

Questão 01: Padrão de Sobreviventes do Titanic

Com base na análise dos dados, os principais padrões de sobrevivência são:

Classe: Passageiros da primeira classe tiveram a maior taxa de sobrevivência, enquanto os da terceira classe tiveram a menor.

Sexo: Mulheres tiveram uma taxa de sobrevivência significativamente maior do que homens.

Idade: Crianças e idosos tiveram prioridade nos botes salva-vidas.

Familiares a bordo: Passageiros com familiares tiveram maior chance de sobrevivência.

Local de embarque: Passageiros que embarcaram em Cherbourg tiveram uma taxa de sobrevivência maior.

Questão 02: Diferenças entre ID3 e C4.5

1. Diferenças entre ID3 e C4.5

- **ID3** Utiliza o ganho de informação (entropia) para selecionar os atributos. Só funciona com atributos categóricos e não lida com valores faltantes.

- **C4.5** Melhoria do ID3, utiliza o ganho de informação normalizado (ganho de razão) para selecionar os atributos. Pode lidar com atributos numéricos e valores faltantes. Além disso, realiza poda (pruning) para evitar overfitting.

2. Como o C4.5 lida com atributos numéricos

- O C4.5 discretiza os atributos numéricos, ou seja, transforma os valores contínuos em intervalos (bins) para que possam ser tratados como categóricos. Ele escolhe o ponto de corte que maximiza o ganho de informação.

Questão 03: Classificação de Instâncias

- Instância 1: Iris_Versicolor
- Instância 2: Iris_Setosa
- Instância 3: Iris_Versicolor
- Instância 4: Iris_Virgínica

A alternativa correta é:

c) Iris_Versicolor, Iris_Setosa, Iris_Versicolor, Iris_Virgínica

Questão 04: Análise das Afirmações

1. "Esta árvore possui 5 regras de classificação."

- Depende da estrutura da árvore. Se a árvore tiver 5 folhas, então há 5 regras.

2. "Das regras geradas, há apenas uma com cobertura por classe de 100%."

- Isso significa que uma das regras cobre todos os exemplos de uma classe específica.

3. "A menor cobertura por classe é de 6.8% e corresponde à classe Iris_Virgínica."

- Isso indica que a classe Iris_Virgínica tem a menor cobertura entre as regras.

Resposta:

Se todas as afirmações forem verdadeiras, a alternativa correta é:

e) I, II e III

Questão 05: Métricas da Matriz de Confusão

Cálculos para cada classe:

Classe A

$$TP = 10$$

$$FP = 4 + 2 + 1 = 7$$

$$FN = 1 + 2 + 4 = 7$$

$$TN = 15 + 20 + 50 + 2 + 3 + 5 + 1 + 2 = 98$$

$$\text{Precisão} = 10 / (10 + 7) \approx 0.588$$

$$\text{Recall} = 10 / (10 + 7) \approx 0.588$$

$$F1\text{-Score} = 2 * (0.588 * 0.588) / (0.588 + 0.588) \approx 0.588$$

$$TVP = 0.588$$

$$TFN = 7$$

$$TFP = 7$$

$$TVN = 98 / (98 + 7) \approx 0.933$$

Classe B

$$TP = 15$$

$$FP = 1 + 2 + 0 = 3$$

$$FN = 4 + 3 + 1 = 8$$

$$TN = 10 + 20 + 50 + 2 + 5 + 2 = 89$$

$$\text{Precisão} = 15 / (15 + 3) = 0.833$$

$$\text{Recall} = 15 / (15 + 8) \approx 0.652$$

$$F1\text{-Score} = 2 * (0.833 * 0.652) / (0.833 + 0.652) \approx 0.731$$

$$TVP = 0.652$$

$$TFN = 8$$

$$TFP = 3$$

$$TVN = 89 / (89 + 3) \approx 0.967$$

Classe C

$$TP = 20$$

$$FP = 2 + 3 + 5 = 10$$

$$FN = 2 + 2 + 2 = 6$$

$$TN = 10 + 15 + 50 + 4 + 1 + 1 = 81$$

$$\text{Precisão} = 20 / (20 + 10) \approx 0.666$$

$$\text{Recall} = 20 / (20 + 6) \approx 0.769$$

$$F1\text{-Score} = 2 * (0.666 * 0.769) / (0.666 + 0.769) \approx 0.714$$

$$TVP = 0.769$$

$$TFN = 6$$

$$TFP = 10$$

$$TVN = 81 / (81 + 10) \approx 0.890$$

Classe D

$$TP = 50$$

$$FP = 1 + 0 + 5 = 6$$

$$FN = 4 + 1 + 2 = 7$$

$$TN = 10 + 15 + 20 + 4 + 2 + 3 = 54$$

$$\text{Precisão} = 50 / (50 + 6) \approx 0.893$$

$$\text{Recall} = 50 / (50 + 7) \approx 0.877$$

$$F1\text{-Score} = 2 * (0.893 * 0.877) / (0.893 + 0.877) \approx 0.885$$

$$TVP = 0.877$$

$$TFN = 7$$

$$TFP = 6$$

$$TVN = 54 / (54 + 6) = 0.9$$

Classe	Precisão	Recall	F1-Score	TVP	TFN	TFP	TVN	
A	0.588	0.588	0.588	0.588	7	7	0.933	
B	0.833	0.652	0.731	0.652	8	3	0.967	
C	0.666	0.769	0.714	0.769	6	10	0.890	
D	0.893	0.877	0.885	0.877	7	6	0.9	