS), pede-se:

1. Calcular o ganho de informação de cada atributo. Que atributo é a raiz da árvore?
2. Que atributo estará no segundo nível da árvore. Faça os cálculos e apresente a árvore gerada até o segundo nível da árvore.

1.1 R1

O atributo raiz da árvore será Cliente. Imagens das contas referente a essa resposta no final!

1.2 R2

O atributo no segundo nível da árvore será Fome. Imagens das contas referente a essa resposta no final!

2 Q2

Orientação: Considerando-se os códigos (**Lendo_e_tratando_arquivo.ipynb**) **DecisionTree.ipynb**) disponibilizados no CANVAS, pede-se:

1. Com a codificação atual dos atributos de entrada, rode-os e plote a árvore de decisão obtida.
2. Altere a codificação dos atributos (preço, cliente e tempo de espera) para um atributo nominal ordinal e faça os experimentos acima novamente. O que você observa? Houve ganho?
3. Utilizando o código (**DecisionTree_com_gridsearch.ipynb**), ajuste os hiperparâmetros da árvore e veja que hiperparâmetros retornaram os melhores resultados. Investigue outros hiperparâmetros da árvore.

2.1 R1

Imagem Q2.1, presente no final do arquivo.

2.2 R2

O fato observado foi que não houve ganho, logo, não foram observadas diferenças ao alterar a codificação dos atributos para um atributo nominal ordinal.

3 Q3

Orientação: Leia o artigo “**A_comparative_study_of_decision_tree_ID3_and_C4.5.pdf**” que está no CANVAS e responda:

1. Quais as diferenças entre os algoritmos de árvore ID3 e C4.5?
2. Como o algoritmo C4.5 lida com os atributos de entrada que são numéricos?

3.1 R1

Diferenças entre ID3 e C4.5:

- **Tipo de Dados:** ID3 lida com atributos categóricos, C4.5 lida com atributos categóricos e numéricos.
- **Ganho de Informação:** Ambos usam ganho de informação, mas C4.5 utiliza *gain ratio* para corrigir o viés do ID3.
- **Valores Desconhecidos:** C4.5 lida com valores desconhecidos, atribuindo pesos, enquanto o ID3 não.
- **Poda:** C4.5 realiza poda para evitar *overfitting*, o ID3 não.
- **Eficiência:** C4.5 é mais eficiente e preciso.

3.2 R2

Como o C4.5 lida com atributos numéricos:

C4.5 converte atributos numéricos em categóricos criando limiares. Ordena os valores numéricos, calcula o ganho de informação para possíveis pontos de divisão, e seleciona o ponto que maximiza o ganho. Isso divide o atributo em duas categorias: valores menores ou iguais ao limiar e maiores.

Lista 1 - IA

$$ganho(x) = Entropia(classe) - Entropia(x)$$

1) Atributo Alternativo

$$1 - \left(\frac{6}{12} \cdot E\left(\frac{3}{6}, \frac{3}{6}\right) + \frac{6}{12} \cdot E\left(\frac{3}{6}, \frac{3}{6}\right) \right)$$

Não Sim

(3) (3) (3) (3)

1 - 1 = 0

Atributo Bar → Mesma coisa! logo, ganho = 0

$$1 - 1 = 0$$

Não Sim

(3) (3) (3) (3)

Atributo lex / lab

$$1 - \left(\frac{5}{12} \cdot E\left(\frac{2}{5}, \frac{3}{5}\right) + \frac{7}{12} \cdot E\left(\frac{4}{7}, \frac{3}{7}\right) \right)$$

Não Sim

(4) (3) (2) (3)

0,970951 0,985228

1 - $\left(\frac{5}{12} \cdot 0,970951 + \frac{7}{12} \cdot 0,985228 \right) = 0,0207207$

Atributo Jome

$$1 - \left(\frac{5}{12} \cdot E\left(\frac{1}{5}, \frac{4}{5}\right) + \frac{7}{12} \cdot E\left(\frac{5}{7}, \frac{2}{7}\right) \right)$$

Não Sim

(1) (4) (5) (2)

0,72198 0,863121

1 - $\left(\frac{5}{12} \cdot 0,72198 + \frac{7}{12} \cdot 0,863121 \right) = 0,195618$

tilibra

Atributo Cliente - Ganador!

