

## Lista 3 de IA

domingo, 10 de março de 2024 18:18

- 1) C
- 2) C I e II apenas.
- 3)

	PRECISÃO	RECALL	F1Score	TVP	TFN	TFP	TVN
A	58,2%	58,82%	58,82%	10/17	7/17	7/115	98/105
B	65,22%	83,33%	73,17%	15/18	3/18	8/104	96/104
C	76,92%	66,67%	71,43%	20/30	10/30	6/92	86/92
D	89,29%	87,72%	88,50%	50/57	7/57	6/65	59/65

A = 10    fn = 7    Fp = 7  
 B = 15    fn = 8    Fp = 8  
 C = 20    fn = 6    Fp = 6  
 D = 50    fn = 6    Fp = 6

Precisão = A = 10/17 = 58,2%  
 B = 15/23 = 65,22%  
 C = 20/26 = 76,92%  
 D = 50/56 = 89,29%  
 Recall = A = 10/17 ≈ 58,82%  
 B = 15/18 ≈ 83,33%  
 C = 20/30 ≈ 66,67%  
 D = 50/57 ≈ 87,72%

F1 = A)  $\frac{2 \cdot 58,82 \cdot 58,82}{58,82 + 58,82} = 58,82\%$   
 B)  $\frac{2 \cdot 65,22 \cdot 83,33}{65,22 + 83,33} \approx 73,17\%$

$$\text{Precisão} = \frac{VP}{VP + FP}$$

$$\text{Recall} = \frac{VP}{VP + FN}$$

$$F1 = \frac{2 \cdot p \cdot r}{p + r}$$

p = precisão  
 r = Recall

C)  $\frac{2 \cdot 76,92 \cdot 66,67}{76,92 + 66,67} \approx 71,43\%$   
 D)  $\frac{2 \cdot 89,29 \cdot 87,72}{89,29 + 87,72} \approx 88,50\%$

4)

Atributos de Classificações ?

História do crédito	Dívida	Garantias	Renda Anual	Risco
Ruim	Alta	Nenhuma	<15000	Alto
Desconhecido	Alta	Nenhuma	>=15000 a <=35000	Alto
Desconhecido	Baixa	Nenhuma	>=15000 a <=35000	Moderado
Desconhecido	Baixa	Nenhuma	>35000	Alto
Desconhecido	Baixa	Nenhuma	>35000	Baixo
Desconhecido	Baixa	Adequada	>35000	Baixo
Ruim	Baixa	Nenhuma	<15000	Alto
Ruim	Baixa	Adequada	>35000	Moderado
Boa	Baixa	Nenhuma	>35000	Baixo
Boa	Alta	Adequada	>35000	Baixo
Boa	Alta	Nenhuma	<15000	Alto
Boa	Alta	Nenhuma	>=15000 a <=35000	Moderado
Boa	Alta	Nenhuma	>35000	Baixo
Ruim	Alta	Nenhuma	>=15000 a <=35000	Alto

A → 6/14

M → 3/14

B → 5/14

$$E \rightarrow -\frac{6}{14} \log_2\left(\frac{6}{14}\right) - \frac{3}{14} \log_2\left(\frac{3}{14}\right) - \frac{5}{14} \log_2\left(\frac{5}{14}\right) \approx 1,53$$

ganho(H.C) → Ruim:  $\frac{4}{14}$   
 B:  $\frac{0}{4}$  M:  $\frac{1}{4}$  A:  $\frac{3}{4}$

Desconhecido:  $\frac{5}{14}$   
 B:  $\frac{2}{5}$  M:  $\frac{1}{5}$  A:  $\frac{2}{5}$

Boa:  $\frac{5}{14}$   
 B:  $\frac{3}{5}$  M:  $\frac{1}{5}$  A:  $\frac{1}{5}$

$$g(H.C) \rightarrow 1,53 - \left(\frac{4}{14} \cdot E\left(\frac{0}{4}, \frac{1}{4}, \frac{3}{4}\right) + \frac{5}{14} \cdot E\left(\frac{2}{5}, \frac{1}{5}, \frac{2}{5}\right) + \frac{5}{14} \cdot E\left(\frac{3}{5}, \frac{1}{5}, \frac{1}{5}\right)\right)$$

$$1,53 - \left(\frac{4}{14} \cdot 0,81 + \frac{5}{14} \cdot 1,52 + \frac{5}{14} \cdot 1,37\right)$$

