

### Prática 3 – Validação Cruzada e Métricas de Avaliação

## Objetivo

Nesta prática, você irá utilizar a função ***cross\_validate*** da biblioteca scikit-learn para estimar o desempenho de diferentes algoritmos de classificação através de validação cruzada. A atividade propõe a comparação entre os resultados obtidos com validação cruzada e com uma simples divisão entre treino e teste.

## Descrição da Atividade

### 1. Modificar o Notebook da Prática 02:

- Acrescente no notebook da Prática 02 o cálculo da validação cruzada (com 10 folds) utilizando a função ***cross\_validate*** do sklearn e altere a métrica de avaliação para a **Macro F1**. Realize os experimentos com os classificadores [Árvore de Decisão](#), [Naive Bayes](#), [KNN \(fazer uma amostragem dos dados para que seja viável executar esse algoritmo\)](#) e [SVM \(usar o hyperparametro max\\_iter para que seja viável executar o SVC\)](#).

- Para cada classificador, calcule a Macro F1 estimada a partir: (a) da validação cruzada (com 10 folds) e (b) da divisão simples entre treino e teste.

- Escolha dois modelos, um baseado na validação cruzada e outro na divisão baseada apenas em *treino e teste*.

- Em seguida, compare qual modelo obteve o melhor score no **leaderboard** do Kaggle. Para isso, crie uma tabela como:

Algoritmo	Macro F1 Validação Cruzada Local (antes)	Macro F1 Treino e Teste (antes)	Kaggle (antes)	Macro F1 Validação Cruzada Local (ajustado)	Macro F1 Treino e Teste (ajustado)	Kaggle (escolhido Treino e Teste) (ajustado)	Kaggle (escolhido Validação Cruzada) (ajustado)
Árvore de Decisão							
Naive Bayes							
KNN							
SVM							

- Escreva de 3 a 5 linhas comentando os resultados

### 2. Entregar:

Entregue um Jupyter Notebook com todos os experimentos executados, comentários explicativos e a sua conclusão sobre o melhor modelo. O código deve estar organizado, com seções bem definidas e resultados claros. Dicas:

- Use ***cross\_validate***(..., **scoring=["f1\_macro"]**, **cv=10**).

- Para submissão no Kaggle, utilize o modelo treinado com os dados completos de treino.

- Comente seus resultados! A interpretação é tão importante quanto os números.