Lista 1 – Inteligência Artificial
Vitor Dias de Britto Militão

1 –

1.1)

Entropia (classe) = 1

Alternativo
Ganho de Informação = 1 – 1 = 0

Bar
Ganho de Informação =1 – 1 = 0

SexSab
Ganho de Informação =1 - 0,97 = 0,02

Fome
Ganho de Informação =1 - 0,803 = 0,19

Cliente
Ganho de Informação =1 - 0,459 = 0,54

Preco
Ganho de Informação =0,196

Chuva
Ganho de Informação =1 – 0,978 = 0,02

Res
Ganho de Informação =1 – 0,978 = 0,02

Tipo
Ganho de Informação =1 – 1 = 0

Tempo
Ganho de Informação =1 – 0,459 + 1/6 + 1/6 = 0,87

O Atributo raiz da Árvore é “Cliente”, com ganho de 0,54.

1.2)

Alternativo
Ganho de Informação = 1.109

Bar
Ganho de Informação =1 – 1 = 0

SexSab
Ganho de Informação =1.109

Fome
Ganho de Informação =0.251 RAIZ 2

Cliente
Ganho de Informação =0.918 BASE

Preco
Ganho de Informação =0,251

Chuva
Ganho de Informação =0.043

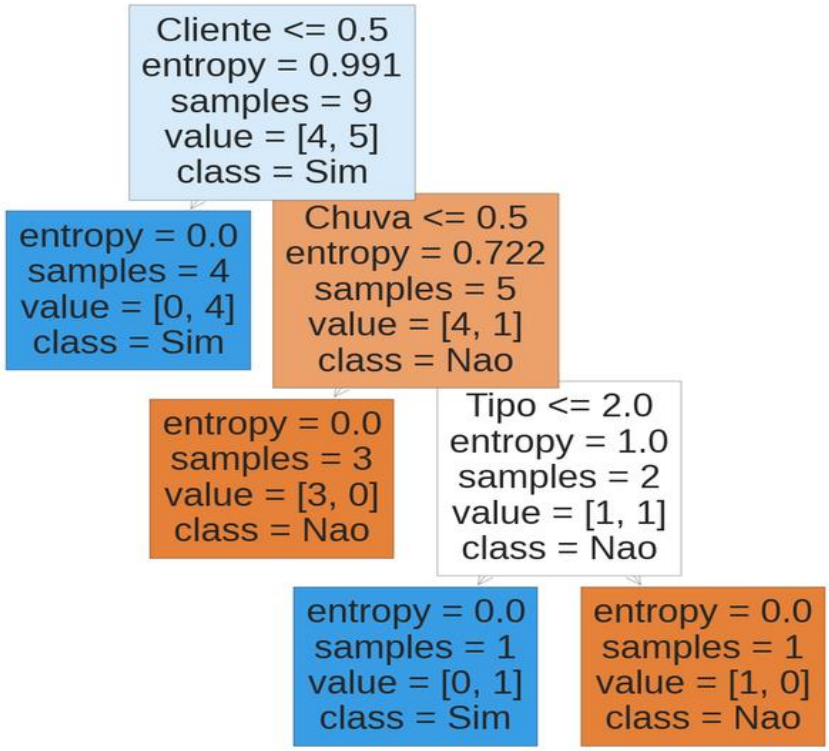
Res
Ganho de Informação =0.251

Tipo
Ganho de Informação =0.251

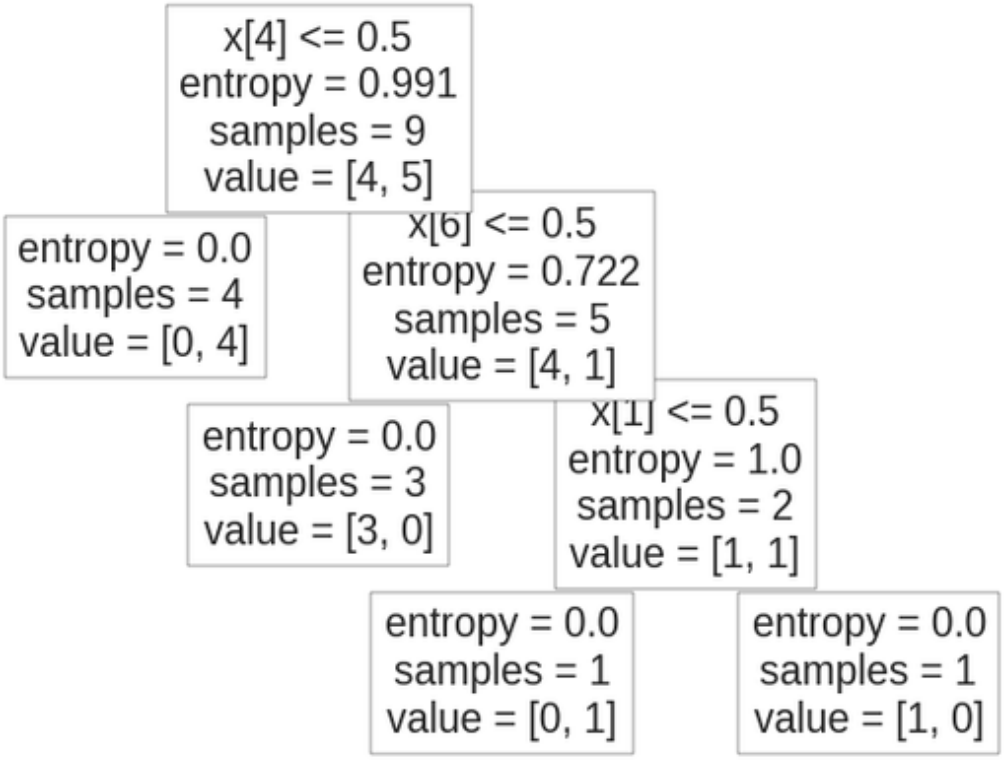
Tempo
Ganho de Informação =0.251

2 –

2.1)



2.2) Apesar das mudanças realizadas nos atributos não se faz evidente nenhuma mudança relevante, talvez pelo fato de ser uma base de dados pequena.



2.3)

```
params = {
   'criterion':  ['gini', 'entropy'],
   'max_depth':  [None, 2, 4, 6, 8, 10],
   'max_features': [None, 'sqrt', 'log2', 0.2, 0.4, 0.6, 0.8],
}
```

- {'criterion': 'entropy', 'max_depth': 10, 'max_features': 0.6} 0.9499999999999998
- {'criterion': 'gini', 'max_depth': 10, 'max_features': 0.4} 0.9583333333333333
- {'criterion': 'gini', 'max_depth': 10, 'max_features': 0.8} 0.9666666666666666

```
params = {
 'criterion': ['gini', 'entropy'],
 'max_depth': [None, 2, 4, 6, 8, 10],
 'max_features': [None, 'sqrt', 'log2', 0.2, 0.4, 0.6, 0.8],
 'min_samples_split': [2, 5, 10, 20],
 'min_samples_leaf': [1, 2, 4, 6],
 'max_leaf_nodes': [None, 5, 10, 20, 30],
}
```

- {'criterion': 'gini', 'max_depth': 4, 'max_features': 'log2', 'max_leaf_nodes': 30, 'min_samples_leaf': 6, 'min_samples_split': 20} 0.9666666666666666
