

Minimizando o Risco de Default na X-Health

Fernando Kuriike Matsumoto





Sumário

01

Visão Geral

02

Insights

03

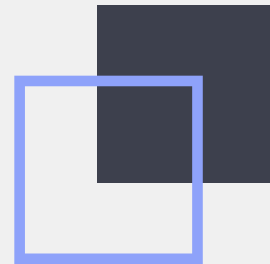
O Modelo

04

Resultados

05

Conclusão



01

Visão Geral

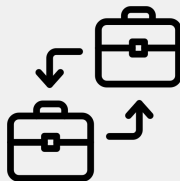


X-Health

- Venda de dispositivos eletrônicos na área da saúde



- Modelo de negócios B2B



- Ampla gama de clientes e produtos, com valores e formas de pagamento variadas



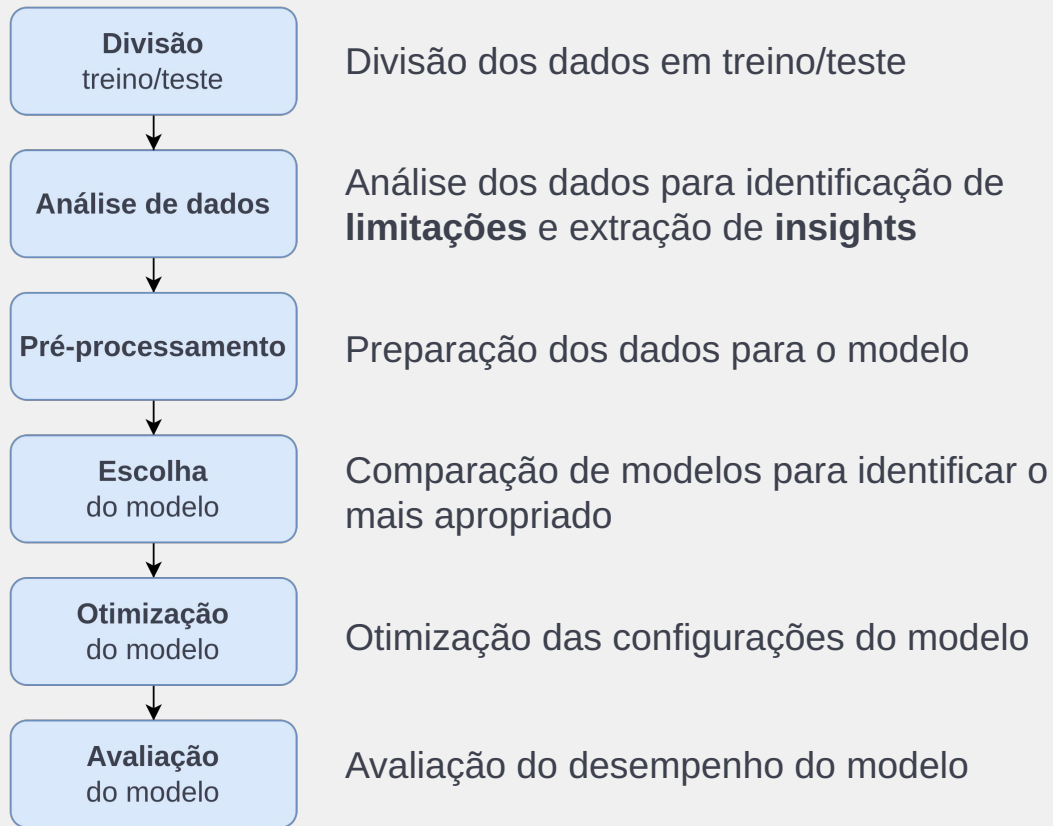
0 Problema

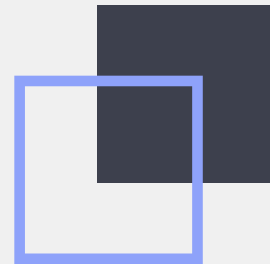
- Grande número de **defaults**: pagamentos não realizados
- Grande **impacto financeiro** na X-Health
 - Projetos realizados, mas sem nenhuma receita
 - Projetos que não puderam ser realizados por falta de capacidade
- Objetivo:
 - Desenvolver uma solução para **identificar clientes com alto risco de default**
 - Minimizar perdas financeiras, viabilizando o crescimento da empresa

Os dados

- Relativos a pedidos realizados de 2017 a 2019
- Variáveis internas, relativas ao **histórico do cliente** e ao **pedido** em questão
- Variáveis externas, provenientes do **Serasa** e relativas a dívidas, protestos, etc.
- Diversos problemas que serão analisados adiante
- Dados desbalanceados, apenas 17% de defaults

