

Questão 3) 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?

Resposta:

1) Tipo de Atributos: O ID3 trabalha apenas com atributos categóricos, enquanto o C4.5 pode lidar tanto com atributos categóricos quanto com atributos contínuos.

Ganho de Informação: O ID3 utiliza o ganho de informação como critério de seleção de atributos. O C4.5, além do ganho de informação, introduz a razão de ganho (gain ratio), que corrige a tendência do ganho de informação em favorecer atributos com muitos valores distintos.

Valores Desconhecidos: O C4.5 pode lidar com valores desconhecidos em atributos, enquanto o ID3 não possui essa capacidade.

Poda: O C4.5 inclui uma fase de poda para reduzir o tamanho da árvore e melhorar a generalização, enquanto o ID3 não realiza poda.

Peso dos Atributos: O C4.5 permite o uso de atributos com pesos diferentes, enquanto o ID3 considera todos os atributos com o mesmo peso.

2) O C4.5 lida com atributos numéricos convertendo-os em atributos categóricos através da definição de um ponto de corte. Ele examina os valores do atributo numérico e os ordena. Em seguida, considera várias possíveis divisões (por exemplo, se os valores de um atributo forem {1, 2, 3, 4, 5}, o C4.5 consideraria divisões como "menor ou igual a 2" versus "maior que 2"). O ponto de corte que maximiza o ganho de informação ou a razão de ganho é escolhido para dividir os dados