

Lista de IA #2

806454 - Yago Almeida Melo

Questão 1)

A partir da árvore gerada, foi possível observar que o padrão de sobreviventes do Titanic foram mulheres que estavam na primeira ou segunda classe.

Questão 2)

1 - O algoritmo ID3 emprega o ganho de informação simples na escolha dos atributos, enquanto o C4.5 aprimora esse procedimento ao adotar a razão de ganho. Isso contribui para prevenir o viés em favor de atributos com muitos valores. Ademais, o ID3 se limita a atributos discretos e não lida com valores ausentes, ao passo que o C4.5 permite a utilização de atributos contínuos, trata dados ausentes atribuindo probabilidades e ainda realiza a poda da árvore para minimizar o overfitting.

2 - O C4.5 ordena os valores dos atributos numéricos e testa vários pontos de corte. Avalia-se em cada ponto o ganho (ou gain ratio) obtido pela divisão dos dados em dois conjuntos - um com valores abaixo ou acima do limiar e o outro com valores acima. O local que maximiza esse benefício é selecionado, transformando o atributo contínuo numa variável de divisão binária.

Questão 3)

c) Iris_Versicolor, Iris_Setosa, Iris_Versicolor, Iris_Virginica

Questão 4)

c) I e II, apenas.

Questão 5)

	Precisão	Recall	F1 Score	TVP	TFN	TFP	TVN
A	0,588	0,588	0,588	10/17	7/17	7/105	98/105
B	0,652	0,833	0,731	15/18	3/18	8/104	96/104
C	0,769	0,667	0,714	20/30	10/30	6/92	86/92
D	0,893	0,877	0,885	50/57	7/57	6/65	59/65