Abordagem





02

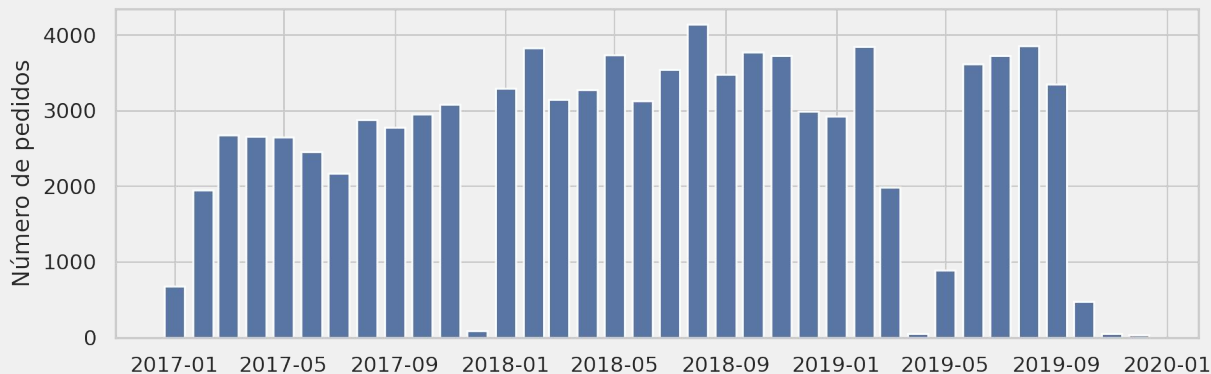
Insights



Problemas na coleta de dados

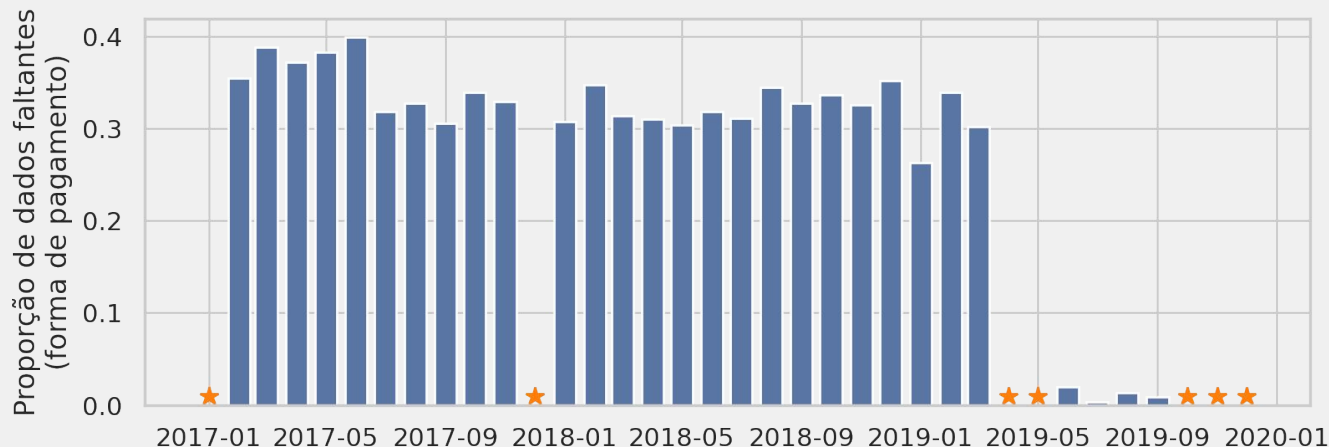
- Problemas nos dados do Serasa:
 - Diversos valores inconsistentes, eg. clientes com:
- Dados faltantes (variáveis internas):
 - 16% sem opção tributária
 - 28% sem forma de pagamento
- Alguns meses com muito poucos dados, em especial os últimos meses:

Quantidade de ações judiciais: 0
Valor das ações judiciais: R\$ 1500,00



Problemas na coleta de dados: melhorias

- Grande melhoria nos dados de forma de pagamento recentemente
- Hipótese: mudanças no processo de entrada de dados?



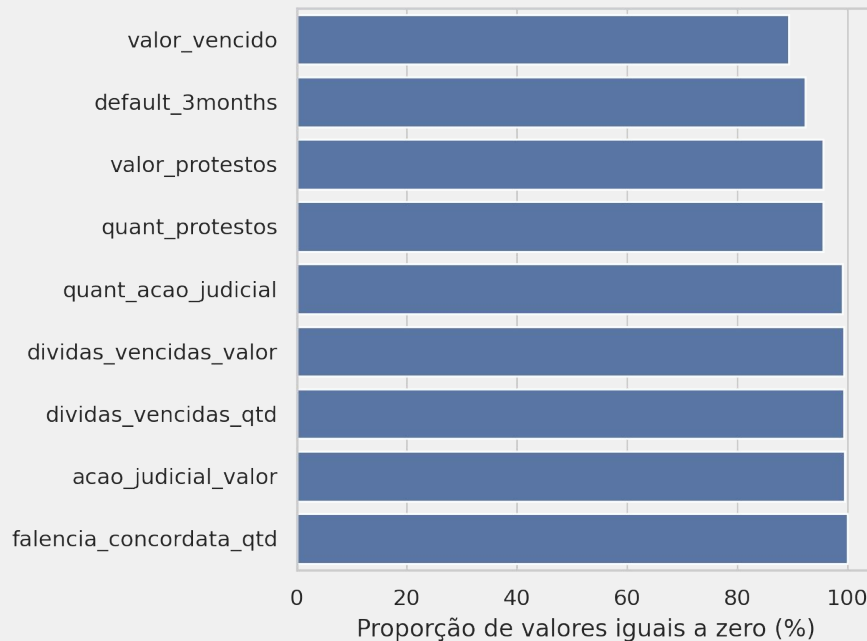
* Estrelas indicam meses nos quais poucos dados foram coletados

