

<MAKE IT>



<LGBM in Churn Prediction>

<APHRODITE>

Make IT | Universidade de Lisboa

by **Millennium**
bcp

Problema

Descrição:

- Perda de clientes e não haver identificação do porquê nem os padrões para conseguir evitar a perda dos mesmos.
- Custa cerca de 7 vezes mais encontrar um cliente novo do que mantê-lo.
- O impacto é perder o mínimo de clientes possíveis e maximizar assim o lucro.
- O nosso problema passa por fazer classificação (binária), ou seja atribuir como '0' ou '1', sobre um conjunto de clientes, em que '0' significa que o cliente não saiu (não fez churn) e '1' significa que o cliente saiu (fez churn).
- Para avaliar a situação, tínhamos um conjunto de clientes com uma série de transações por categoria ao longo de 4 meses.

Abordagem

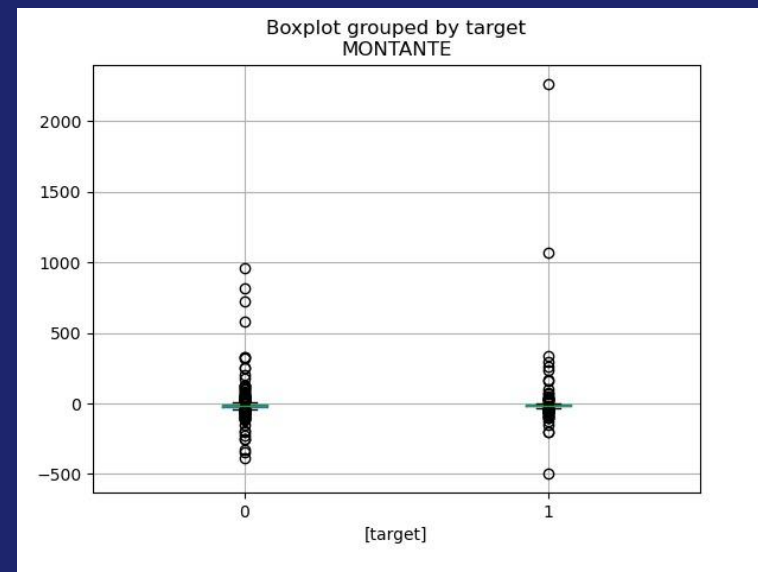
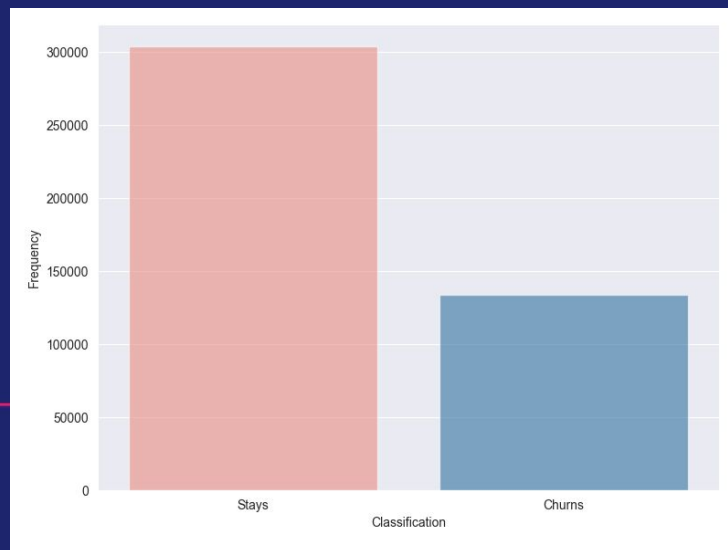
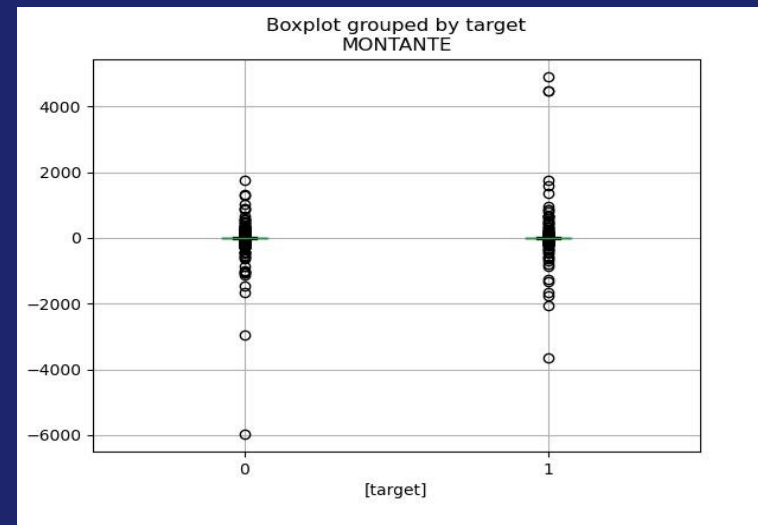
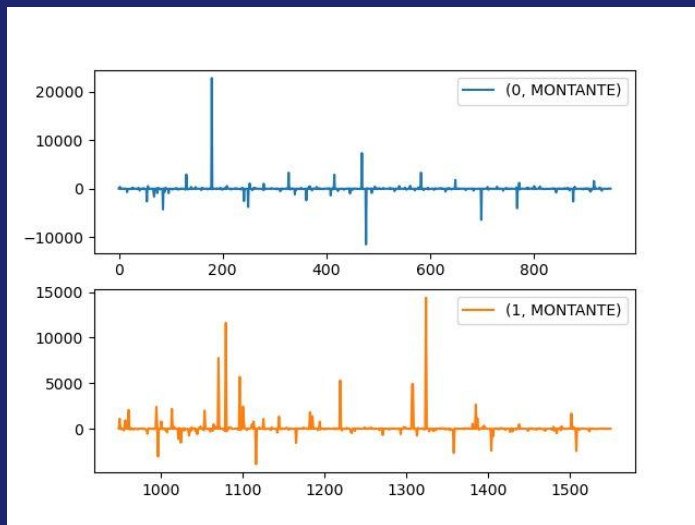
Descrição:

