

# Teste Técnico Estag em IA

## Objetivo

Avaliar suas habilidades em análise de dados, aplicação de métodos de classificação e interpretação de resultados. Você trabalhará com um dataset fictício contendo informações de perfis de usuários de um site imobiliário e deverá prever se um usuário realizará a compra de uma casa ou não.

---

## Descrição do Problema

Você recebeu um dataset com informações sobre usuários que visitaram o site imobiliário. Sua tarefa é construir um modelo de classificação para prever se um usuário comprará uma casa (**Compra = 1**) ou não (**Compra = 0**), com base nas seguintes informações:

### Estrutura do Dataset

O dataset possui as seguintes colunas:

- **Idade:** Idade do usuário.
  - **Renda Anual (em \$):** Renda anual do usuário em dólares.
  - **Gênero:** Gênero do usuário (Feminino ou Masculino).
  - **Tempo no Site (min):** Tempo que o usuário passou navegando no site (em minutos).
  - **Anúncio Clicado:** Indica se o usuário clicou em um anúncio no site (Sim ou Não).
  - **Compra:** Indica se o usuário comprou uma casa (1 para sim, 0 para não). **Essa é a variável alvo.**
- 

## Tarefas

### 1. Análise Exploratória dos Dados

Realize uma análise inicial dos dados:

- Verifique a distribuição das variáveis (idade, renda, tempo no site, etc.).
- Explore possíveis relações entre as variáveis independentes e a variável alvo (**Compra**).
- Identifique valores ausentes ou inconsistências nos dados.

### 2. Pré-processamento dos Dados

- Normalizar ou padronizar variáveis numéricas, se necessário.

- Realizar codificação para variáveis categóricas, transformando-as em valores numéricos.
  - Dividir os dados em conjuntos de treino e teste.
  - 3. **Construção do Modelo de Classificação**
    - Treinar um modelo simples de classificação (como Regressão Logística, Árvore de Decisão, ou Random Forest).
    - Avaliar o modelo utilizando métricas apropriadas
  - 4. **Interpretação dos Resultados**
    - Identificar quais variáveis mais influenciaram na decisão do modelo.
    - Explicar o desempenho do modelo e propor possíveis melhorias.
- 

## Extras (Não Obrigatórios)

- Adicionar visualizações gráficas para complementar a análise.
  - Comparar o desempenho de diferentes modelos de classificação.
  - Usar técnicas de validação cruzada para uma avaliação mais robusta.
  - Explorar o impacto de hiperparâmetros no desempenho do modelo.
- 

## Requisitos Técnicos

1. A análise pode ser feita em Python ou R.
  2. Utilize bibliotecas como:
    - Python: `Pandas`, `NumPy`, `Scikit-learn`, `Matplotlib`, `Seaborn`.
    - R: `dplyr`, `ggplot2`, `caret`.
  3. O projeto deve incluir:
    - O código-fonte usado.
    - Um relatório em PDF ou Markdown explicando os resultados e as conclusões.
- 

## Entrega

- Envie um repositório público no GitHub contendo:
    - O código-fonte do projeto (Jupyter Notebook ou script).
    - O dataset utilizado.
    - Um relatório explicando os resultados obtidos.
- 

## Avaliação

1. **Análise Estatística:** Qualidade da exploração dos dados.
2. **Modelo:** Implementação correta de um modelo básico de classificação.
3. **Interpretação:** Clareza nas explicações e justificativas das escolhas.
4. **Extras:** Implementações adicionais ou insights inovadores.