

Template de Modelos

Nome do Modelo

Versão

DD/MM/YYYY



Preditiva.ai

1. Introdução e escopo

- a. Objetivo do modelo
- b. Histórico de versões
- c. Visão regulatória
- d. Papéis e responsabilidades
- e. Bases de dados utilizadas
- f. Público alvo
- g. Target
- h. Descrição das variáveis
- i. Premissas e Limitações

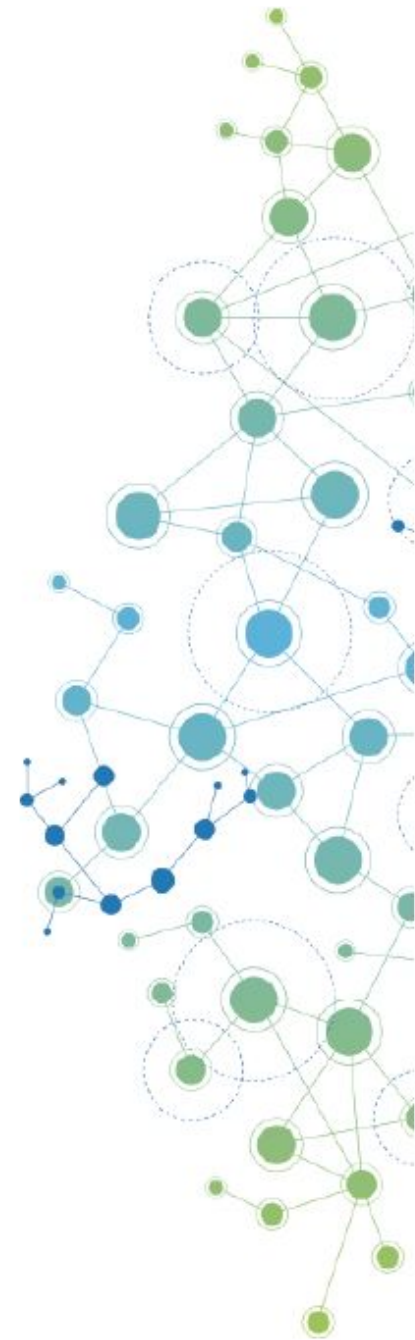
2. Metodologia

- a. Análise exploratória
- b. Descrição da técnica utilizada
- c. Resultados obtidos
- d. Macro fluxo da solução

3. Conclusão e próximos passos

- a. Deploy em produção
- b. Acompanhamento do modelo
- c. Roadmap das melhorias

Introdução e Escopo



Introdução e escopo

Objetivo do modelo



Preditiva.ai

Este modelo possui como objetivo estimar a probabilidade de churn (abandono) de um cliente de cartão de crédito.

O churn é definido como o abandono do cartão pelo cliente em um período de até 4 meses após a mensuração.

Com este modelo, a squad anti-churn entrará em contato com o cliente realizando ofertas que reduzam a probabilidade de abandono.

É esperado que a utilização do modelo ajude a reduzir a taxa de churn em pelo menos 25%, e por consequência aumente as receitas em pelo menos 20%.

- Descrever o que o modelo busca responder e/ou resolver?
- Quais os usos?
- Quais expectativas?

Introdução e escopo

Histórico de versões



Versão	Data	Responsável	Descrição
V1.0 - Inicial	15/01/2019	Nome do Cientista	Desenvolvimento do modelo de Churn Cartões.
V1.1	27/06/2020	Nome do Cientista	Recalibragem dos coeficientes com base de Jan a Abr/20.

- Descrever o histórico de versões do modelo.
- Incluir mesma tabela no Readme do repositório.

Introdução e escopo

Visão regulatória



Preditiva.ai

No escopo do churn, não existem regulamentações que tenham impacto no tema.

Opcional

- Incluir as resoluções, portarias e/ou instruções em que o processo alvo está submetido.

Ex: Res. 2682 (PDD)

Introdução e escopo

Papéis e responsabilidades



O projeto de desenvolvimento de um modelo para estimativa da probabilidade de churn será realizado conjuntamente pelas áreas indicadas abaixo, assim como suas respectivas responsabilidades e ponto focal.

Papel/Atividade	Área responsável	Ponto focal
Descrição do processo	Produtos Cartões de Crédito	Cláudio Costa
Disponibilização dos dados	Engenharia de Dados	Edson Nakami
Desenvolvimento do modelo	Data Science	Marcia Souza
Validação da metodologia	Validação de Modelos	João Fernandes
Validação do valuation	Finanças	Flavia Bonna

Incluir as atividades e responsáveis para deixar clara a atuação de cada um dos stakeholders do projeto.