- Fazer análise estatística e EDA do dataset.
- Encontrar as features que têm mais impacto no facto de haver “churn”.
- Testar vários modelos que se enquadram com o problema.
- Escolher o que têm melhores resultados.

Problemas do dataset

- Dados não balanceados por churn: Não houve balanceamento dos dados.
- Dados repetidos: Eliminamos os dados duplicados.
- Outliers: não fizemos a verificação
- Falta de features: criamos features a partir de uma análise estatística.

EDA



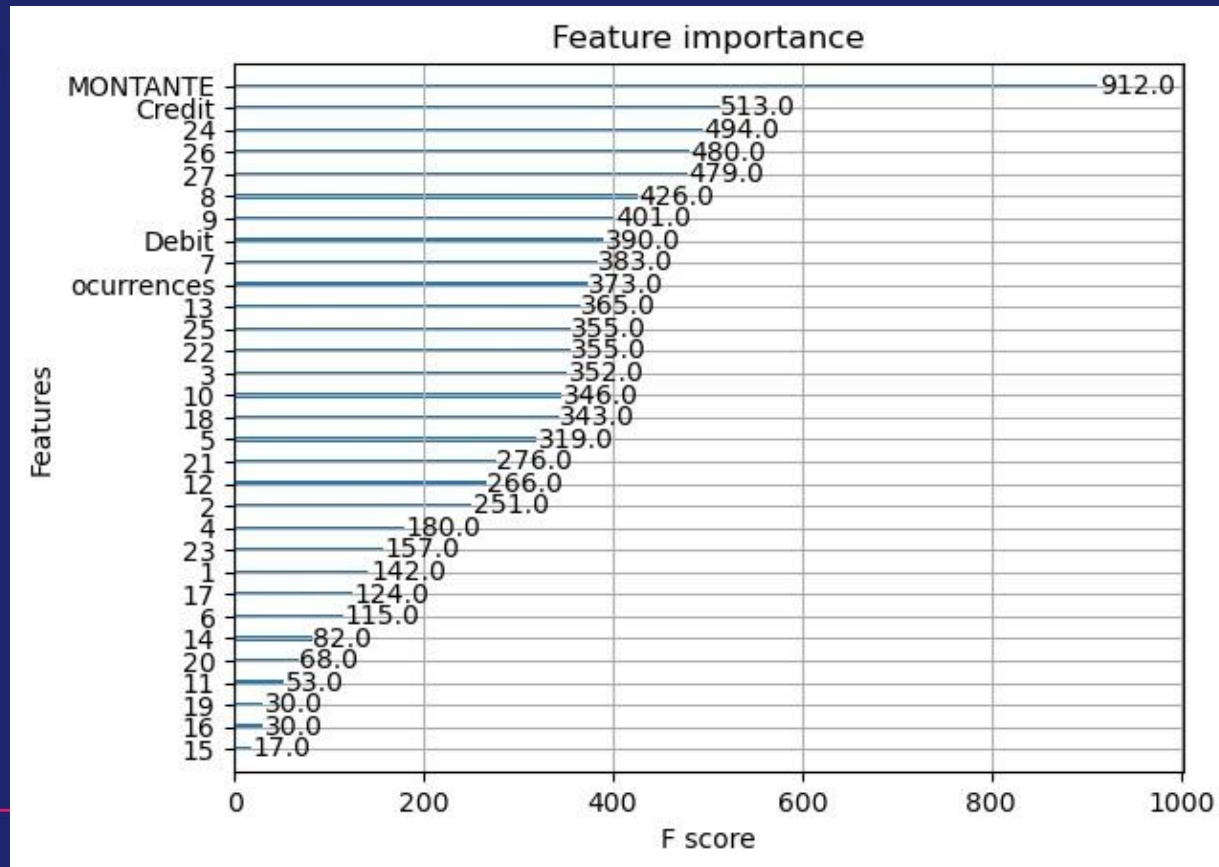
◀MAKE IT▶

Solução/Resultados

Descrição:

- Inicialmente, limpamos os dados repetidos do dataset.
- Posteriormente, criou-se um novo dataset onde cada linha representa um diferente cliente, com 31 features diferentes.
- Fez-se one-hot encoding de um total de 27 categorias (27 features)
- Adicionar também como features: montante final, número de transações em crédito, número de transações em débito e quantidade total de transações
- Aplicamos 3 modelos diferentes: naive bayes classifier, KNN e LightGBM

Variáveis mais relevantes



1. Montante
2. Credit
3. Taxas & Impostos
4. Levantamentos
5. Não Categorizadas

Top Performer

Light GBM

	precision	recall	f1-score	support
0	0.73	0.79	0.76	250
1	0.57	0.49	0.53	141

Quais foram as nossas limitações?

Falta de dados - Dados limitados ao nível de informação (falta de dados pessoais).

Categorias que não trazem informação ao problema (“Eg: Não categorizadas”).

Dificuldades em definir a segmentação dos clientes para interceptar os seus interesses.

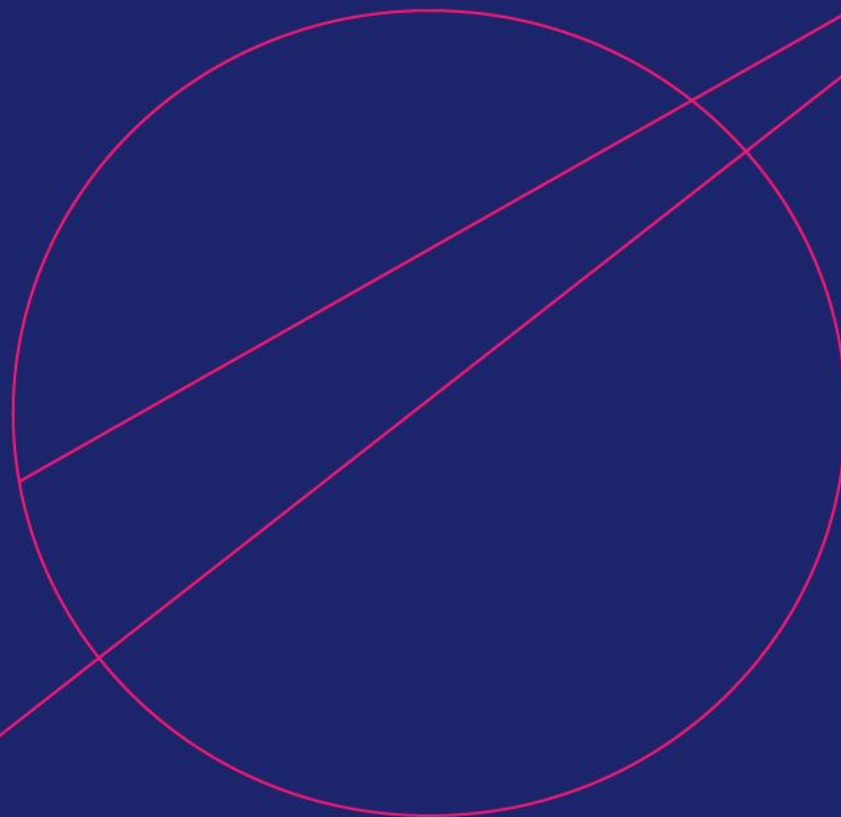
Fazer uma análise estatística mais extensa dos clientes.

Falta de uma equipa maior e mais estruturada com um plano bem definido.



Ideias de trabalho futuro

- Detetar e tratar de outliers.
- Balancear dados.
- Testar mais modelos.
- Fazer mais inferência estatística.



<MAKE IT>

Aplicação no negócio

Segmentar os clientes e detectar em que áreas é que o dinheiro é gasto.

Detetar padrões de churn o mais rápido possível.

Focar nos clientes ou na segmentação que mais dinheiro trazem ao banco.

Encontrar parcerias com instituições e serviços que tragam benefícios aos clientes do banco.

Definir descontos e benefícios que pelo menos igualem a concorrência.

Comunicar constantemente com o cliente para perceber quais as suas preocupações e utilizá-las para completar as a inferência estatística.



◀MAKE IT▶

>OBRIGADO :)

◀MAKE IT▶