



## Querido Participante

Você está prestes a embarcar em um desafio que vai testar suas habilidades técnicas, seu pensamento analítico e sua capacidade de transformar dados em decisões estratégicas. No mundo real, prever o comportamento dos clientes é um diferencial competitivo essencial para qualquer empresa. Sua missão aqui é desenvolver um modelo de aprendizado de máquina capaz de identificar quais clientes possuem maior risco de cancelamento, permitindo que a empresa tome medidas preventivas e aumente a retenção.

Este desafio não se trata apenas de programar um modelo preditivo, mas de compreender o problema, preparar os dados de maneira inteligente e apresentar soluções. Queremos ver como você pensa, estrutura seu código e comunica suas ideias.

Dê o seu melhor, explore diferentes abordagens e, acima de tudo, aproveite o processo! Boa sorte!

## Introdução

A rotatividade de clientes (churn) é um dos problemas mais críticos enfrentados por empresas de diversos setores. Identificar antecipadamente quais clientes têm maior probabilidade de cancelar seus serviços permite que as empresas tomem medidas proativas para retê-los. Neste desafio, você desenvolverá um modelo de Machine Learning supervisionado para prever a rotatividade de clientes com base em dados históricos.

## Descrição do Cenário

A TelecomPlus, uma empresa fictícia de telecomunicações, está enfrentando uma alta taxa de cancelamento de serviços. O CEO está preocupado com o impacto financeiro e solicitou à equipe de dados que desenvolva uma solução para identificar clientes com alto risco de cancelamento antes que eles tomem essa decisão.

Você foi contratado como cientista de dados para desenvolver um modelo preditivo que identifique clientes com maior probabilidade de cancelar seus serviços nos próximos 3 meses. A empresa fornecerá um conjunto de dados históricos contendo informações sobre clientes que permaneceram e que cancelaram seus serviços.

## Objetivos de Aprendizado

- Dominar o processo completo de desenvolvimento de um modelo supervisionado, desde a preparação dos dados até a avaliação.
- Aplicar técnicas de pré-processamento de dados e engenharia de features.
- Implementar e comparar diferentes algoritmos de classificação.

Interpretar resultados e métricas de modelos para tomada de decisão de negócio.

## Requisitos Funcionais

- Desenvolver um pipeline de pré-processamento de dados que inclua, caso necessário:
  - Tratamento de valores ausentes
  - Codificação de variáveis categóricas
  - Normalização/padronização de features numéricas
  - Engenharia de features para criar variáveis derivadas relevante.
- Implementar pelo menos dois algoritmos de classificação diferentes (ex: Regressão Logística, Random Forest, XGBoost) e comparar seus desempenhos.
- Desenvolver uma análise de feature importance para identificar os principais fatores que influenciam o cancelamento de clientes.

## Requisitos Técnicos

- Utilizar Python como linguagem de programação principal
- Implementar o projeto usando bibliotecas de ciência de dados (pandas, scikit-learn, numpy)
- Documentar o código com comentários explicativos
- Utilizar Notebook, como formato de entrega
- Subir seu projeto de forma pública no GitHub

## Estrutura da Entrega

- Escrever um README.md para explicar as etapas feitas e insights adquiridos durante o desenvolvimento do case.
- Estrutura Exemplo de Notebook: Exploração de dados, Pré-processamento, insights/hipóteses, treino, teste e comparação de modelos.
- O treino será feito na "dados\_clientes.csv" e o teste em "desafio.csv".
- Seu Notebook deve criar, ao fim do processo, um arquivo CSV com nome: "resultado\_nome\_sobrenome.csv", contendo apenas as colunas Id e Target ('churn'), com os resultados.
- Nós vamos avaliar a acurácia fazendo uma comparação do seu CSV de resultado, feito a partir do "desafio.csv", com outro CSV contendo as classificações corretas de cada linha.
- A entrega deve ser feita a partir de um link de um repositório público no GitHub. Ele deve conter o notebook desenvolvido, os CSV de teste e treino e o de resultado.

## CrITÉrios de Avaliação

- **Cumprimento dos requisitos funcionais e técnicos:** É preciso que seja cumprido todos os requisitos funcionais e técnicos definidos previamente.
- **Performance do modelo:** O modelo final deve atingir no mínimo 70% de acurácia.
- **Qualidade da Documentação:** Clareza na documentação e explicação do processo de criação do modelo e identificação dos fatores mais importantes para a previsão de churn.
- **Cumprimento da estrutura da entrega:** As etapas mínimas definidas no tópico "Estrutura da Entrega", foram apresentadas.