Indicadores de Inadimplência

- Clientes em geral cumprem com os pagamentos
- Indicadores de inadimplência quase sempre iguais a zero
- Dados não muito úteis, mas:

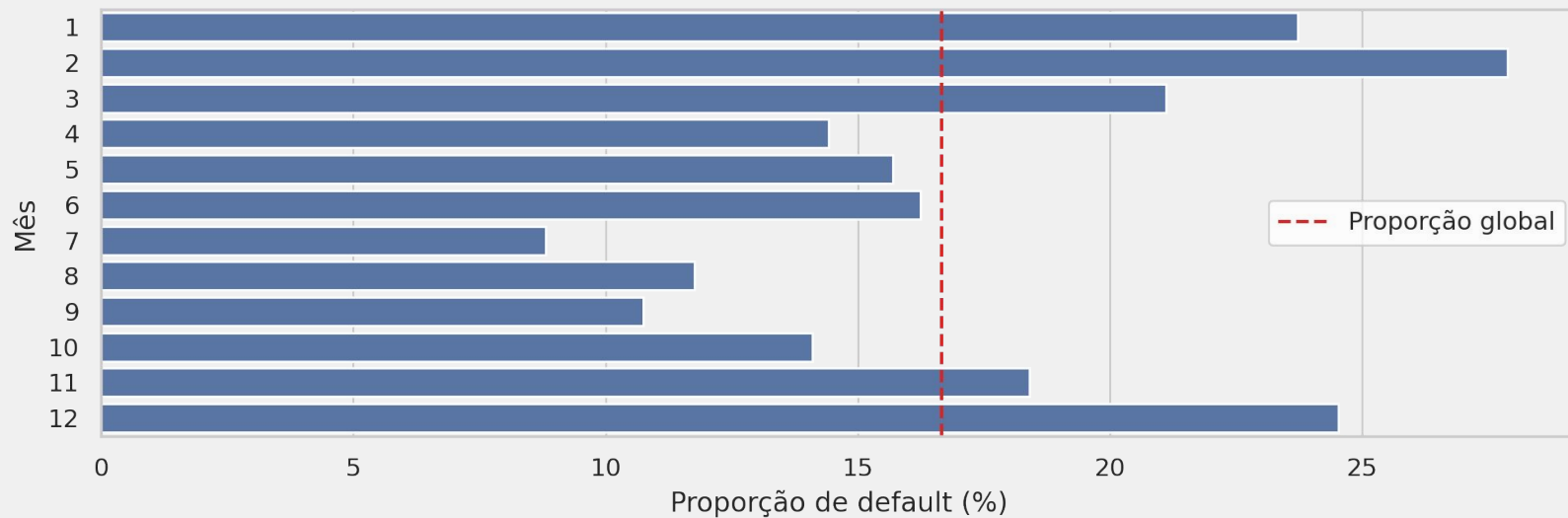
A quantidade de indicadores iguais a zero é um bom indicador de risco

