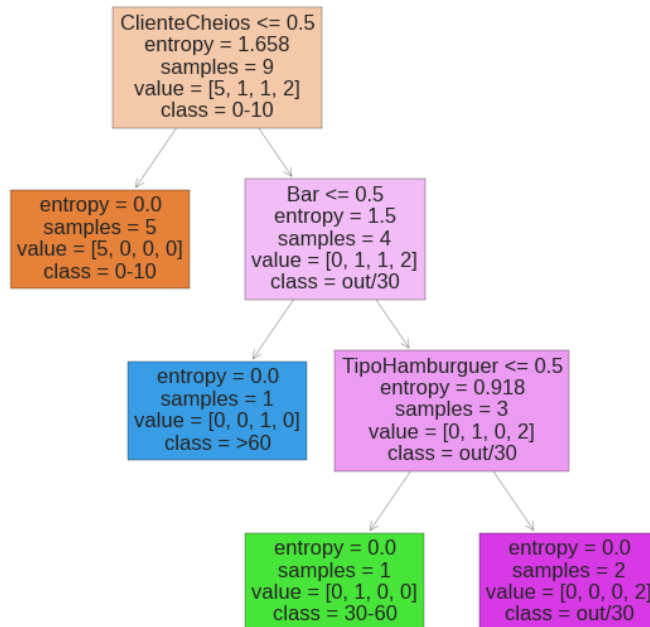


Lista 2

Questão 1:

1. Árvore gerada:



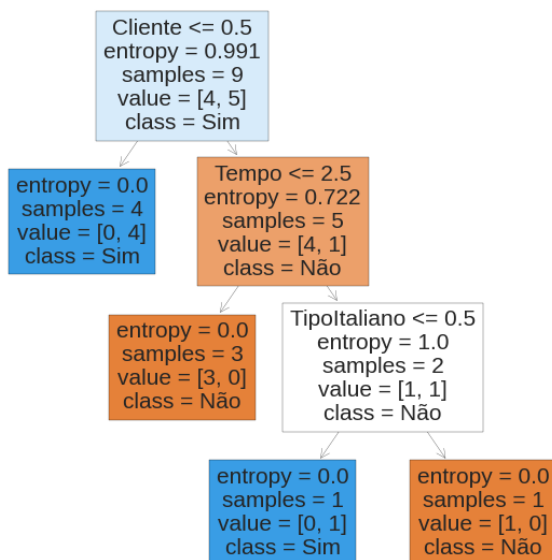
Lendo e tratando o arquivo:

<https://colab.research.google.com/drive/13HCiLZLpx0pngYS9oyW3XRhLPyV0VHtf?usp=sharing>

Decision Tree com árvore gerada:

<https://colab.research.google.com/drive/1QONMWA0KwqVVmcNli14ztDQjP-HNbokG?usp=sharing>

2. Árvore gerada:



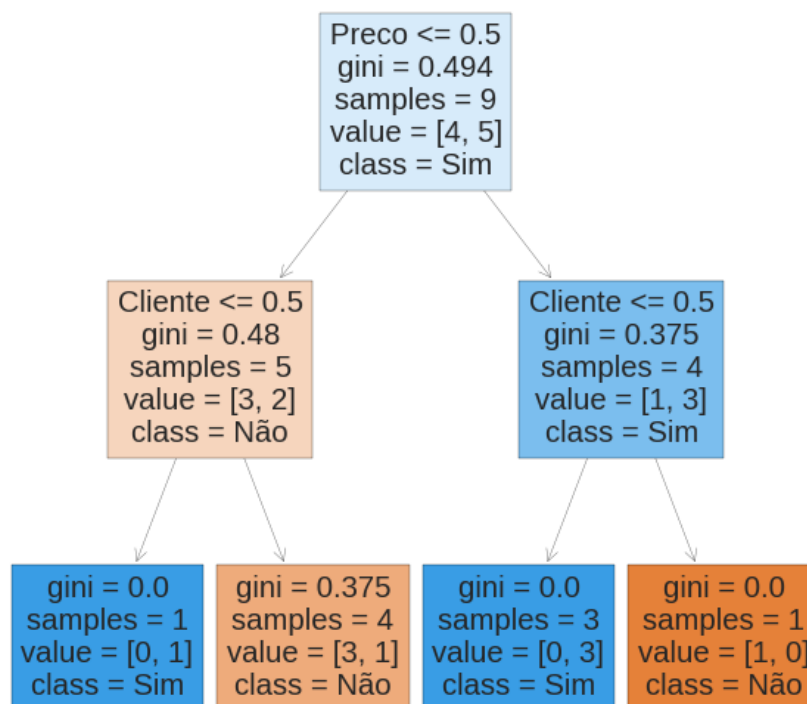
Lendo e tratando o arquivo:
https://colab.research.google.com/drive/1wD7JTn7AejJeB6SN_KYI5rKSWjVcea3-?usp=sharing

Decision Tree com árvore gerada:

<https://colab.research.google.com/drive/14TVqdsrPSvvja09MuSwHgJDN1CszCK5k?usp=sharing>

Pode-se observar que os atributos abaixo da raiz da árvore gerada são diferentes da primeira questão. Houve um ganho de 0.33333 na matriz de confusão; antes era 0.33333, agora é 0.66666.

3. Árvore gerada:



Decision Tree com Grid:
<https://colab.research.google.com/drive/1A4ASq93TrGrb1Jx4cRgsJXQgyhSJBg0B?usp=sharing>

Foram alterados os hiperparâmetros de cv e profundidade máxima da árvore; cv de 10 para 5 e profundidade máxima de 10 para 2, tendo assim uma árvore de 2 níveis abaixo da raiz.

Questão 2:

1. As diferenças entre os algoritmos de ID3 e C4.5 são:
 - a. C4.5 é uma extensão de ID3, criado para superar as limitações do ID3;
 - b. ID3 é mais sensível para números grandes, enquanto C4.5 é capaz de calcular esse tipo numérico.
 - c. As árvores de decisão do C4.5 são construídas através do treinamento de dados do ID3.
2. O algoritmo de C4.5 lida com a entrada numérica através do conceito de ganho de informação, na qual permite a medida do raio numérico, com uma equação do ganho e uma divisão da informação. O C4.5 escolhe um atributo a cada nó da árvore que seja mais efetivo de acordo com as equações citadas anteriormente, divide em amostras menores focadas em uma classe.