$$1,53 - (0,23 + 0,54 + 0,49)$$

$$1,53 - 1,26 = 0,27$$

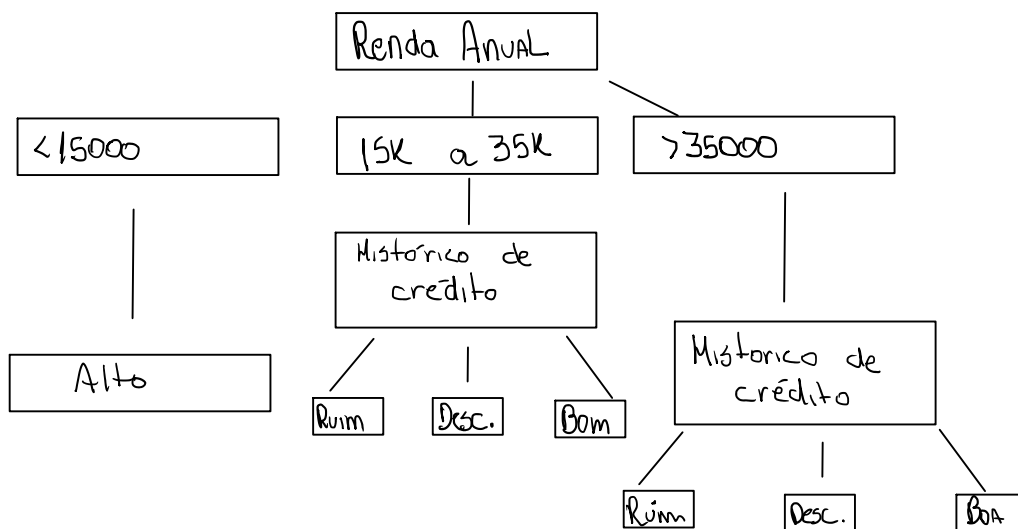
ganho(Dívida) = Alta:  $\frac{7}{14}$  Baixo:  $\frac{7}{14}$

$$\begin{aligned} \text{ganho(Dívida)} &= \text{Alta: } \frac{7}{14} \quad \text{Baixo: } \frac{7}{14} \\ &B: \frac{2}{7} \quad M: \frac{1}{7} \quad A: \frac{4}{7} \quad B: \frac{3}{7} \quad M: \frac{2}{7} \quad A: \frac{2}{7} \\ g(0) &= 1,53 - \left( \frac{7}{14} \cdot E\left(\frac{2}{7}, \frac{1}{7}, \frac{4}{7}\right) + \frac{7}{14} \cdot E\left(\frac{3}{7}, \frac{2}{7}, \frac{2}{7}\right) \right) \\ &1,53 - \left( \frac{7}{14} \cdot 1,38 + \frac{7}{14} \cdot 1,56 \right) \\ &1,53 - (0,69 + 0,78) \\ &1,53 - 1,47 = 0,06 \end{aligned}$$

$$\begin{aligned} \text{ganho(Garantia)} &= \text{Nenhuma: } \frac{11}{14} \quad \text{Adequada: } \frac{3}{14} \\ &B: \frac{3}{11} \quad M: \frac{2}{11} \quad A: \frac{6}{11} \quad B: \frac{2}{3} \quad M: \frac{1}{3} \quad A: \frac{0}{3} \\ g(6) &= 1,53 - \left( \frac{11}{14} \cdot E\left(\frac{3}{11}, \frac{2}{11}, \frac{6}{11}\right) + \frac{3}{14} \cdot E\left(\frac{2}{3}, \frac{1}{3}, \frac{0}{3}\right) \right) \\ &1,53 - \left( \frac{11}{14} \cdot 1,44 + \frac{3}{14} \cdot 0,92 \right) \\ &1,53 - (1,13 + 0,20) = 1,53 - 1,33 = 0,2 \end{aligned}$$

$$\begin{aligned} \text{ganho(Renda Anual)} &= <15000: \frac{3}{14}; >15000 \text{ a } <=35000 = \frac{4}{14} \\ \text{RAIZ} \quad &B: \frac{0}{3} \quad M: \frac{0}{3} \quad B: \frac{0}{4} \quad M: \frac{2}{4} \quad A: \frac{2}{4} \\ &>35000: \frac{7}{14} \\ &B: \frac{5}{7} \quad M: \frac{1}{7} \quad A: \frac{1}{7} \end{aligned}$$