- Correlação: -56%



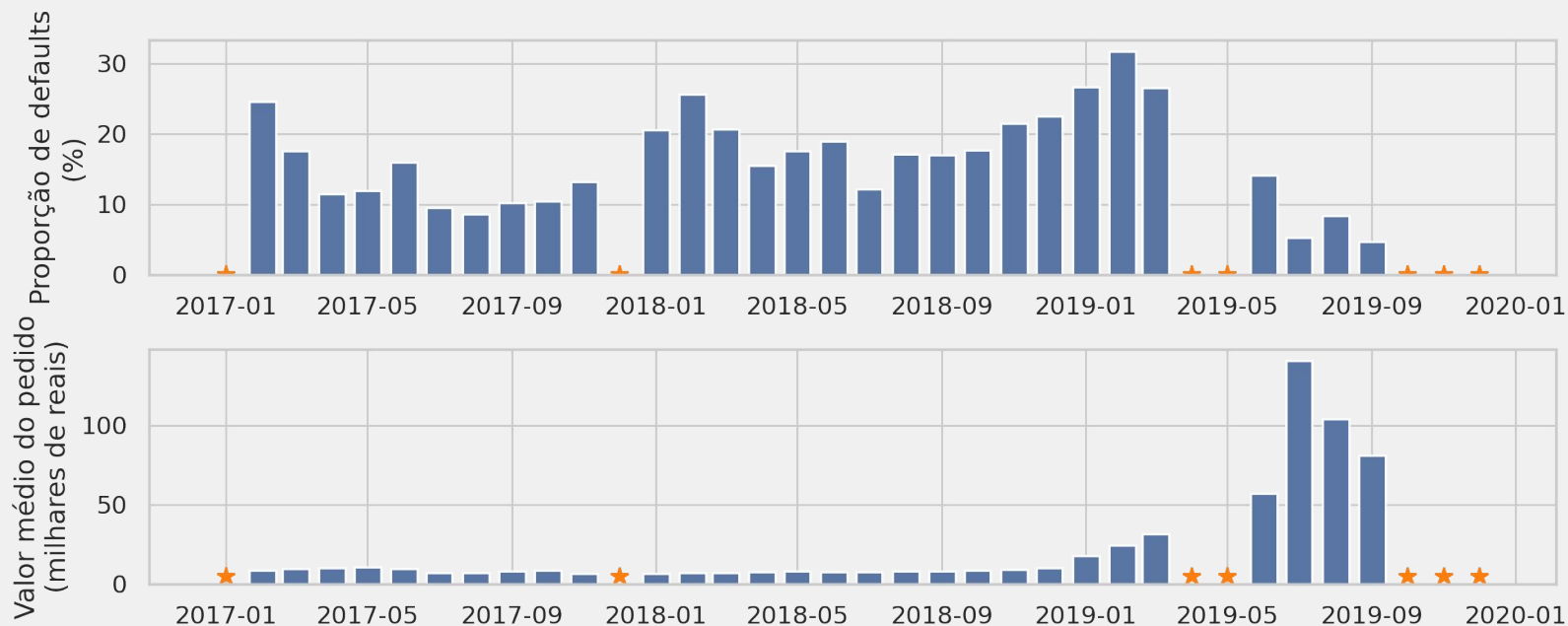
Efeito Sazonal

- Menos defaults em pedidos realizados no meio do ano



Mudanças no final de 2019

- Redução de defaults
- Crescimento no valor dos pedidos



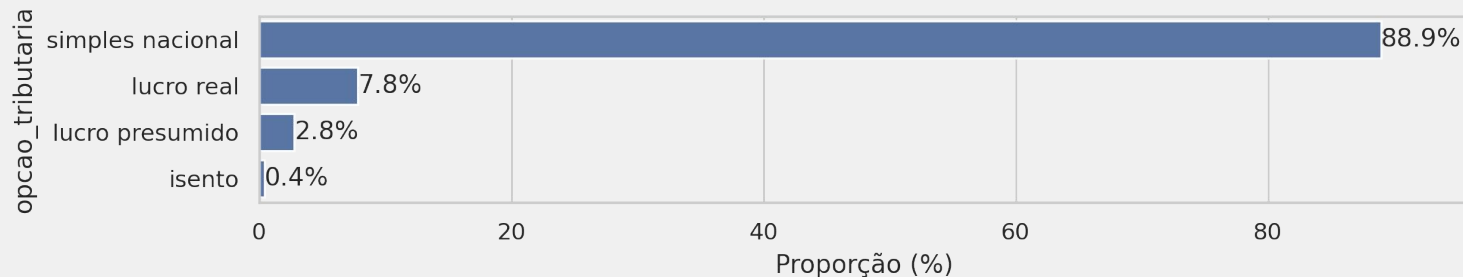
Variáveis categóricas

- Desbalanceadas, algumas categorias dominam...

Tipo da sociedade: sociedade empresaria limitada, empresario (individual), ...

Opção tributária: simples nacional, lucro real, ...

