

**Questão 01:**

| Coluna   | Descrição                             |
|----------|---------------------------------------|
| Survived | 1 = Sobreviveu, 0 = Não sobreviveu    |
| Pclass   | Classe (1ª, 2ª, 3ª)                   |
| Sex      | Masc/Femi                             |
| Age      | Idade                                 |
| SibSp    | Quantidade de irmãos/cônjuges a bordo |
| Parch    | Quantidade de pais/filhos a bordo     |
| Fare     | Tarifa paga                           |
| Embarked | Porto de Embarque (C, Q, S)           |

| Atributo | Impacto na Sobrevivência  |
|----------|---|
| Sexo     | Mulheres sobreviveram mais  |
| Classe   | 1ª Classe sobreviveu mais   |
| Idade    | Crianças sobreviveram mais  |
| Tarifa   | Passageiros que pagaram tarifas mais altas sobreviveram mais                |
| Embarque | Pessoas que embarcaram em Cherbourg (C) tiveram maior taxa de sobrevivência |

**Conclusão:**

A análise da base de dados mostrou que o sexo e a classe do passageiro e idade foram os atributos que mais influenciaram na sobrevivência.

Mulheres e crianças tinham maior chance de sobreviver, especialmente aquelas na 1ª classe. Passageiros que pagaram tarifas mais altas também tiveram maior taxa de sobrevivência.

**Questão 02:**

1) As diferenças entre ID3 e C4.5 são:

**Tipos de dados:** onde ID3 é Apenas categórico e o C4.5 é Categóricos e Numéricos;

**Dados Faltantes:** ID3 Ignora e C4.5 Lida usando probabilidades;

**Critério de Divisão:** ID3 usa Ganho de Informação e o C4.5 usa Razão de Ganho;

**Poda:** ID3 Não possui e o C4.5 Possui poda para evitar overfitting;

**Tratamento de Ruídos:** ID3 Não tem e o C4.5 tem.

2) O algoritmo C4.5 converte atributos numéricos em intervalos de valores. Isso é feito calculando o ponto de corte que maximiza o ganho de informação. Por exemplo, para o atributo idade, C4.5 pode criar um intervalo:

Idade  $\leq$  30 anos

Idade > 30

**Questão 03:**

**Dados de Teste:**

| INSTÂNCIA | TAMANHO DA PÉTALA | LARGURA DA PÉTALA | TAMANHO DA SÉPALA | LARGURA DA SÉPALA |
|-----------|-------------------|-------------------|-------------------|-------------------|
| 1         | 3.46              | 0.87              | 2.45              | 1.78              |
| 2         | 1.67              | 1.89              | 0.78              | 1.32              |
| 3         | 2.56              | 2.34              | 2.45              | 1.78              |
| 4         | 6.67              | 2.34              | 2.45              | 1.78              |

Usando o critério Gini com profundidade de 3 níveis, a resposta correta é:

c) Iris\_Versicolor, Iris\_Setosa, Iris\_Versicolor, Iris\_Virgínica;

**Questão 04:**

- I. Afirmação 1 ? correta
- II. Afirmação 2 ? correta
- III. Afirmação 3 ? correta

Resposta: **Letra e) I, II, III**

**Questão 05:**

|          | <b>precisão</b> | <b>recall</b> | <b>f1_score</b> | <b>TVP</b>   | <b>TFN</b>   | <b>TFP</b>  | <b>TVN</b>    |
|----------|-----------------|---------------|-----------------|--------------|--------------|-------------|---------------|
| <b>A</b> | 10/17<br>59%    | 10/17<br>59%  | 0.588           | 10/17<br>59% | 7/17<br>41%  | 7/105<br>7% | 98/105<br>93% |
| <b>B</b> | 15/23<br>65%    | 15/18<br>83%  | 0.731           | 15/18<br>83% | 3/18<br>17%  | 8/104<br>8% | 96/104<br>92% |
| <b>C</b> | 20/26<br>77%    | 20/30<br>67%  | 0.713           | 20/30<br>67% | 10/30<br>33% | 6/92<br>7%  | 86/92<br>93%  |
| <b>D</b> | 50/56<br>88%    | 50/57<br>88%  | 0.884           | 50/57<br>87% | 7/57<br>13%  | 6/65<br>9%  | 59/65<br>91%  |

**Classe A:**

- $VP = 10$
- $FN = 4 + 2 + 1 = 7$
- $FP = 1 + 2 + 4 = 7$
- $VN = 15 + 2 + 0 + 3 + 20 + 5 + 1 + 2 + 50 = 98$

**Classe B:**

- $VP = 15$
- $FN = 1 + 2 = 3$
- $FP = 4 + 3 + 1 = 8$
- $VN = 10 + 2 + 1 + 20 + 5 + 2 + 4 + 50 = 94$

**Classe C:**

- $VP = 20$
- $FN = 2 + 3 + 5 = 10$
- $FP = 2 + 2 = 4$
- $VN = 10 + 4 + 1 + 15 + 0 + 1 + 50 = 81$

**Classe D:**

- $VP = 50$
- $FN = 4 + 1 + 2 = 7$
- $FP = 1 + 0 + 5 = 6$
- $VN = 10 + 4 + 2 + 15 + 2 + 3 + 20 = 56$

$$\text{Precisão} = VP / (VP + FP)$$

$$\text{Recall} = VP / (VP + FN)$$

$$F1 = 2 \cdot (\text{Precisão} \times \text{Recall}) / (\text{Precisão} + \text{Recall})$$