

Teste Técnico — Cientista de Dados Sênior (Duração Máxima: 3 Horas)

Introdução

Este teste técnico avalia a capacidade do candidato em resolver um problema de dados de ponta a ponta em um período limitado. O objetivo é demonstrar proficiência em manipulação de dados, modelagem preditiva e comunicação de resultados, simulando um cenário de projeto ágil e focado.

1) Objetivos

- **Preparação de Dados:** Demonstrar habilidades em limpeza, transformação e engenharia de features.
- **Modelagem Preditiva:** Construir e avaliar um modelo de Machine Learning para um problema específico.
- **Análise e Comunicação:** Interpretar resultados e apresentar insights de forma clara e concisa.

2) Cenário do Problema: Previsão de Churn de Clientes

Uma empresa de telecomunicações deseja prever quais clientes estão propensos a cancelar seus serviços (churn) para implementar estratégias de retenção proativas. Você receberá um dataset simplificado contendo informações de clientes e seu status de churn.

3) Requisitos

- **3.1) Preparação de Dados (45 minutos)**

- Carregue o dataset fornecido (simulado ou pequeno dataset público, ex: [Telco Customer Churn](#)).
- Realize a limpeza de dados, tratando valores ausentes e inconsistências.
- Execute engenharia de features básicas (ex: one-hot encoding para variáveis categóricas).
- **3.2) Modelagem Preditiva (90 minutos)**
 - Divida os dados em conjuntos de treino e teste.
 - Treine um modelo de classificação (ex: Regressão Logística, Random Forest, Gradient Boosting) para prever o churn.
 - Avalie o desempenho do modelo usando métricas apropriadas (ex: Acurácia, Precisão, Recall, F1-Score, AUC-ROC).
- **3.3) Análise e Comunicação (45 minutos)**
 - Analise os resultados do modelo, identificando as features mais importantes.
 - Prepare um breve resumo (máximo 1 página) dos seus achados, incluindo a performance do modelo e insights sobre os fatores que influenciam o churn.

4) Entregáveis

- **Código:** Um script Python (`main.py`) ou Jupyter Notebook (`notebook.ipynb`) contendo todo o código-fonte para a preparação de dados, modelagem e avaliação.
- **Relatório:** Um arquivo Markdown (`REPORT.md`) ou PDF (`REPORT.pdf`) com o resumo da análise e insights.

5) Requisitos Técnicos

- **Linguagem:** Python 3.8+
- **Bibliotecas:** Pandas, Scikit-learn, Matplotlib/Seaborn (opcional para visualização).

6) Critérios de Avaliação

- **Corretude:** A solução funciona conforme o esperado e atende a todos os requisitos.
- **Qualidade do Código:** Clareza, organização, comentários e boas práticas de programação.
- **Análise:** Profundidade dos insights e interpretação dos resultados.
- **Comunicação:** Clareza e concisão na apresentação dos resultados.
- **Eficiência:** A solução é executada dentro do tempo limite (3 horas).

7) Dataset (Exemplo - usar um dataset público ou simulado)

- **Nome:** Telco Customer Churn (Kaggle)
 - **Link:** <https://www.kaggle.com/blastchar/telco-customer-churn>
-