$$\begin{aligned} g(RA) &= 1,53 - \left( \frac{3}{14} \cdot E\left(\frac{0}{3}, \frac{0}{3}, \frac{3}{3}\right) + \frac{4}{14} \cdot E\left(\frac{0}{4}, \frac{2}{4}, \frac{2}{4}\right) + \frac{7}{14} \cdot E\left(\frac{5}{7}, \frac{1}{7}, \frac{1}{7}\right) \right) \\ &1,53 - \left( \frac{3}{14} \cdot 0 + \frac{4}{14} \cdot 1 + \frac{7}{14} \cdot 1,16 \right) \\ &1,53 - (0 + 0,29 + 0,68) = 1,53 - 0,97 = 0,56 \end{aligned}$$



Entropia  $>=15000$  a  $<=35000 = 1$

$$\begin{aligned} \text{ganho(Historico de Crédito)} &= 1 - \left( \frac{1}{4} \cdot E\left(\frac{0}{1}, \frac{0}{1}, \frac{1}{1}\right) + \frac{2}{4} \cdot E\left(\frac{0}{2}, \frac{1}{2}, \frac{1}{2}\right) + \frac{1}{4} \cdot E\left(\frac{0}{1}, \frac{1}{1}, \frac{0}{1}\right) \right) \\ &1 - \left( \frac{1}{4} \cdot 0 + \frac{2}{4} \cdot 1 + \frac{1}{4} \cdot 0 \right) = 1 - 0,5 = 0,5 \end{aligned}$$

$$\begin{aligned} \text{Ruim: } &\frac{1}{4} \\ \text{Baixo: } &\frac{0}{1} \quad \text{Moderado: } \frac{0}{1} \quad \text{Baixo: } \frac{0}{2} \quad \text{Moderado: } \frac{1}{2} \quad \text{Baixo: } \frac{0}{1} \quad \text{Moderado: } \frac{1}{1} \end{aligned}$$

$$\begin{array}{l} \text{Baixo: } \frac{0}{1} ; \text{ Moderado: } \frac{0}{1} \quad \text{Baixo: } \frac{0}{2} ; \text{ Moderado: } \frac{1}{2} \quad \text{Baixo: } \frac{0}{1} \text{ Moderado: } \frac{1}{1} \\ \text{Alto: } \frac{1}{1} \quad \text{Alto: } \frac{1}{2} \quad \text{Alto: } \frac{0}{1} \end{array}$$

$$\text{ganho(Dívida)} = 1 - \left( \frac{3}{4} \cdot E\left(\frac{0}{3}, \frac{1}{3}, \frac{2}{3}\right) + \frac{1}{4} \cdot E\left(\frac{0}{1}, \frac{1}{1}\right) \right) \\ 1 - \left( \frac{3}{4} \cdot 0,92 + \frac{1}{4} \cdot 0 \right) = 1 - 0,69 = 0,31$$

$$\begin{array}{l} \text{Alta: } \frac{2}{4} \quad \text{Baixa: } \frac{1}{4} \\ B: \frac{0}{3} \quad M: \frac{2}{3} \quad B: \frac{0}{1} \quad M: \frac{1}{1} \\ A: \frac{2}{3} \quad A: \frac{0}{1} \end{array}$$

$$\text{ganho(Garantias)} = 1 - \left( \frac{4}{4} \cdot E\left(\frac{0}{4}, \frac{2}{4}, \frac{2}{4}\right) + \frac{0}{4} \cdot E\left(\frac{0}{0}, \frac{0}{0}, \frac{0}{0}\right) \right) \\ 1 - (1 \cdot 1 + 0) = 1 - 1 = 0$$

$$\text{Nenhuma: } \frac{4}{4}$$

$$\begin{array}{l} B: \frac{0}{4} \quad M: \frac{2}{4} \quad \text{Adequado: } \frac{0}{4} \\ A: \frac{2}{4} \quad \text{Tudo 0} \end{array}$$

