

Previsão de Churn

Contexto

A atualidade caracteriza-se pela crescente competitividade dos mercados, que acelera a banalização dos produtos e inibe a diferenciação, tão ambicionada pelas marcas junto do consumidor.

Além deste clima de concorrência, o impacto a longo prazo de uma reputação prejudicada, o abandono por parte de um cliente e as despesas para o readquirir - relacionadas com esforços adicionais em vendas, marketing e publicidade - somam uma perda significativa, em termos de receita e reputação. É sabido que conquistar novos clientes pode custar cinco vezes mais do que manter os existentes, e uma grande parte das empresas já interiorizou esta realidade.

Perante este cenário, onde é mais caro atrair novos clientes do que manter os atuais, urge aproveitar as oportunidades geradas pelas novas tecnologias com os dados que temos ao nosso dispor, para conhecer cada vez melhor o perfil dos nossos clientes, prever o seu comportamento num futuro próximo e antecipar a criação de campanhas estratégicas de retenção. Desta forma, a tarefa de prever e prevenir o abandono de clientes (o chamado *churn*) reveste-se de uma importância assinalável para o nosso negócio.

Informações adicionais

Consideramos que um cliente “faz *churn*” quando encerra todas as suas contas à ordem num determinado mês, e por isso, deixa de ter qualquer vínculo com o Banco a partir desse momento. Neste caso, o *churn* do cliente é definido como o fecho das suas contas num intervalo de três meses após o período transaccional disponível.

A informação disponibilizada encontra-se desagregada por transação. No decorrer do exercício, deverá ser agregada por cliente de uma forma que seja explicativa para a previsão de *churn*, uma vez que este evento é marcado ao nível do cliente. Nenhum dos Clientes abandonou o Banco durante o período das transações disponibilizadas.

Objetivos

Os objetivos do desafio são:

1. Contribuir para a previsão do *churn* do cliente, i.e., que um dado cliente irá fechar a sua conta;

2. Discutir a aplicação do modelo no dia-a-dia do Banco e estudar o seu possível impacto no negócio.

Resultados esperados

São esperados os seguintes resultados:

- A identificação dos fatores transacionais que podem influenciar a previsão do *churn* de cada cliente;
- Um modelo preditivo validado que permita prever se, potencialmente, irá ocorrer *churn* para cada cliente;
- Uma (ou mais) ideia(s) de aplicabilidade da solução e quais as vantagens que esta traz ao negócio do banco.

Recursos disponíveis

As equipas terão acesso ao dataset de transações no ficheiro *churn_train.csv* com os seguintes campos:

- ID do cliente;
- Montante da transação: negativa se transação a débito, positiva se transação a crédito;
- Número da categoria da transação;
- Descrição da categoria da transação (e.g. “Compras e Serviços”, “Levantamentos”, “Combustível”);
- Data da transação;
- Variável target. Apesar de o dataset estar desagregado em transações, esta variável está construída por cliente:
 - o 1 - identifica clientes que no espaço de 3 meses deixaram de ter qualquer conta aberta.
 - o 0 - identifica os clientes que continuaram a ter a(s) sua(s) conta(s) aberta(s)

O ficheiro *churn_train.csv* está compreendido entre janeiro e abril de 2022 (período das transações). Também será fornecido o ficheiro *churn_test.csv* que deverá ser utilizado para efetuar as *predictions* pelos participantes. Todos os ficheiros podem ser encontrados [aqui](#).

As equipas participantes podem ainda recorrer e/ou utilizar outra informação, desde que esta seja aberta, disponível e legal.

Entregáveis

É esperado que o participante entregue:

1. As previsões resultantes usando o dataset *churn_test.csv*;

2. Elaboração de uma apresentação (.ppt/PDF) e de um *pitch* com o máximo de 5 minutos

Avaliação

Todos os critérios são avaliados numa escala de 0 a 3 e são distribuídos da seguinte forma:

- 45% Critério técnico:
 - F1 Score obtido no ficheiro de testes * 3
 - Nota: o F1 é calculado com o parâmetro *average* definido como *macro*.
- 40% Critério de negócio:
 - A solução está enquadrada no problema;
 - A equipa atinge os resultados esperados;
 - A solução inclui ideias inovadoras para aplicar no negócio;
 - A equipa apresenta ideias de trabalho futuro para o desenvolvimento da solução;
- 15% Critério de comunicação:
 - A solução foi apresentada de uma forma explícita e coerente.

Apenas as 10 melhores equipas com o melhor F1 Score irão apresentar a sua solução ao júri.

Processo de Entrega

Upload da apresentação em formato *ppt* ou *pdf* e o ficheiro *churn_predictions.csv* com a coluna *prediction* preenchida através deste [formulário](#) da Google.

Dicas para o desafio

Recomendamos:

- Dada a dificuldade em trabalhar com os dados desagregados, começar a agregar a informação ao nível do cliente: por exemplo, construir variáveis para o número de transações, montante total, rácio do montante pelo número de transações, etc (por categoria ou não).
- A partir da análise de dados e/ou do modelo preditivo, identificar as variáveis mais importantes na previsão de *churn*. Se foram identificadas, apresentar a análise executada e sugerir ações de prevenção e para retenção de clientes atempadas.