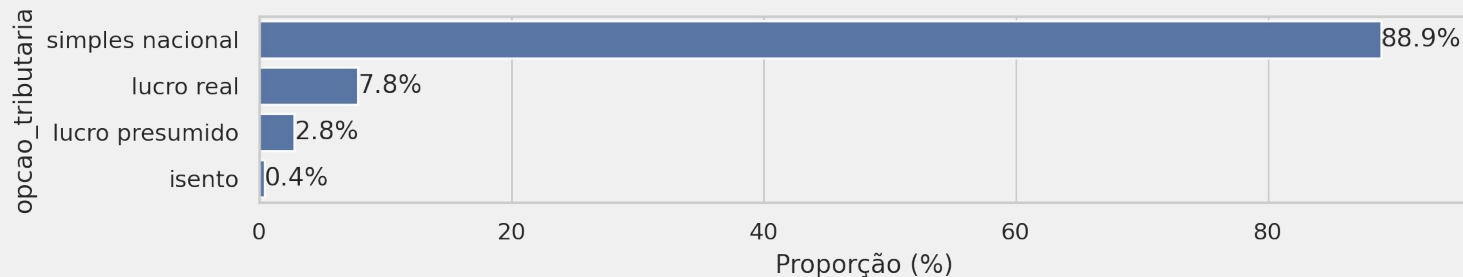
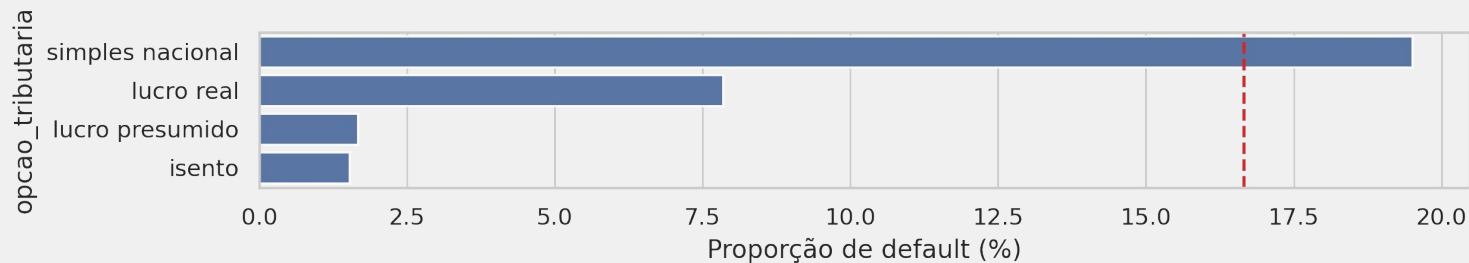
Atividade principal: comunicação de equipamentos de informática, papelaria, ...



Variáveis categóricas

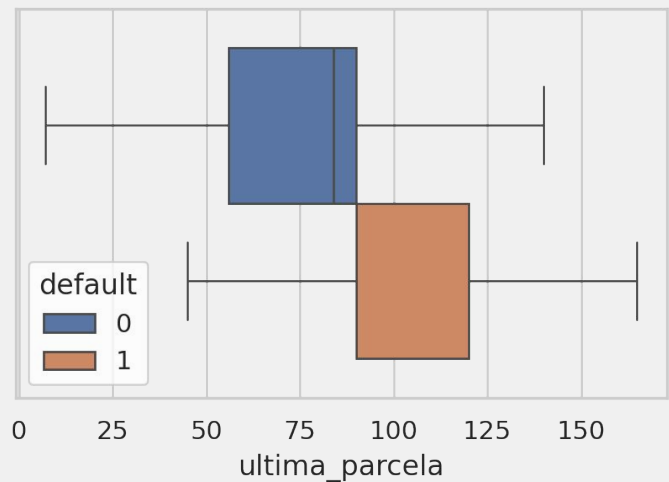
- Desbalanceadas, algumas poucas categorias dominam...

... mas há uma forte relação com default



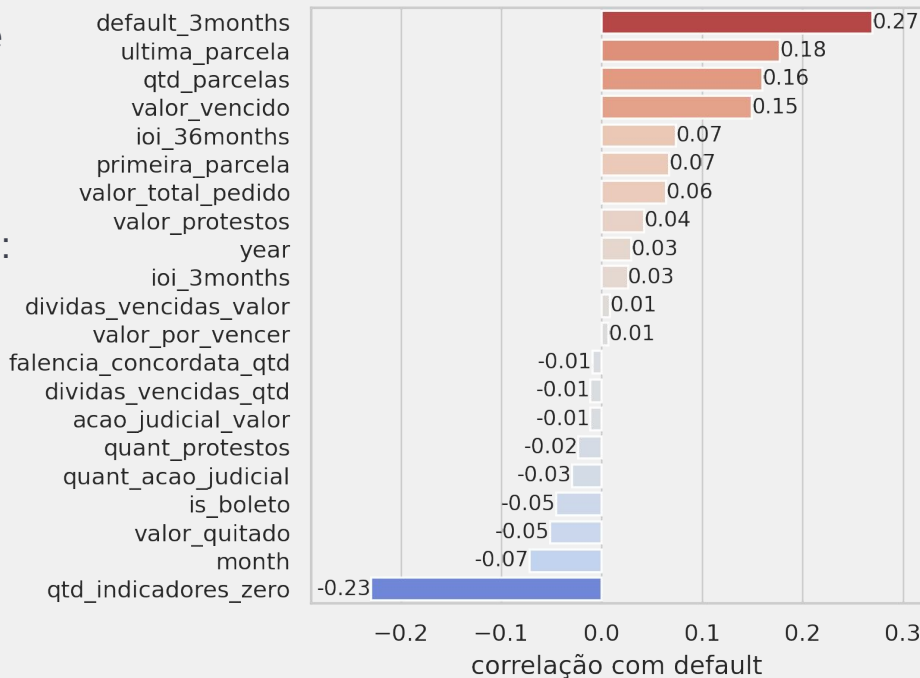
Forma de pagamento

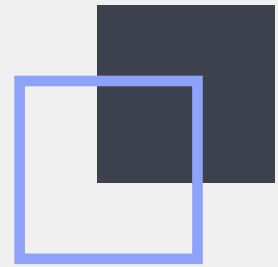
- Tem valores como: 30/60/90, boleto 7 dias, 24x
- Extração de informações úteis:
 - Número de parcelas
 - Data das parcelas
 - Quantidade de parcelas
 - Boleto



