Previsão de Churn de clientes em empresa de Telecom

Diego Martins Faria
Cientista de dados

Introdução ao problema

O que é churn?

Churn é o evento em que o cliente encerra a relação ativa com a empresa. Em outras palavras, churn é quando se perde o cliente.

Por que é um problema crítico para o negócio?

Porque o churn impacta a receita futura da empresa, custo de aquisição(trazer clientes novos é mais caro que manter os atuais), impacta a imagem da marca, já que cliente insatisfeito fala mal. Enfim, impacta a saúde da empresa como um todo

Objetivo do projeto:

Desenvolver um modelo de machine learning capaz de prever clientes com maior chance de cancelamento

Objetivos

- Detectar padrões que indiquem churn
- Ajudar time de marketing a agir preventivamente
- Reduzir a taxa de cancelamento de clientes

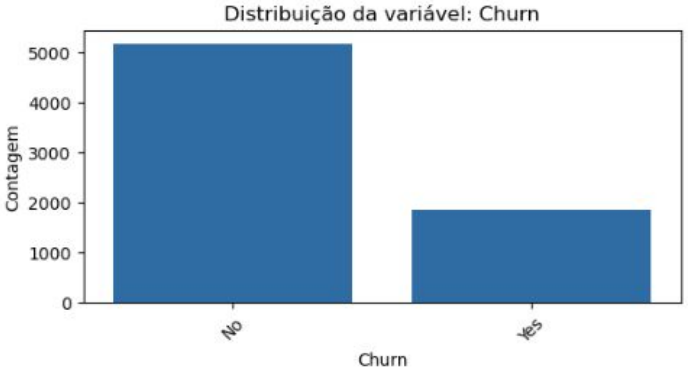
Conhecendo os dados

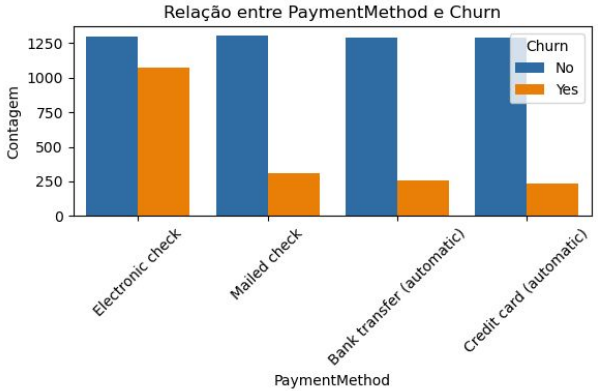
Dicionário de dados

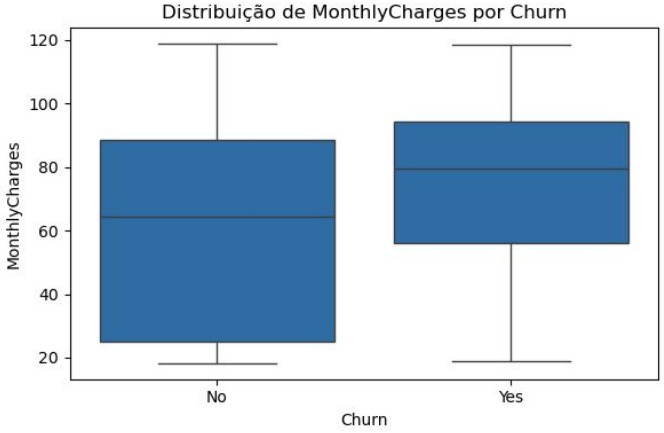
Variável	Descrição	Tipo
customerID	Identificador único do cliente	Categórica
gender	Gênero do cliente (Female/Male)	Categórica
SeniorCitizen	Indica se o cliente é idoso (1) ou não (0)	Numérica
Partner	Indica se o cliente possui parceiro(a) (Yes/No)	Categórica
Dependents	Indica se o cliente possui dependentes (Yes/No)	Categórica
tenure	Meses de permanência do cliente	Numérica
PhoneService	Indica se o cliente possui serviço telefônico (Yes/No)	Categórica
MultipleLines	Indica se o cliente possui múltiplas linhas telefônicas	Categórica
InternetService	Tipo de serviço de internet (DSL/Fiber optic/No)	Categórica
OnlineSecurity	Indica se o cliente possui segurança online (Yes/No/No internet service)	Categórica
OnlineBackup	Indica se o cliente possui backup online (Yes/No/No internet service)	Categórica
DeviceProtection	Indica se o cliente possui proteção de dispositivo (Yes/No/No internet service)	Categórica
TechSupport	Indica se o cliente possui suporte técnico (Yes/No/No internet service)	Categórica
StreamingTV	Indica se o cliente possui serviço de streaming de TV (Yes/No/No internet service)	Categórica
StreamingMovies	Indica se o cliente possui serviço de streaming de filmes (Yes/No/No internet service)	Categórica
Contract	Tipo de contrato do cliente (Month-to-month/One year/Two year)	Categórica
PaperlessBilling	Indica se o cliente utiliza faturamento sem papel (Yes/No)	Categórica
PaymentMethod	Método de pagamento do cliente	Categórica
MonthlyCharges	Valor cobrado mensalmente do cliente	Numérica
TotalCharges	Valor total cobrado do cliente	Numérica
Churn	Indica se o cliente deixou a empresa (Yes/No)	Categórica

