



# Bootcamp Data Analytics 2025

Challenge de Classificação





# Desafio de Classificação: Diagnóstico de Alzheimer com Dados Cognitivos

## Contexto

Este desafio utiliza o conjunto de dados OASIS (Open Access Series of Imaging Studies), que contém informações clínicas e demográficas de pacientes com potencial risco de Alzheimer. O objetivo é prever se o paciente apresenta sinais de demência.

A variável alvo é binária:

- 0: Sem demência
- 1: Com demência

## Dicionário de Dados

- Age: Idade do paciente (numérico)
- Sex: Gênero (F: feminino, M: masculino)
- EDUC: Anos de escolaridade (numérico)
- SES: Status socioeconômico (1 a 5)
- MMSE: Escore do Mini Exame do Estado Mental (0 a 30)
- CDR: Clinical Dementia Rating (0 a 3)
- eTIV: Volume intracraniano estimado
- nWBV: Proporção de volume cerebral normalizado
- ASF: Fator de escala anatômica
- Group (target): 0 = sem demência, 1 = com demência

## Perguntas

1. Faça uma análise exploratória dos dados e identifique os principais preditores da variável Group. Trate valores ausentes, codifique variáveis categóricas e normalize os dados, se necessário.
2. Treine um modelo de classificação (Regressão Logística, Árvore de Decisão ou Random Forest).
3. Avalie o modelo utilizando:
  - Matriz de confusão
  - Acurácia
  - Precisão
  - Revocação (Recall)
  - F1-score (opcional, mas recomendado)
4. Qual a importância relativa de cada variável no modelo final?
5. O modelo apresenta bom desempenho em ambos os grupos (com e sem demência)?
6. É possível identificar pacientes em estágio inicial de demência com esse modelo?
7. O modelo apresenta overfitting ou underfitting? Como isso pode ser ajustado?