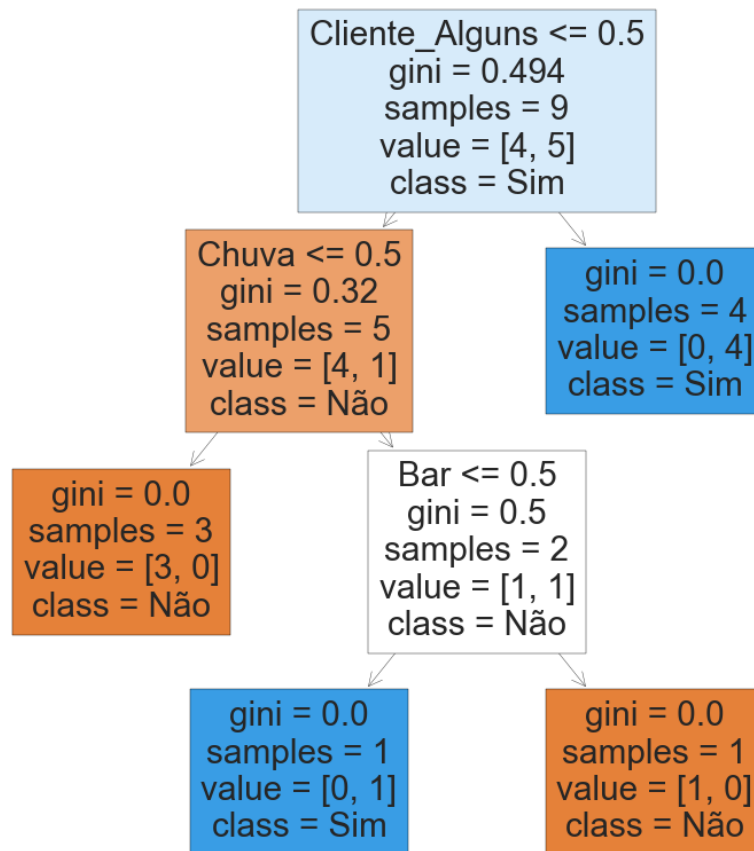


Nome: Edmar de Oliveira Melandes Junior

Questão 1

1.



2. Não houve ganho, manteve a mesma accuracy e precisão, porém a árvore resultante ficou diferente. Antes, era clientes na raiz, agora ficou tempo.
3. {'criterion': 'gini', 'max_depth': 2, 'max_features': 0.2, 'max_leaf_nodes': 4, 'min_samples_leaf': 2, 'min_samples_split': 2, 'random_state': None, 'splitter': 'best'}
Esses foram os parâmetros escolhidos pelo grid search

Questão 2

1. Em resumo são 4 grandes diferenças: o algoritmo C4.5 possui poda após a criação da árvore, removendo os ramos que não ajudam no resultado, consegue trabalhar com dados faltantes na entrada, lida com atributos contínuos além dos discretos e implementa a razão de ganho ao invés do ganho de informação.
2. Para lidar com os atributos contínuos, o C4.5 aplica uma discretização nesses atributos para convertê-los em discreto. Essa discretização consiste em criar pontos de corte e atribuir os intervalos sendo tratados como discreto, por exemplo um atributo "idade" que varia de 18 a 50 poderia ter um corte no 30 e criar os valores discretos "idade < 30" e "idade >= 30".