Introdução e escopo

Público alvo



Preditiva.ai

O público considerado para a construção do target deste modelo foi selecionado a partir das características abaixo:

- Possuir um cartão ativo no período de Jan/19 a Jun/20;
- Possuir um cartão da categoria Internacional ou superior;
- Possuir gasto médio mensal superior a R\$ 3.000,00;
- Possuir todas as faturas pagas sem atraso.

Descrever o **público alvo** que será usado para chegar no Target do modelo.

Importante colocar todos os **filtros realizados** nos dados brutos coletados.

Introdução e escopo

Target



Preditiva.ai

A variável Target do modelo é do tipo binária e será construída avaliando-se as características do cliente, cartão e demais variáveis em uma data de referência.

- **Churn:** cliente abandona o cartão de crédito em até 4 meses após a data de referência.
- **Não churn:** caso contrário.

Se modelo supervisionado, descrever o evento de interesse (Target).

Observação: Incluir volumetrias e curvas de Média de Target no tempo. Ex: Bad Rate / Good Rate etc.

Introdução e escopo

Bases de dados utilizadas



As bases de dados utilizadas com as respectivas informações são apresentadas na tabela abaixo:

Base de dados	Datas bases de referência	Qtde de observações	Fonte da informação	Responsável pela disponibilização
Transações no cartão de crédito	Jan/19 a Dez/19	12.025.352	Sistema MVCC	Produtos Cartões de Crédito
Vigência dos cartões de crédito	Jan/19 a Jun/20	2.064.964	Sistema MVCC	Produtos Cartões de Crédito

Repositório GitHub:

Descrever os **tipos de informação** utilizada, o **período** de tempo da query de extração, a **fonte** e o **responsável** por esta extração, para referências futuras. Incluir os acessos necessários para extrair a base.

Incluir os links para o **GitHub/Bitbucket** do projeto.

Importante: com esse arquivo, acesso ao S3/DW e ao repositório, uma pessoa deveria ser capaz de recriar o modelo do zero.

Introdução e escopo

Descrição das variáveis



As bases de dados utilizadas com as respectivas informações são apresentadas na tabela abaixo:

Variável	Tipo	Descrição
Score Crédito	Texto	Capacidade financeira do cliente
Estado	Número	UF do cliente
Gênero	Texto	Gênero do cliente
Idade	Número	Idade do cliente
Tempo como cliente (meses)	Número	Tempo que o cliente tem o cartão de crédito em meses
Saldo	Número	Saldo da conta corrente do cliente
Número de produtos	Número	Quantidade de produtos que o cliente possui
Assina mensalidade	Texto	Indicação se o cliente assina a mensalidade
Renda mensal presumida	Número	Renda presumida do cliente
Data Ultima Transacao	Data	Data em que o cliente fez a última compra no cartão de crédito

Descrever as **principais variáveis** a serem estudadas neste slide. Caso exceda o espaço, deixar a lista de variáveis em anexo.

Introdução e escopo

Premissas e Limitações



Preditiva.ai

Premissas

- Clientes que possuem cartão adicional mantém ou abandonam todos os cartões de uma só vez.
- Todas as contratações e abandonos estão registrados adequadamente no sistema MVCC.

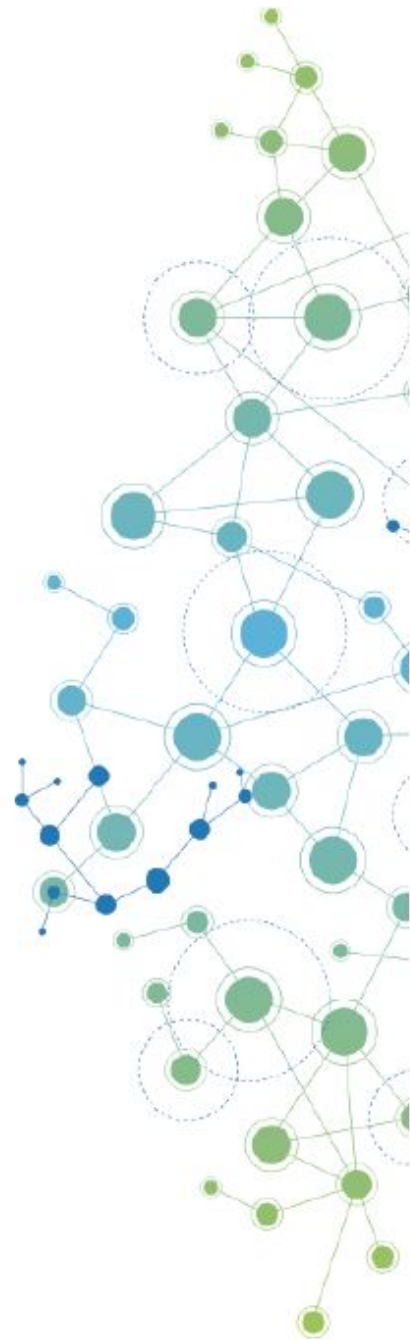
Descrever as premissas do estudo. Ou seja, todas as hipóteses aceitas de forma a continuar com os estudos/modelagem.