$$1 - \left(\frac{2}{12} \cdot E\left(\frac{0}{2}, \frac{2}{2}\right) + \frac{4}{12} \cdot E\left(\frac{4}{1}, \frac{0}{4}\right) + \frac{6}{12} \cdot E\left(\frac{2}{6}, \frac{4}{6}\right) \right)$$

$$1 - \left(\frac{2}{12} \cdot 0,918296 + \frac{4}{12} \cdot 0,918296 + \frac{6}{12} \cdot 0,918296 \right) = 0,540867$$

Atributo Preço

$$1 - \left(\frac{7}{12} \cdot E\left(\frac{4}{7}, \frac{3}{7}\right) + \frac{2}{12} \cdot E\left(\frac{2}{2}, \frac{0}{2}\right) + \frac{3}{12} \cdot E\left(\frac{1}{3}, \frac{2}{3}\right) \right)$$

$$1 - \left(\frac{7}{12} \cdot 0,985228 + \frac{2}{12} \cdot 0,918296 + \frac{3}{12} \cdot 0,918296 \right) = 0,19571$$

Atributo Chama

$$1 - \left(\frac{5}{12} \cdot E\left(\frac{3}{5}, \frac{2}{5}\right) + \frac{7}{12} \cdot E\left(\frac{3}{7}, \frac{4}{7}\right) \right)$$

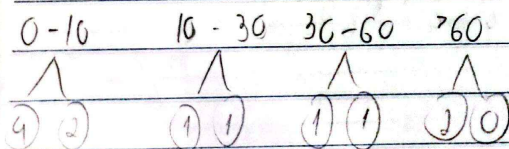
$$1 - \left(\frac{5}{12} \cdot 0,0207208 + \frac{7}{12} \cdot 0,0207208 \right) = 0,0207208$$

Atributo Reserva

$$1 - \left(\frac{5}{12} \cdot E\left(\frac{3}{5}, \frac{2}{5}\right) + \frac{7}{12} \cdot E\left(\frac{2}{7}, \frac{5}{7}\right) \right)$$

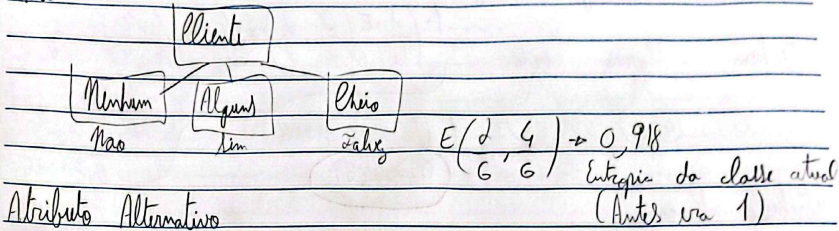
$$1 - \left(\frac{5}{12} \cdot 0,0919503 + \frac{7}{12} \cdot 0,0919503 \right) = 0,0919503$$

Atributo Tempo





1.2



Atributo Alternativo

$$0,918 - \left(\frac{5}{6} E\left(\frac{2}{5}, \frac{3}{5}\right) + \frac{1}{6} E(0,1) \right) =$$

Diagram for Alternativo:

```
graph TD
    A((1)) --- Nao1[Nao]
    A --- Lim1[lim]
    B((2)) --- Nao2[Nao]
    B --- Lim2[lim]
    C((3)) --- Nao3[Nao]
    C --- Lim3[lim]
```

$$= -0,000295834$$

0,918296

Atributo Bar

$$0,918 - \left(\frac{3}{6} E\left(\frac{1}{3}, \frac{2}{3}\right) + \frac{3}{6} E\left(\frac{1}{3}, \frac{2}{3}\right) \right) =$$

Diagram for Bar:

```
graph TD
    A((1)) --- Nao1[Nao]
    A --- Lim1[lim]
    B((2)) --- Nao2[Nao]
    B --- Lim2[lim]
    C((1)) --- Nao3[Nao]
    C --- Lim3[lim]
    D((2)) --- Nao4[Nao]
    D --- Lim4[lim]
```

$$= -0,000295834$$

0,918276

Atributo No/Tab

$$0,918 - \left(\frac{5}{6} E\left(\frac{2}{5}, \frac{3}{5}\right) + \frac{1}{6} E(1,0) \right) =$$

Diagram for No/Tab:

```
graph TD
    A((0)) --- Nao1[Nao]
    A --- Lim1[lim]
    B((1)) --- Nao2[Nao]
    B --- Lim2[lim]
    C((2)) --- Nao3[Nao]
    C --- Lim3[lim]
    D((3)) --- Nao4[Nao]
    D --- Lim4[lim]
```

$$= 0,108875$$

Atributo Zome

$$0,918 - \left(\frac{4}{6} E\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{2}{6} E(0,1) \right) =$$

