

### **Roteiro Aula 3 Lab**

1. Carregue o dataset breast cancer disponível no sklearn e obtenha a distribuição de classes deste dataset.
2. Use o classificador Vizinho Mais Próximo e uma divisão percentual com 30% dos exemplos do dataset breast cancer usados para teste para obter o valor da acurácia e f1 macro, f1 micro e f1 weighted deste classificador neste dataset.
3. Use o classificador Vizinho Mais Próximo e uma validação cruzada estratificada com 10 folds no dataset breast cancer para obter a matriz de confusão, a acurácia, o desvio padrão, o intervalo de confiança a 95%, a precision micro, o recall micro e a f1 macro deste classificador neste dataset.
4. Use o classificador K Vizinhos Mais Próximos com o método de validação cruzada com busca em grade (use 10 partições) no dataset breast cancer para realizar a busca pelo melhor valor de K. Apresente a melhor acurácia obtida, o desvio padrão, o intervalo de confiança a 95% e o valor de K correspondente.
5. Use o classificador K Vizinhos Mais Próximos no dataset breast cancer para obter seu desempenho estimando o melhor valor de K em cada partição da validação cruzada com 10 partições mas sem possibilidade de superajuste. Repita o procedimento por 3 vezes. Apresente a melhor acurácia obtida, o desvio padrão e o intervalo de confiança a 95%.
6. Repita o procedimento proposto no exercício 5 somente trocando o classificador para uma Árvore de Decisão. Compare o desempenho obtido pelo K Vizinhos Mais Próximos com a Árvore de Decisão utilizando o teste T e o de Wilcoxon.