

### Trabalho - Árvore de Decisão

1. No programa feito em aula, usamos todos os exemplos para treinar o classificador de árvore de decisão. Mas podemos dividir o conjunto de exemplos em conjuntos de treinamento e teste usando ***sklearn.model\_selection.train\_test\_split***.

a. Divida o conjunto de exemplos em treinamento e teste das seguintes formas:

- i. 60% treinamento e 40% teste
- ii. 80% treinamento e 20% teste

Em cada um dos casos, gere a árvore de decisão correspondente e indique qual a acurácia obtida no treinamento e no teste. Compare as árvores obtidas e comente suas observações.

2. Utilizamos a entropia para escolher os atributos. Use o ***índice Gini*** (<https://medium.com/analytics-steps/understanding-the-gini-index-and-information-gain-in-decision-trees-ab4720518ba8>) ao invés da entropia. Você deve repetir o experimento considerando que o conjunto de exemplos não foi dividido em treinamento e teste, e considerando as divisões do exercício anterior. Indique a acurácia em cada caso (preencha a seguinte tabela). Compare as árvores obtidas quando usamos a entropia e o índice gini.

Treinamento / Teste	Entropia	Gini
100%-0%		
60%-40%		
80%-20%		

3. Podemos fazer o treinamento usando a validação cruzada k-fold ([https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.KFold.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html)). Aplique a validação cruzada para k = 3,8,14 gerando a seguinte tabela:

3-fold		8-fold		14-fold	
Treinamento	Teste	Treinamento	Teste	Treinamento	Teste
acurácia média	acurácia média	acurácia média	acurácia média	acurácia média	acurácia média

O que você pode concluir a partir dos resultados obtidos?