

Processo seletivo Data Scientist - Case para Análise de Viabilidade da Adoção de um Novo Modelo Externo de Prevenção de Fraudes para o Boticário

Lisandra Melo

10 de agosto de 2024



Este relatório apresenta uma análise do impacto financeiro e estratégico da adoção de uma nova ferramenta de prevenção de fraudes na empresa Boticário. São discutidas as métricas de desempenho do modelo atual, bem como estatísticas e projeções para determinar a viabilidade da proposta da empresa externa.

Conteúdo

1	Introdução	3
2	Descrição do Problema	3
3	Materiais Construídos	3
4	Implementação e Aspectos Técnicos	3
5	Análise Descritiva	4
5.1	Métricas de Pedidos Reprovados	4
5.2	Métricas de Pedidos Aprovados	5
5.3	Taxa de Aprovação	5
5.4	Fraudes Aprovadas	5
6	Análise de Impacto Financeiro	6
6.1	Análise de ROI	6
7	Análise de Escalabilidade	6
8	Conclusões e Recomendações	8
8.1	Viabilidade da Compra	8
8.2	Estratégia de Implementação	8

1 Introdução

O objetivo deste relatório é avaliar a viabilidade da adoção de uma nova ferramenta de prevenção de fraudes oferecida por uma empresa externa, em comparação com o modelo atual utilizado pelo Boticário. A análise inclui uma revisão das métricas atuais de aprovação e fraude, uma avaliação financeira do custo-benefício, e uma projeção de escalabilidade em cenários de crescimento.

2 Descrição do Problema

Atualmente, o Boticário utiliza um modelo interno de escoragem para a aprovação ou reprovação de pedidos. Com uma taxa de aprovação de 87.9%, a proposta da nova ferramenta sugere um aumento de 4.4 p.p. nesta taxa, ao custo de R\$6 por transação avaliada. Este relatório busca determinar se tal investimento é justificável e qual seria a melhor estratégia para sua implementação.

3 Materiais Construídos

Para a execução do case, foi utilizado um Jupyter Notebook do Google Colaboratory, para acessá-lo [clique aqui](#).

4 Implementação e Aspectos Técnicos

A implementação foi realizada utilizando ferramentas do Google. Especificamente, o BigQuery Studio da Google Cloud Platform e o Jupyter Notebook do Google Colaboratory foram utilizados para tratamento de dados, cálculo de métricas e construção de objetos.

Os dados presentes na planilha disponibilizada foram importados para o BigQuery Studio e três tabelas - Pedidos, Fraude e Backtest - foram criadas para representar cada uma das planilhas de dados fornecidas. Após a implementação deste processo, os dados foram acessados utilizando um Jupyter Notebook com acesso à API do Big Query.

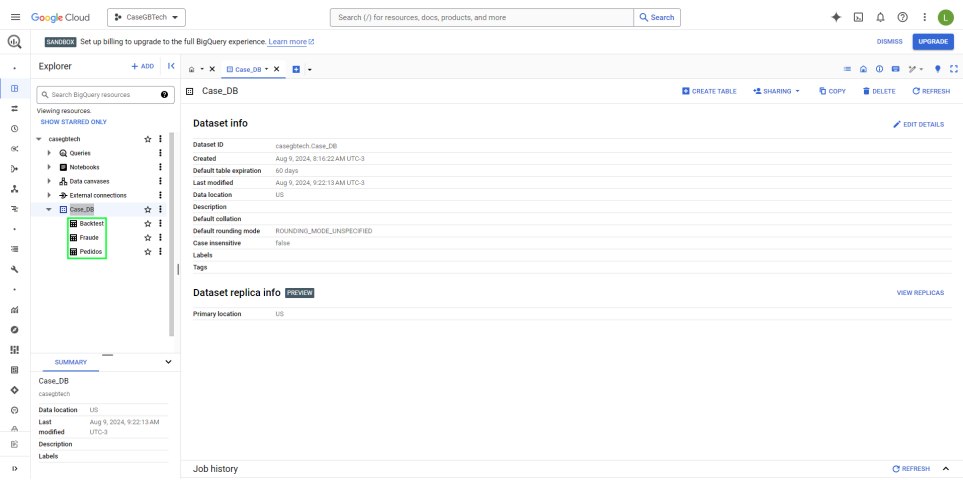


Figura 1: Painel do BigQuery Studio

Para a análise dos dados providos foi utilizado Python e as bibliotecas Pandas, Matplotlib e Numpy. O estudo de viabilidade foi conduzido em algumas etapas: uma Análise Descritiva nos dados atuais e obtidos com a nova ferramenta, Análise de Impacto financeiro e Análise de Escalabilidade.