- {'criterion': 'gini', 'max_depth': 2, 'max_features': 'log2', 'max_leaf_nodes': 30, 'min_samples_leaf': 2, 'min_samples_split': 10} 0.9666666666666666

3 –

3.1) Os algoritmos de Árvore de Decisão ID3 e C4.5 foram desenvolvidos por J. R. Quinlan e possuem algumas diferenças. O C4.5 pode ser reconhecido como um aprimoramento do ID3, desenvolvido para resolver suas limitações em relação a tópicos como o Tratamento de Atributos Contínuos, Valores Ausentes, Poda e Critério de Divisão, consequentemente alterando o formato de sua Árvore de Decisão.

Em relação à manipulação de atributos, enquanto o ID3 lida apenas com atributos categóricos, o C4.5 provê também o tratamento de atributos contínuos. Além Disso, o ID3 não possui tratamento de valores ausentes em dados, exigindo que todos os atributos possuam valores completos. Já o C4.5, pode lidar com valores ausentes durante o processo de construção da árvore, usando métodos baseados em probabilidade para distribuir amostras com valores ausentes.

No que se refere a mecanismos de critério e eficiência, o C4.5 introduz um mecanismo de poda pós-processamento que ajuda a reduzir o sobreajuste, removendo ramos que possuem pouca relevância ou são baseados em dados ruidosos, o que não está presente no ID3. A poda no C4.5 melhora a generalização da árvore ao torná-la mais robusta e eficiente. Essa característica torna o C4.5 mais adequado para conjuntos de dados maiores e mais complexos, onde o risco de sobreajuste é maior. Dessa forma, o C4.5 se posiciona como um algoritmo mais versátil e eficaz em comparação ao ID3, superando várias de suas limitações originais.

3.2) O algoritmo C4.5 lida com atributos numéricos ao transformá-los em atributos contínuos e ao encontrar pontos de divisão para separá-los em intervalos. Para isso, o algoritmo busca o melhor ponto de corte que divide os dados de forma que maximizem o ganho de informação, assim como faria para atributos categóricos. Esse ponto de corte é determinado avaliando os valores numéricos dos atributos e identificando o valor que melhor separa os dados em subconjuntos mais homogêneos.