Diagram for Zome:

```
graph TD
    A((0)) --- Nao1[Nao]
    A --- Lim1[lim]
    B((2)) --- Nao2[Nao]
    B --- Lim2[lim]
    C((2)) --- Nao3[Nao]
    C --- Lim3[lim]
```

$$= 0,25133...$$



Atributo Preço

$$\begin{array}{c}
 R \quad RRR \\
 \wedge \quad \wedge \\
 (2) (2) \quad (0) (2)
 \end{array}
 \quad 0,918 - \left(\frac{4}{6} E\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{2}{6} E(0, 1) \right)$$

$$= 0,251333...$$

Atributo Chama

$$\begin{array}{c}
 Mo \quad Sim \\
 \wedge \quad \wedge \\
 (1) (1) \quad (1) (3)
 \end{array}
 \quad 0,918 - \left(\frac{2}{6} E\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{6} E\left(\frac{1}{4}, \frac{3}{4}\right) \right)$$

$$= 0,0438146$$

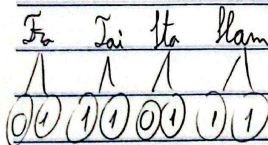
Atributo Reserva

$$\begin{array}{c}
 Mo \quad Sim \\
 \wedge \quad \wedge \\
 (0) (2) \quad (2) (2)
 \end{array}
 \quad 0,918 - \left(\frac{2}{6} E\left(\frac{0}{2}, \frac{1}{2}\right) + \frac{4}{6} E\left(\frac{2}{4}, \frac{2}{4}\right) \right)$$

$$= 0,254333...$$

$$\begin{array}{c}
 Atributo Tipo \\
 0,918 - \left(\frac{2}{6} E\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{6} E\left(\frac{1}{2}, \frac{1}{2}\right) \right)$$