5 Análise Descritiva

5.1 Métricas de Pedidos Reprovados

A análise das reprovações pelos modelos é essencial para entender os aspectos de melhoria, ou não, implementadas pelo modelo oferecido à empresa. A Tabela 3 apresenta as principais métricas.

Tabela 1: Métricas dos Pedidos Reprovados Obtidas no Jupyter Notebook Produzido

Modelo	Pedidos Reprovados	Valor Total	Valor Médio	Mediana	Desvio Padrão
Atual	121	R\$24314.00	R\$200.94	R\$204.00	R\$53.78
Novo Modelo	74	R\$14900.00	R\$201.35	R\$209.50	R\$56.50
Diferença	47	R\$9414.00	-	-	-

A partir dos valores obtidos podemos afirmar que o novo modelo oferece uma diminuição de 47 pedidos anteriormente reprovados, com isso, utilizando o modelo conseguimos uma diminuição de R\$9414.00 em valores perdidos em pedidos reprovados. Abaixo, um gráfico apresenta uma comparação entre os valores perdidos em pedidos reprovados.

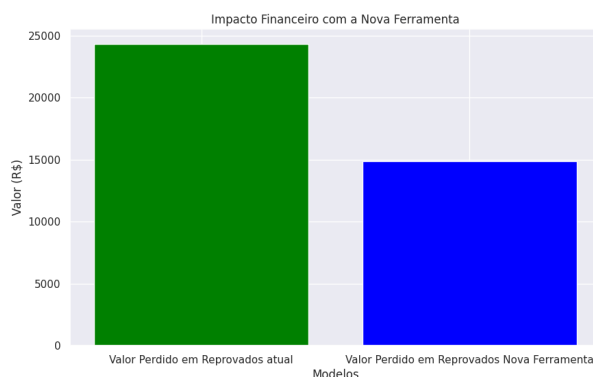


Figura 2: Comparação de valores perdidos em pedidos reprovados entre Modelo Atual, em verde, e Novo Modelo, em azul

5.2 Métricas de Pedidos Aprovados

Já análise das aprovações fornece uma visão real do impacto dos modelos dentro dos pedidos que foram realizados com sucesso. A Tabela 2 apresenta as principais métricas encontradas.

Tabela 2: Métricas dos Pedidos Aprovados Obtidas no Jupyter Notebook Produzido

Modelo	Pedidos Aprovados	Valor Total	Valor Médio	Mediana	Desvio Padrão
Atual	879	R\$175335.00	R\$199.47	R\$200.00	R\$56.72
Novo Modelo	926	R\$184749.00	R\$199.51	R\$200.00	R\$56.37
Diferença	47	R\$9414.00	-	-	-

Comparando os dados, percebemos que o novo modelo consegue aprovar 47 pedidos adicionais, proporcionando um aumento de R\$9414.00 no valor total de pedidos aprovados, sem alterar substancialmente as outras métricas descritivas, como valor médio, mediana e desvio padrão.

5.3 Taxa de Aprovação

Na proposta realizada, a empresa propõe que o modelo gerará um aumento de 4.4 pontos percentuais na taxa de aceitação de pedidos. Para verificar se isso se confirma, a taxa de aceitação de cada modelo foi calculada usando os dados apresentados na seção anterior.

A taxa de aprovação do modelo atual está em 87.90%, enquanto o modelo externo apresenta uma taxa de 92.60%. Dessa forma, o novo modelo apresenta um aumento de 4.7 pontos percentuais, um valor ainda maior que o proposto pela empresa. Portanto, a empresa não somente cumpre com o proposto como proporciona um aumento maior do número de pedidos aprovados.

5.4 Fraudes Aprovadas

Já com relação à aceitação de fraudes, a empresa propõe a manutenção de das fraudes aprovadas. Nesse sentido, se torna necessário calcular os valores de registros como esses

nos modelos.