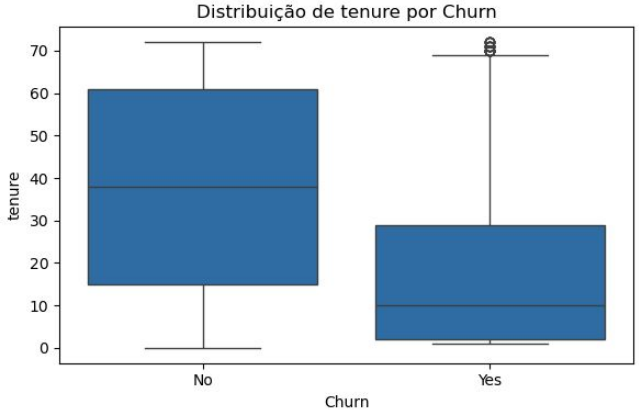
- O dataset tem 7043 linhas e 21 colunas

Análise Exploratória (EDA)









Análise Exploratória

Observamos nos gráficos anteriores que:

- Clientes com cobranças mensais mais altas, tendem a churn
- Métodos de pagamento eletrônicos têm mais churn se comparados com cobranças automáticas
- Clientes novos tendem a churn

Pré-processamento

- Retirei a coluna ‘customerID’ por não representar nenhum dado importante para o modelo
- Fiz o encoding das variáveis categóricas

Separação dos Dados

- **Fiz a separação em treino, teste e validação, ficando:**
  - Treinamento: (4500, 40), Validação: (1125, 40), Teste: (1407, 40)

- **Identifiquei também que havia desbalanceamento nos dados, o que poderia influenciar nosso modelo.**

Modelo Base

- **Random Forest Classifier**
- **Principais métricas:**
  - Acurácia: **79%**
  - Recall da classe 1 (churn): **48%**

Classification Report

	precision	recall	f1-score	support
0	0.82	0.90	0.86	822
1	0.64	0.48	0.55	303
accuracy			0.79	1125
macro avg	0.73	0.69	0.70	1125
weighted avg	0.77	0.79	0.78	1125

- **Limitações:**
  - Baixo recall
  - Classes desbalanceadas

Estratégias para Melhorar

- Ajustei o hiperparâmetro class_weight do modelo, mas não obtive resultados significativos

- Fiz ajustes com threshold, sem resultados significativos

- Solução:
  - Usar outro modelo (XGBoost), por ser mais robusto a classes desbalanceadas

Resultados Finais com XGBoost

- **Principais métricas:**
  - Acurácia: **75%**
  - Recall da classe 1 (churn): **84%**

Classification Report

	precision	recall	f1-score	support
0	0.93	0.72	0.81	822
1	0.53	0.84	0.65	303
accuracy			0.75	1125
macro avg	0.73	0.78	0.73	1125
weighted avg	0.82	0.75	0.77	1125

Preferimos maximizar o recall para evitar churns não detectados

Conclusões

- Modelo alcançou bom desempenho para detectar clientes em risco
- Business value: ação preventiva para retenção de clientes
- Recomendações para o time:
  - Focar em clientes com alto risco
  - Melhorar canais de comunicação para clientes insatisfeitos

Próximos Passos

- Colocar modelo em produção
- Monitoramento contínuo do desempenho
- Integração com times de marketing e atendimento

Obrigado!

LinkedIn: https://www.linkedin.com/in/diego-martins-faria/

Github: https://github.com/diegomartf