Limitações

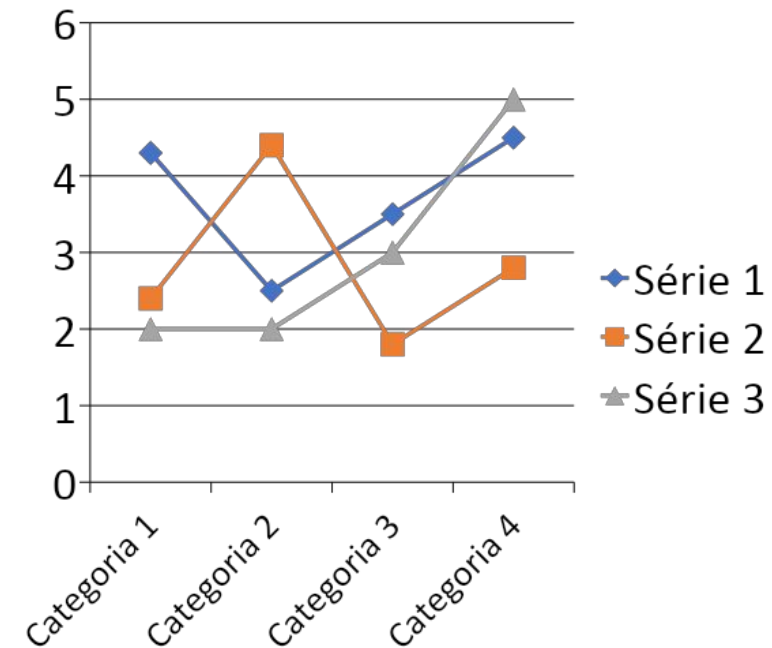
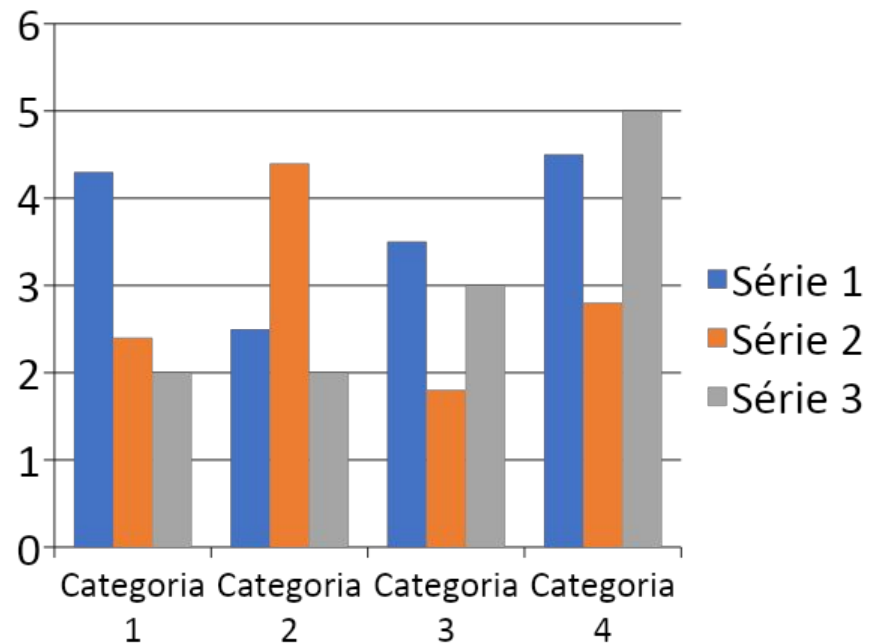
- Clientes que abandonarem o cartão de crédito e retornarem dentro do período de 4 meses não serão corretamente identificados como tendo abandonado.
- Variável resposta do modelo mudou de definição no meio do ano, pois antes o período para considerar o churn era de 3 meses.
- O modelo escolhido não é diretamente interpretável, porém é acurado.
- Devido a limitações sistêmicas, a base de treinamento é uma amostra da população de interesse.