$$\text{Entropia} > 35000 = 1,15$$

$$\text{ganho(Histórico de Crédito)} = 1,15 - \left( \frac{1}{4} \cdot E\left(\frac{0}{1}, \frac{1}{1}, \frac{0}{1}\right) + \frac{3}{4} \cdot E\left(\frac{2}{3}, \frac{0}{3}, \frac{1}{3}\right) + \frac{3}{4} \cdot E\left(\frac{2}{3}, \frac{0}{3}, \frac{0}{3}\right) \right) \\ \text{Ruim: } \frac{1}{4} \quad \text{Desconhecido: } \frac{3}{4} \quad \text{Boa: } \frac{3}{4} \\ 1,15 - \left( \frac{1}{4} \cdot 0 + \frac{3}{4} \cdot 0,92 + \frac{3}{4} \cdot 0 \right) = 1,15 - 0,39 = 0,76$$

$$\text{ganho(Dívida)} = 1,15 - \left( \frac{2}{4} \cdot E\left(\frac{2}{2}, \frac{0}{2}, \frac{0}{2}\right) + \frac{5}{4} \cdot E\left(\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right) \right) \\ 1,15 - \left( \frac{2}{4} \cdot 0 + \frac{5}{4} \cdot 1,37 \right) = 1,15 - 0,98 = 0,17$$

$$\begin{array}{l} \text{Alta: } \frac{2}{4} \quad \text{Baixa: } \frac{2}{4} \\ B: \frac{2}{2} \quad M: \frac{0}{2} \quad B: \frac{2}{3} \quad M: \frac{1}{3} \\ A: \frac{0}{2} \quad A: \frac{1}{3} \end{array}$$

$$\text{ganho(Garantia)} = 1,15 - \left( \frac{4}{4} \cdot E\left(\frac{2}{4}, \frac{0}{4}, \frac{1}{4}\right) + \frac{3}{4} \cdot E\left(\frac{2}{3}, \frac{1}{3}, \frac{0}{3}\right) \right) \\ 1,15 - \left( \frac{4}{4} \cdot 0,81 + \frac{3}{4} \cdot 0,92 \right) = 1,15 - (0,46 + 0,39) = 1,15 - 0,85 = 0,3$$

$$\begin{array}{l} \text{Nenhuma: } \frac{4}{4} \quad \text{Adequada: } \frac{3}{4} \\ B: \frac{2}{4} \quad M: \frac{0}{4} \quad B: \frac{2}{3} \quad M: \frac{1}{3} \\ A: \frac{1}{4} \quad A: \frac{0}{3} \end{array}$$

5)

```
# Defina a função para calcular a entropia def calculando_entropia(y):
# Conta a ocorrência de cada classe no conjunto de dados valores_unicos, contagens = np.unique(y,
return_counts=True) # Calcula as probabilidades das classes
probabilidades = contagens / len(y) # Calcula a entropia
entropia = -np.sum(probabilidades * np.log2(probabilidades)) return entropia

# Defina a função para calcular o ganho de informação def calculando_ganho_informacao(X, y, atributo):
entropia_anterior = calculando_entropia(y) valores_unicos = np.unique(X[atributo])
entropia_ponderada_apos_divisao = 0
```

```

    for valor in valores_unicos:
        subset_y = y[X[atributo] == valor] entropia_ponderada_apos_divisao += len(subset_y) / len(y) *
calculando_entropia(subset_y)

    ganho_informacao = entropia_anterior - entropia_ponderada_apos_divisao return ganho_informacao

# Defina a função recursiva para construir a árvore ID3 def construindo_arvore_id3(X, y, atributos):
    # Caso base: se todos os rótulos são iguais, retornando esse rótulo if len(np.unique(y)) == 1:
        return np.unique(y)[0]

    # Caso base: se não houver mais atributos para dividir, retornando a classe majoritária if len(atributos)
    == 0:
        return np.unique(y)[np.argmax(np.bincount(y))]

    # Escolhendo o atributo com o maior ganho de informação
    melhor_atributo = max(atributos, key=lambda atributo: calculando_ganho_informacao(X, y, atributo))

    # Criando um nó de árvore com o atributo escolhido arvore = {melhor_atributo: {}}

    # Removendo o atributo escolhido da lista de atributos
    atributos = [atributo for atributo in atributos if atributo != melhor_atributo]

    # Construindo subárvores recursivamente para cada valor do atributo escolhido for valor in
    np.unique(X[melhor_atributo]):
        subset_X = X[X[melhor_atributo] == valor] subset_y = y[X[melhor_atributo] == valor]
        arvore[melhor_atributo][valor] = construindo_arvore_id3(subset_X, subset_y, atributos)

    return arvore # Exemplo de uso
if name == " main ":
    # Criando a base de dados de exemplo data = pd.DataFrame({
        'História do Crédito': ['Ruim', 'Desconhecida', 'Desconhecida', 'Ruim', 'Desconhecida', 'Baixa',
        'Baixa', 'Desconhecida', 'Ruim', 'Desconhecida'],
        'Divida': ['Alta', 'Alta', 'Baixa', 'Alta', 'Baixa', 'Baixa', 'Baixa', 'Baixa', 'Baixa', 'Alta', 'Alta'],
        'Garantias': ['Nenhuma', 'Nenhuma', 'Nenhuma', 'Nenhuma', 'Nenhuma', 'Nenhuma', 'Nenhuma', 'Adequada', 'Nenhuma',
        'Nenhuma', 'Adequada'],
        'Renda Anual': ['<15000', '>=15000 a <=35000', '>=15000 a <=35000', '>35000',
        '>35000', '>=15000 a <=35000', '>35000', '<15000', '>35000', '>=15000 a <=35000'],
        'Risco': ['Alto', 'Alto', 'Moderado', 'Alto', 'Baixo', 'Baixo', 'Baixo', 'Alto', 'Baixo', 'Alto']
    })

    X = data.drop(columns=['Risco']) y = data['Risco']
    atributos = X.columns

    # Construindo a árvore de decisão
    arvore_decisao = construindo_arvore_id3(X, y, atributos)

    # Imprimindo a árvore de decisão import pprint pprint.pprint(arvore_decisao)