Tabela 3: Métricas de Fraudes Aprovadas Obtidas no Jupyter Notebook Produzido

Modelo	Fraudes Aprovadas	Taxa de Aprovação de Fraudes
Atual	36	3.60%
Novo Modelo	36	3.60%
Diferença	0	0

Com isso, é notável que a empresa cumpre com o prometido com relação à aprovação de Fraudes, mantendo a mesma taxa de 3.60% para fraudes aceitas.

Portanto, é possível apontar que a empresa cumpre com tudo proposto, seu modelo realmente melhora a aceitação de pedidos não afetando o número de fraudes aceitas. Isso indica que o modelo pode melhorar significativamente o processo de análise de fraudes na empresa.

6 Análise de Impacto Financeiro

Contudo, antes de apontarmos o modelo oferecido como o ideal para implementação é preciso observar que este possui um custo adicional de R\$ 6.00 por transação avaliada. Tal fato pode impactar sua rentabilidade para a empresa. Para analisar essa questão, iremos calcular o custo total do modelo à empresa e obter qual o valor líquido oferecido pelo novo modelo.

Considerando que a quantidade de pedidos na nossa amostra é de 1000, temos que o custo total da implementação da ferramenta seria de R\$ 6000.00. Agora tomando posse do valor de R\$9414.00 que foi gerado pelo modelo teríamos que a empresa aumentaria R\$3414.00 em valor líquido recebido na amostra atual de dados.

Isto é, mesmo com o custo adicional, o modelo continua vantajoso para a empresa.

6.1 Análise de ROI

O cálculo do retorno sobre o investimento (ROI) considera o valor dos pedidos adicionais aprovados e o custo da nova ferramenta. Por isso, pode apontar se há sucesso financeiro na implementação.

Com o Cálculo do ROI, temos que o valor da métrica para o modelo é de 56.90%, isso indica que a implementação da ferramenta gerará um bom retorno financeiro.

Com a amostra fornecida, percebe-se que o sucesso financeiro, junto ao aumento de número de pedidos aprovados e a manutenção de fraudes aprovadas, indicam que a mudança será positiva para a área de avaliação de risco de fraudes nas compras realizadas no Boticário.

7 Análise de Escalabilidade

Além de avaliar a rentabilidade do modelo a partir da amostra é importante entender o comportamento dos valores em diferentes números de pedidos. Para isso, iremos propor a análise do valor esperado para o valor recebido em pedidos a partir do uso da nova

ferramenta e utilizando o modelo atual. Consideraremos a média de valor de valores aprovados e a taxa de aprovação apresentadas pelos modelos na amostra na construção dos modelos.

Uma equação geral do valor esperado pode ser denotada abaixo.

$$\text{Valor Esperado} = \text{numeroPedidos} \times \text{mediaValorAprovados} \times \text{taxaAprovacao} - (\text{numeroPedidos} \times \text{custo})$$

Para o modelo atual teremos que

$$\begin{aligned} \text{Valor Esperado Atual} &= \text{numeroPedidos} \times 199.47 \times 0.879 \\ \Rightarrow \text{Valor Esperado Atual} &= 175,33413 \times \text{numeroPedidos} \end{aligned}$$

Já para o novo modelo,

$$\begin{aligned} \text{Valor Esperado Nova Ferramenta} &= \text{numeroPedidos} \times 199.51 \times 0.926 - (\text{numeroPedidos} \times 6) \\ \Rightarrow \text{Valor Esperado Nova Ferramenta} &= 178,74626 \times \text{numeroPedidos} \end{aligned}$$

Com posse dessas funções podemos apontar que o ponto de inflexão entre as duas funções ocorre quando o número de pedidos é zero e que valor esperado com a nova ferramenta sempre será maior que o atual. Indicamos também que com o crescimento de pedidos o impacto positivo da ferramenta se tornará ainda maior.

No gráfico de linhas abaixo, é possível observar com bastante clareza o comportamento das duas curvas, que ilustra o que foi observado a partir das funções propostas.

