

# Projeto AM 2025-2

Francisco de A. T. de Carvalho<sup>1</sup>

1 Centro de Informatica-CIn/UFPE  
Av. Prof. Luiz Freire, s/n -Cidade Universitaria, CEP 50740-540, Recife-PE, Brasil,  
*fatc@cin.ufpe.br*

## Questão 1

- Considere a base de dados "Yeast" (Arquivos anexados com o projeto, <https://archive.ics.uci.edu/dataset/110/yeast> )
- Use a ferramenta da sua escolha (scikit learn <https://scikit-learn.org/stable/>, WEKA <https://www.cs.waikato.ac.nz/ml/weka/>, etc) e use os 5 (cinco) classificadores abaixo:
  - i) Árvores de decisão; ii) Bayesiano ingenuo; iii) Regressão logística; iv) k-vizinhos (Use a distância Euclidiana para definir a vizinhança. Use conjunto de validação para fixar o número de vizinhos  $k$ ); v) Random forest.
- Separe de maneira aleatoria e estratificada (proporcional ao numero de exemplos de cada classe) os dados em treinamento, teste e, se necessário, validação. Sugestão: 70% para treinamento + validação e 30% para teste. Treine o classificador no conjunto de treinamento. Teste o classificador no conjunto de teste. Se o modelo possuir hiper-parametros, use validação cruzada para fixar os hiper-parametros. O conjunto de teste não pode ser usado no treinamento e na escolha dos valores dos hiper-parametros.

## Questão 1 (cont.)

Tentar melhorar cada um dos classificadores escolhido, procurando os melhores hiperparâmetros para alcançar a melhor precision (precisão), cobertura (recall) e f1-score no conjunto de teste.

- Para cada metrica de avaliação, plot a curva de aprendizagem para os classificadores com a melhor configuração de hiper-parametros. Mais precisamente, considere conjuntos de treinamento e teste de (5%, 95%) a (95%, 5%) do conjunto original de dados, com passo de 5% (usando amostragem estratificada). Para cada par de conjuntos de treinamento e teste, compute as metricas de avaliação em cada um desses conjuntos. Comente.

## Observações Finais

- No Relatório deve estar bem claro como foram organizados os experimentos de tal forma a realizar corretamente a avaliação dos modelos e a comparação entre os mesmos. Fornecer também uma descrição sucinta dos dados. No relatório mostrar os detalhes da obtenção dos hiper-parametros do modelo, se houver.
- Data de apresentação e entrega do projeto: **QUARTA-FEIRA 29/10/2025 a partir das 15:00 horas, sala E112.**
- Um dos membros da equipe: **colocar no google classroom o relatório do projeto, e os slides da apresentação. Me avisar por e-mail quando colocar.**
- **NÃO COLOCAR ARQUIVO ZIPADO**
- Tempo de apresentação: **10 minutos** para cada equipe (**rigoroso**).
- **A presença e a participação de todos os membros de cada equipe é obrigatória durante a apresentação;**