

Exercício - Árvores de Decisão

Neste exercício você irá colocar em prática seus conhecimentos em classificação com árvores de decisão utilizando o scikit-learn. Você irá encontrar o dataset disponível em: <https://www.kaggle.com/ronitf/heart-disease-uci>.

Utilizando o dataset acima, faça o que se pede:

- 1** - Carregue o dataset e realize uma análise exploratória dos atributos presentes no dataset. O que você consegue perceber através das suas análises? Os dados seguem algum padrão na sua distribuição? O atributo classe é balanceado?
- 2** - Realize o pré-processamento e divida o dataset em atributos (X) e classe (Y). Se necessário, escale X.
- 3** - Divida o dataset em conjuntos de treino (80%) e teste (20%).
- 4** - Crie e treine a sua árvore de decisão utilizando o **DecisionTreeClassifier** do scikit-learn.
- 5** - Faça previsões utilizando a árvore criada e obtenha as métricas de avaliação do modelo (kappa, acurácia, recall, precisão, etc). Plote também a matriz de confusão.
- 6** - Altere os valores da profundidade da árvore (**max_depth**) e da quantidade de samples para split (**min_samples_split**) na etapa de treinamento. Faça testes com diversos valores e avalie se existem quedas ou aumentos da acurácia após a alteração desses parâmetros.
- 7** - Refaça as etapas **4** e **5** utilizando os valores de parâmetros encontrados na etapa **6**. Plote as árvores obtidas.
- 8** - Observando os resultados obtidos antes e depois das alterações dos parâmetros, o que você consegue concluir sobre o seu modelo?