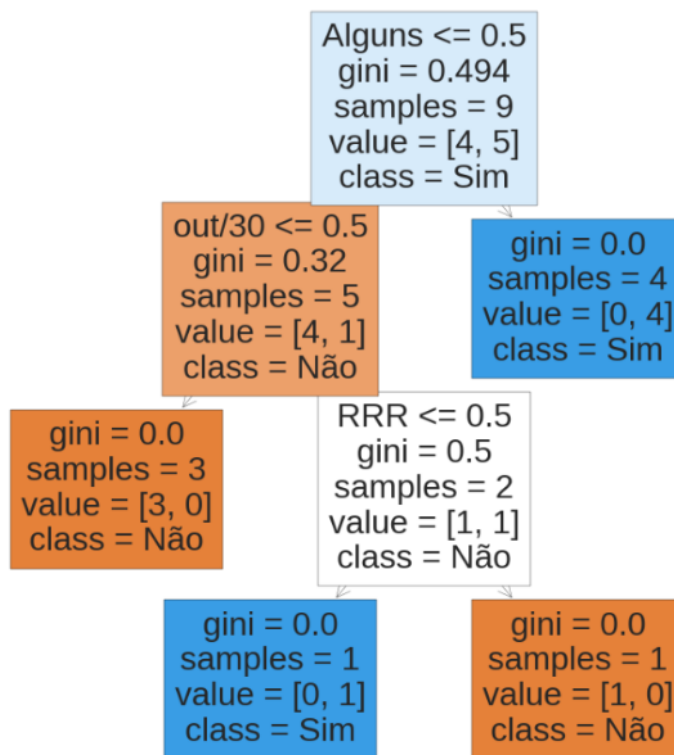
$$= 0,251333...$$



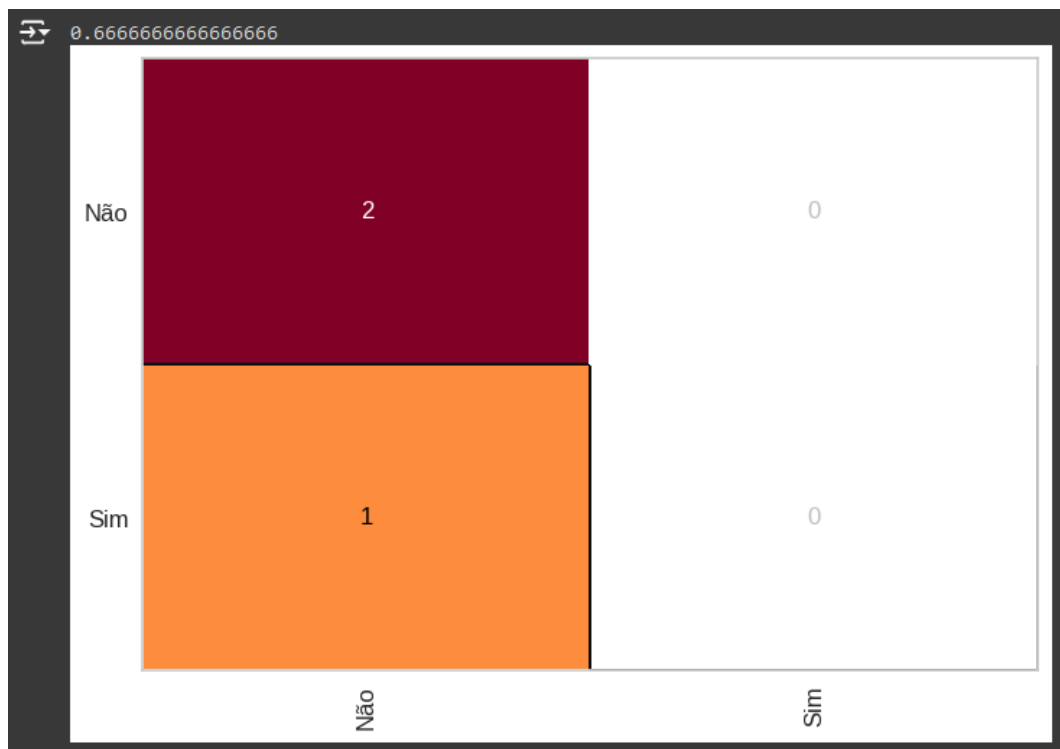
Atributo Tempo

$$\begin{array}{c}
 10-30 \quad 30-60 \quad >60 \\
 \wedge \quad \wedge \quad \wedge \\
 (1) (1) \quad (1) (1) \quad (0) (2)
 \end{array}
 \quad 0,918 - \left(\frac{2}{6} E\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{6} E\left(\frac{1}{2}, \frac{1}{2}\right) \right)$$

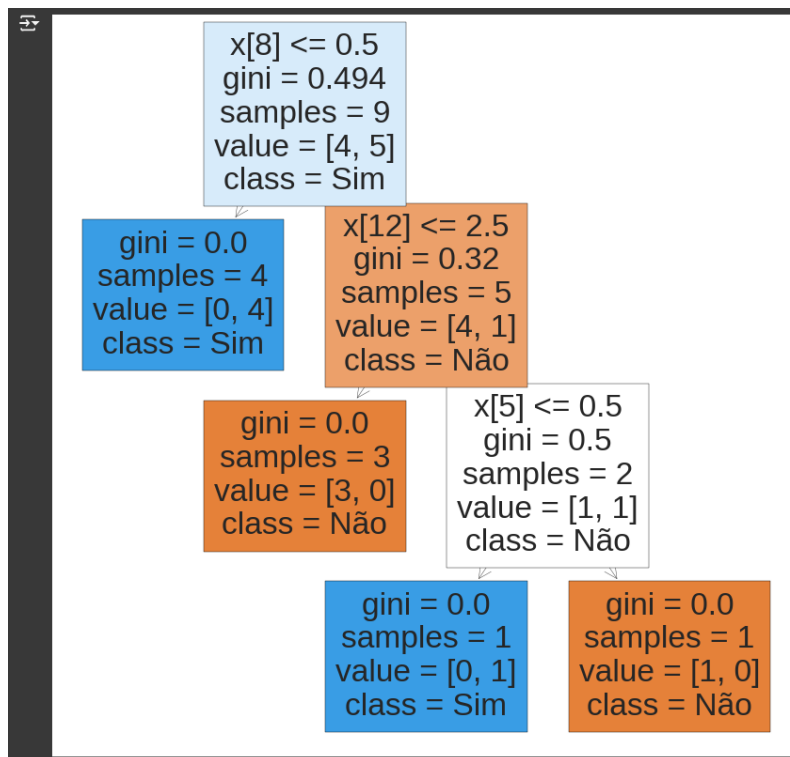
$$= 0,2513333...$$



Q2.1 - Árvore de Decisão Obtida ao rodar o código



Q2.2 Imagem 1



Q2.2 - Árvore gerada com atributo **Tipo** nominal ordinal