Correlação com default

- As maioria das variáveis tem **correlação fraca**
- As variáveis internas têm correlação maior que as externas:
 - ex: número de defaults nos últimos 3 meses
- As variáveis novas extraídas dos dados estão, em geral, entre as mais correlacionadas:
 - ex: última parcela
 - ex: quantidade de indicadores zero



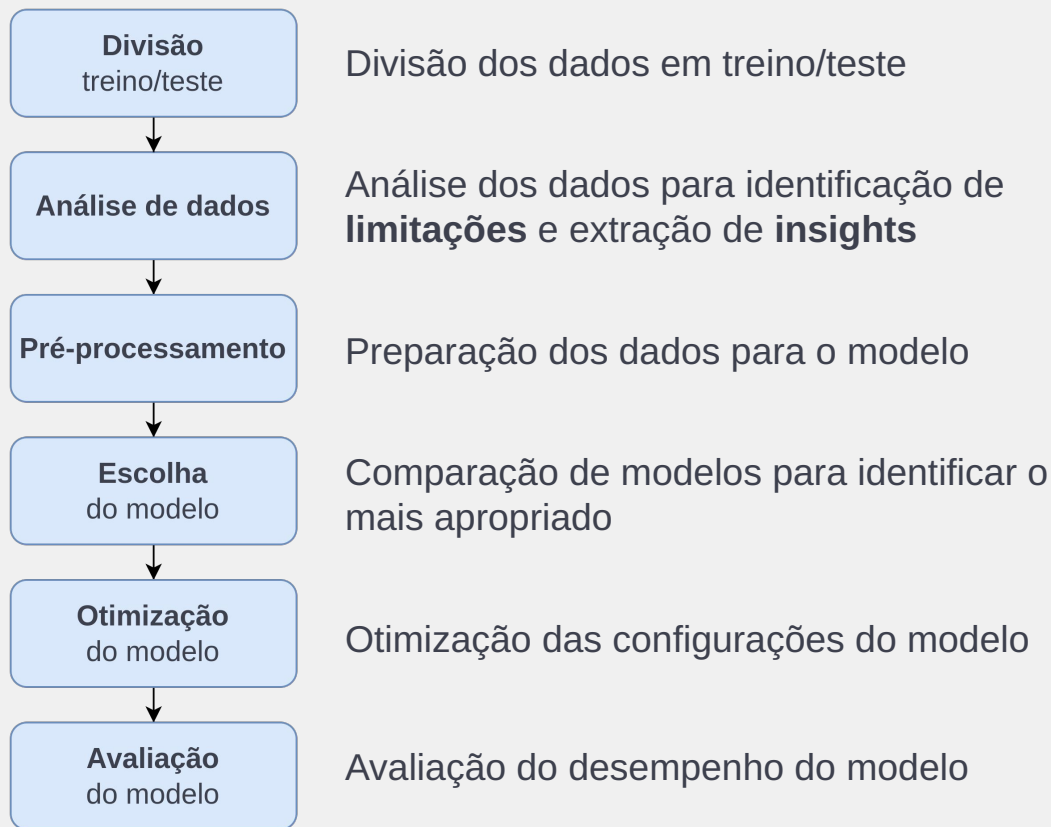


03

0 Modelo

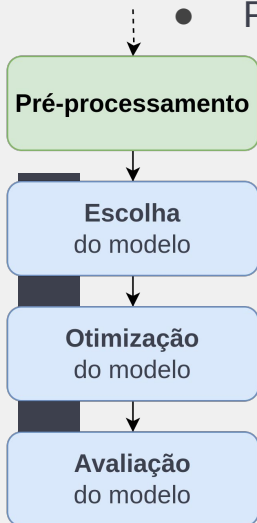


Lembrete: Abordagem



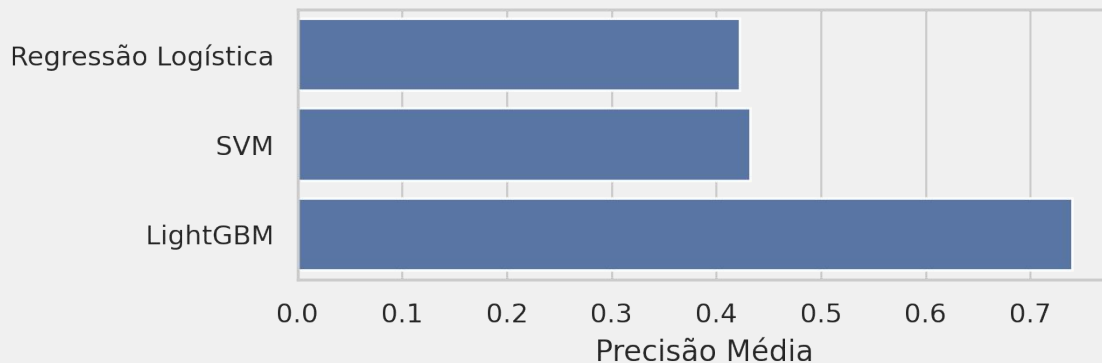
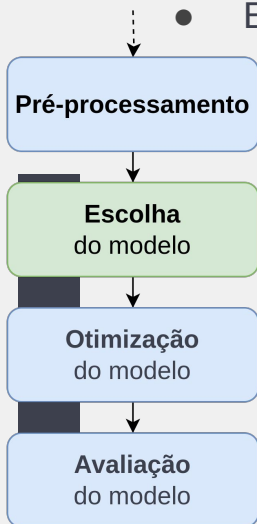
Pré-processamento