Descrever as limitações do estudo. Ou seja, as situações que podem trazer “pontos cegos” ao modelo/estudo.

Metodologia



Inserir os resultados das análises da distribuição das variáveis explicativas em relação ao target (se modelo linear) e estabilidade da variável ao longo do tempo. Importante para detectar problemas de performance do modelo na fase de acompanhamento.





Descrever de forma resumida a ideia das técnicas utilizadas. Essa descrição deve conter, minimamente, os seguintes itens:

- **Tipo:** Regressão, Classificação, Clusterização etc
- **Tipo de aprendizado:** Supervisionado e Não Supervisionado
- **Variável resposta:** binária, multi-classe, multi-label, numérica.
- **Variáveis explicativas/features:** variáveis originais e engenhadas.
- **Método de validação:** Holdout, K-fold, Out-of-time etc.
- **Métrica de performance:** KS, Acurácia, R2 ajustado, MSE, ROC, Gini, F1 Score, Precision e Recall etc.

Apresentar o diagnóstico do modelo selecionado em relação a:

- **Overfitting / Underfitting:** performance em treino, validação e teste
- **Importância das variáveis**
- **Intepretação do modelo:** LIME / Shapley Values
- **Distribuição dos scores** ou grupos finais
- **Estabilidade dos scores** ao longo do tempo

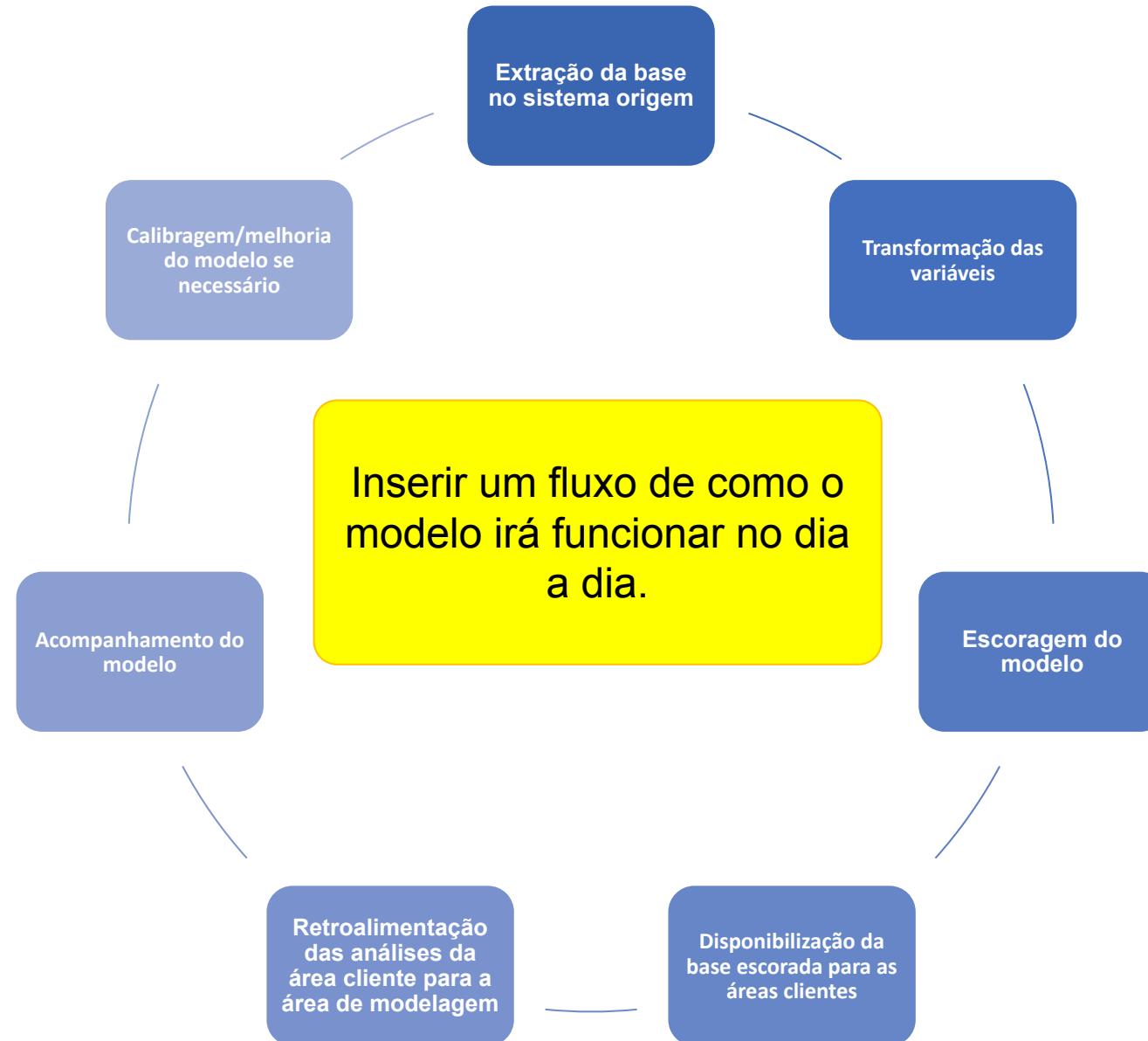
Incluir também os potenciais de **resultados financeiros e não financeiros**.