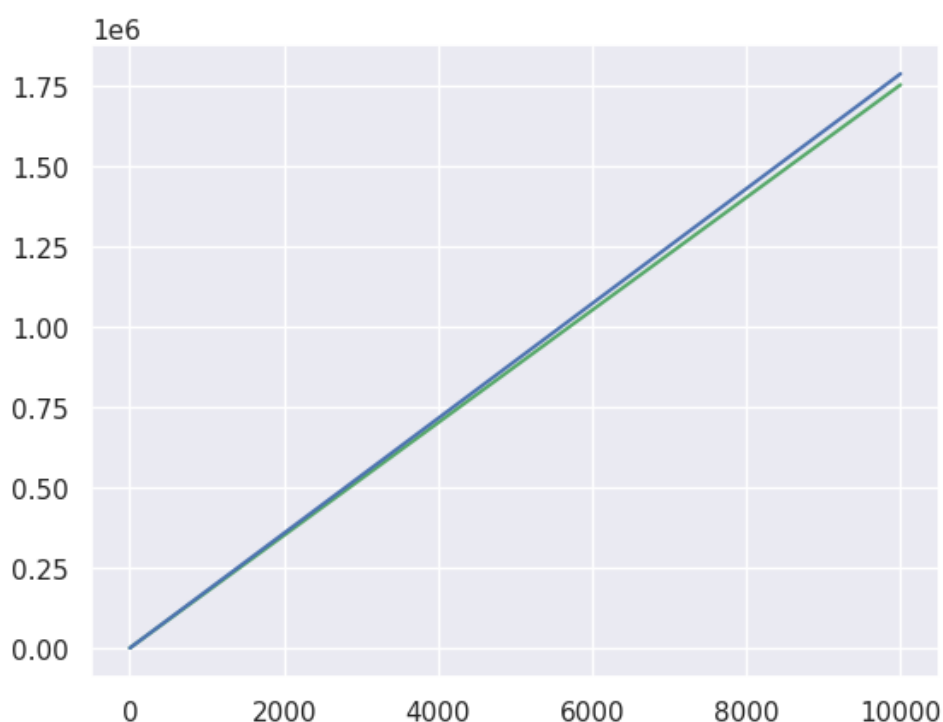


Figura 3: Comportamento do Modelo Atual, em verde, e Novo Modelo, em azul, com diferentes quantidades de pedidos

8 Conclusões e Recomendações

8.1 Viabilidade da Compra

A partir das análises realizadas, podemos concluir que a compra do produto oferecido pela empresa externa é financeiramente viável e vantajosa para o Boticário. O modelo proposto não apenas cumpre, mas excede as promessas feitas, ao aumentar significativamente a taxa de aprovação de pedidos sem impactar negativamente a taxa de fraudes aprovadas. O cálculo do ROI demonstra que a implementação gerará um retorno positivo, com um aumento líquido significativo no valor recebido, mesmo após considerar o custo adicional de R\$ 6,00 por transação avaliada.

Além disso, a análise de escalabilidade aponta que a aplicação do novo modelo é escalável e impactará a empresa positivamente o processo de análise de fraudes em pedidos realizados no Boticário.

Dessa maneira, a compra do produto oferecido será vantajosa para o Boticário, a curto e longo prazo, oferecendo um impacto financeiro positivo com o aumento nas aceitações de pedidos na análise de fraudes.

8.2 Estratégia de Implementação

Para garantir uma transição eficaz e que maximize os benefícios do novo modelo de prevenção de fraudes, é recomendável adotar uma estratégia de implementação gradual e continuamente monitorada. Inicialmente, sugere-se aplicar o modelo externo em um grupo piloto de transações. Essa abordagem permitirá uma avaliação real de seu desempenho antes de uma implementação completa.

Outra alternativa interessante pode ser a utilização de uma abordagem híbrida, na qual o modelo interno existente e o novo modelo externo são combinados. Isso pode ser particularmente vantajoso na segmentação de transações de maior risco, onde o modelo interno pode funcionar como uma segunda camada de verificação. Essa integração também pode facilitar a transição, minimizando qualquer impacto negativo durante o período de adaptação.

Além disso, é essencial estabelecer um sistema robusto de monitoramento contínuo. A performance do modelo externo deve ser regularmente avaliada para garantir que ele continue entregando os resultados prometidos e mantendo o ROI positivo. Essa análise constante permitirá que a empresa responda rapidamente a qualquer variação no desempenho, garantindo a manutenção dos benefícios esperados.

Por fim, recomenda-se a realização de análises periódicas de custo-benefício, em intervalos periódicos. Essas revisões devem verificar se o custo adicional por transação continua sendo justificável à luz dos resultados obtidos. Dessa forma, o Boticário poderá assegurar que o investimento no novo modelo de prevenção de fraudes permaneça financeiramente vantajoso no longo prazo.