- Correção de erros nos dados
- Extração de novas variáveis:
 - A partir da forma de pagamento
 - A partir dos indicadores de inadimplência
- Preparação dos dados para processamento pelo modelo



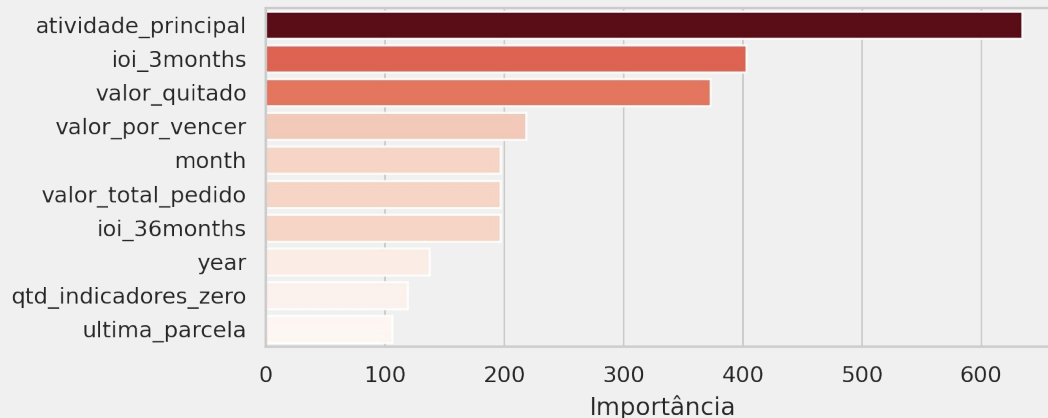
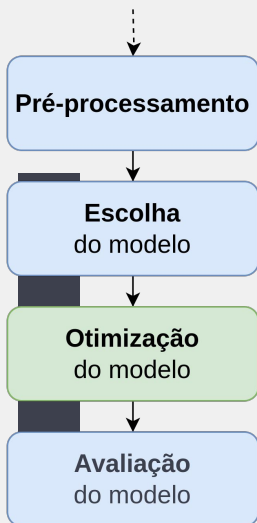
Escolha do modelo

- A complexidade e qualidade dos dados sugere um modelo complexo
 - **LightGBM**: modelo frequentemente utilizado com bons resultados em dados complexos
- Dúvida: precisamos de um modelo complexo?
 - **Regressão Logística** e **SVM** são bons em dados menos complexos
- Escolha: **LightGBM**



Otimização do modelo

- Identificação das variáveis com impacto no desempenho do modelo
- Ajuste das configurações do modelo para otimizar o desempenho





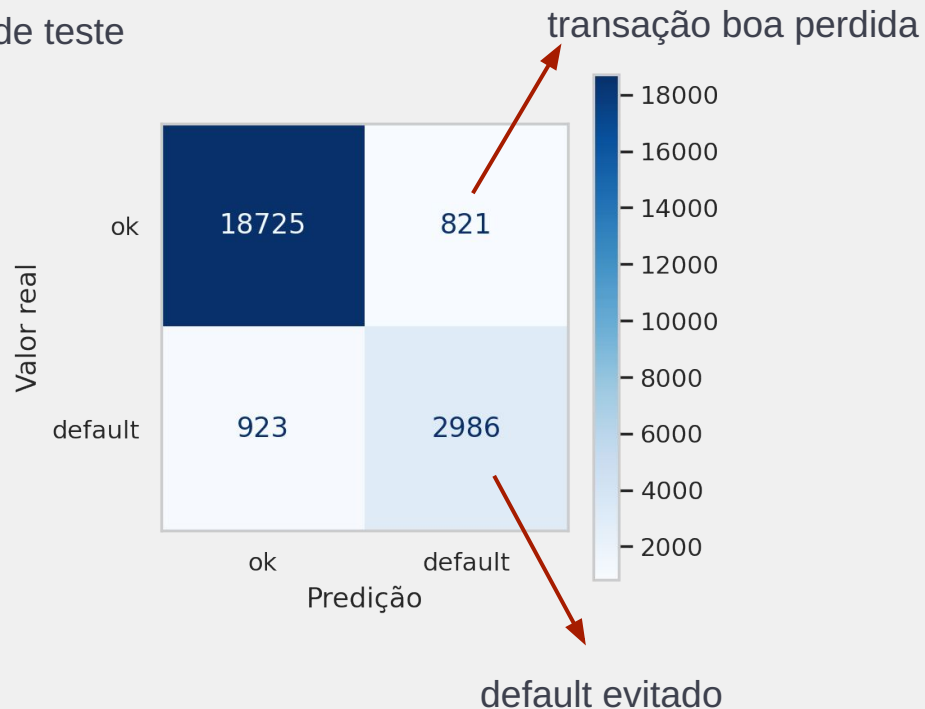
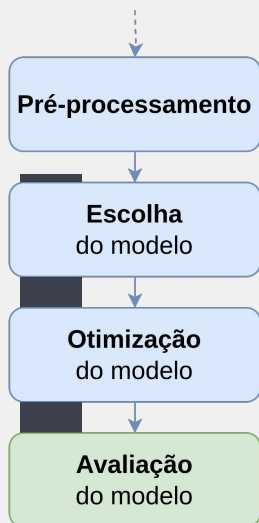
04