```

6)

**Preparação dos Dados:** Coleta-se um conjunto de dados com observações (exemplos) e rótulos de classe. Cada observação possui atributos.

**Escolha da Melhor Divisão:** O algoritmo busca dividir o conjunto de dados em subconjuntos mais homogêneos. Ele escolhe o atributo e valor de divisão que maximizam a pureza dos subconjuntos.

**Construção da Árvore:** A árvore começa com um nó raiz que contém todos os exemplos. Ela é dividida em nós filhos com base nas divisões escolhidas. O processo é repetido para cada nó filho até atingir critérios de parada.

**Crítérios de Parada:** O algoritmo pode incluir critérios de parada, como profundidade máxima da árvore ou impureza mínima em um nó. Quando um nó não pode mais ser dividido, ele se torna um nó folha.

**Classificação ou Regressão:** A árvore construída pode ser usada para classificar novos exemplos (em árvores de classificação) ou prever valores numéricos (em árvores de regressão). Basta percorrer a árvore da raiz até um nó folha com base nos atributos do exemplo e tomar a decisão com base na classe ou valor associado ao nó folha.

7)

#### Árvore de Decisão (ID3):

Melhor atributo para divisão: História do Crédito

```
|
|- Ruim:
|   |
|   |- Resultado: Não
|
|- Desconhecida:
|   |
|   |- Resultado: Sim
|
|- Boa:
|   |
|   |- Resultado: Sim
|
```

#### Árvore de Decisão (C4.5):

Melhor atributo para divisão: Renda Anual

```
|
|- <15000:
|   |
|   |- Resultado: Alto
|
|- >=15000 a <=35000:
|   |
|   |- Resultado: Moderado
|
|- >35000:
|   |
|   |- Resultado: Baixo
|
```

#### Árvore de Decisão (CART):

Melhor atributo para divisão: História do Crédito

```
|
|- Ruim:
|   |
|   |- Melhor atributo para divisão: Dívida
|       |
|       |- Alta:
|           |
|           |- Resultado: Alto
|       |
|       |- Baixa:
|           |
|           |- Resultado: Baixo
|
|- Desconhecida:
|   |
|   |- Resultado: Moderado
|
|- Boa:
|   |
|   |- Resultado: Baixo
|
```

#### Algoritmo de Construção:

ID3: Usa ganho de informação.  
C4.5: Usa ganho de informação ponderado (gain ratio).  
CART: Usa índice Gini.  
Primeira Divisão:

ID3: Divide com base na "História do Crédito".  
C4.5: Divide com base na "Renda Anual".  
CART: Divide com base na "História do Crédito".  
Níveis de Profundidade:

ID3 e C4.5: Têm uma profundidade de 2.  
CART: Tem uma profundidade de 3.  
Número de Divisões:

ID3: 2 divisões.  
C4.5: 2 divisões.  
CART: 3 divisões.  
Resultados Finais:

ID3 e C4.5: Resultados diretamente ligados aos nós folha.  
CART: Resultados finais nas folhas após várias divisões.