Metodologia

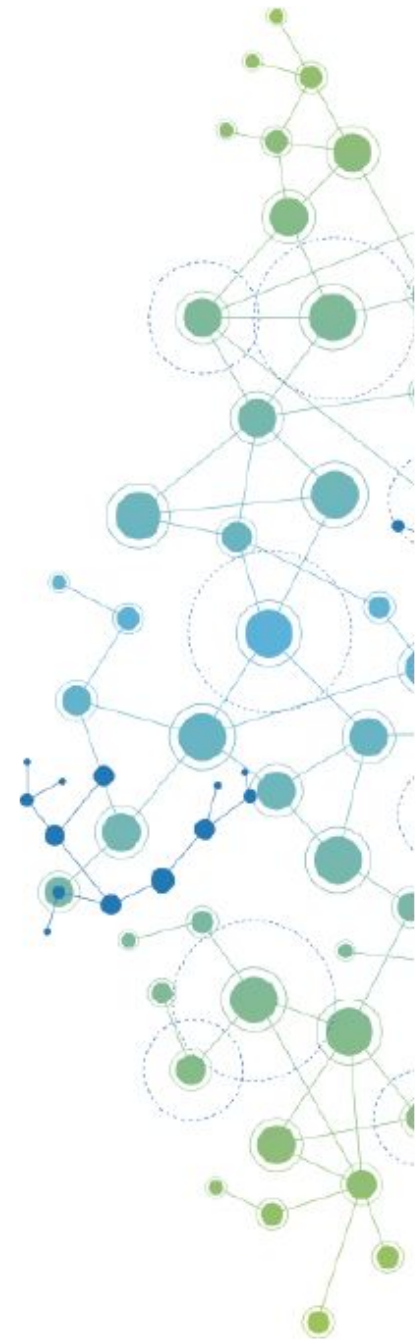
Macro fluxo da solução



Preditiva.ai



Conclusão e próximos passos



Conclusão e próximos passos

Deploy



Descrever em detalhes como o processo será implantado em produção.

- **Periodicidade da escoragem:** Diário, Semanal, mensal etc.
- **Tipo de disponibilização:** Arquivo CSV, Tabela SQL, AWS, Arquivo Excel etc.
- **Local da disponibilização:** Rede, E-mail, Servidor SQL etc.
- **Ponto focal da disponibilização:** José de Almeida -
jose.almeida@meubanco.com.br, ramal 1234, Área de Engenharia de Dados etc.

Conclusão e próximos passos

Acompanhamento do modelo



Descrever os KPIs que devem ser utilizados para o monitoramento de performance do modelo com os respectivos referenciais.

- **Performance do Modelo:** KS/ROC ou qualquer métrica de performance no conjunto de treino.
- **Distribuição dos scores/clusters:** no conjunto de treino.
- **Taxa do Target:** no conjunto de treino.

Importante: Definir também os gatilhos propostos para iniciar um processo de re-treino ou calibragem do modelo.

Conclusão e próximos passos

Roadmap de melhorias



Descrever em atividades e datas os próximos passos para as novas releases do modelo.

- Plano de ação para coletar novos dados.
- Número do Chamado para o problema sistêmico que gerou uma limitação do escopo.
- Novas reuniões de Brainstorming.
- Novas variáveis
- Etc.



Preditiva.ai