Resultados



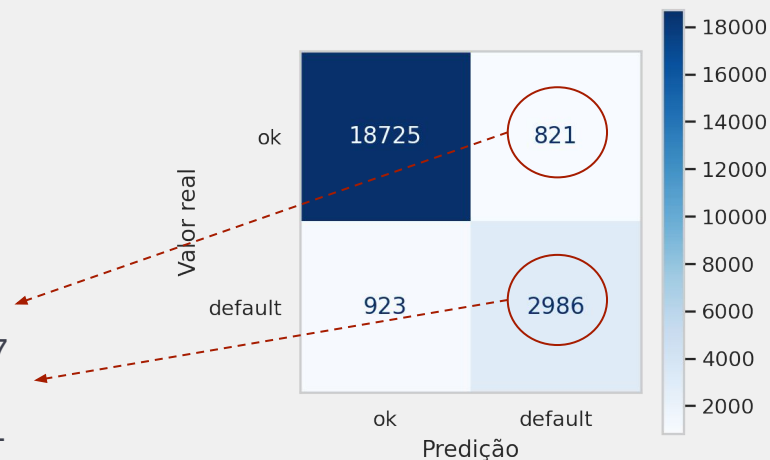
Avaliação do Modelo

- Avaliação nos dados de teste
- Matriz de confusão



Avaliação do Modelo

- Acurácia: 92% | F1: 0,77
- 74% dos defaults identificados
- Considerando uma margem de lucro de 20%, nos dados de teste (20% dos dados):
 - Perdas evitadas pelo modelo: R\$ 47.749.456,97
 - Perdas causadas pelo modelo: R\$ 2.931.539,11
 - Líquido: economia de R\$ 44.817.917,86
- Em todos os dados: **economia de R\$ 224.085.767,69**



Impacto

- Grande **redução nas perdas com default**
- Abertura para **novos clientes**
- Maior **confiança e melhor relação** com o cliente

Limitações e Riscos

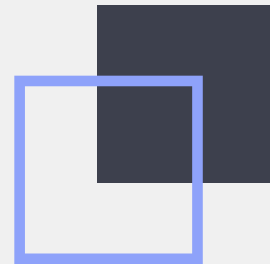
- **Grandes variações** identificadas no final de 2019, possíveis impactos nos resultados
- O comportamento dos clientes pode **variar com o tempo**, sendo necessário re-treinar e re-avaliar os modelos
- A rejeição incorreta de uma transação pode ter impactos negativos em outras transações com o mesmo cliente

Considerações Éticas

- Importância de informar o cliente sobre o uso dos dados e, se necessário, obter consentimento
- Estabelecimento de um mecanismo para revisão das decisões tomadas pelo modelo
- Estabelecimento de processos para identificar falhas consistentes no modelo

Recomendações

- Utilização do modelo como um mecanismo de auxílio
 - Decisão final tomada por **humanos**
- **Atualização contínua** do modelo para identificação de novas tendências
- Implementação de **estratégias de mitigação** para clientes de risco elevado, quando apropriado:
 - Exemplo: termos de pagamento mais flexíveis
 - Principalmente no caso de clientes de longo prazo
- Revisão dos processos de coleta de dados



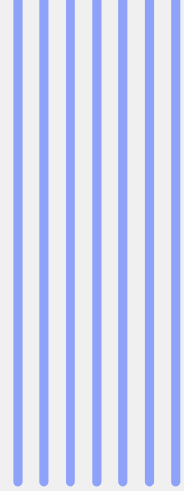
05

Conclusão



Conclusão

- Extração de insights relacionados à ocorrência de defaults
- Desenvolvimento de um modelo de predição de default com desempenho elevado
- Economias estimadas de R\$ 224.085.767,69 de 2017 a 2019
- Recomendações para a implementação e manutenção do modelo



Obrigado

