

Análise de dados e caminhos para resolução das transações fraudulentas

Camila Gomes Vilaça

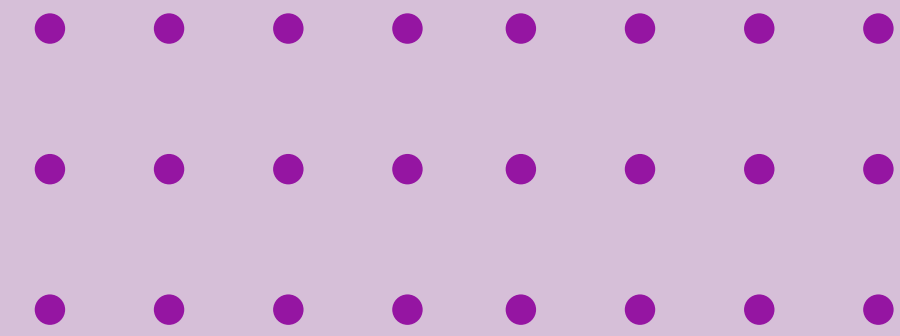
University of Chicago



Diagnóstico

Como nós vimos com a atual equipe aumentaram-se o número de transações fraudulentas, e precisamos analisar mais a fundo e buscar uma resolução para o problema.

- Precisamos de um algoritmo preditivo para detectar transações fraudulentas antecipadamente e desativar os cartões até que se possa verificar as operações suspeitas.
- O algoritmo anterior apresentava muitos erros de falso positivo no qual detectava quase todas as transações fraudulentas, mas também classificava erroneamente algumas operações normais como fraudes, forçando os clientes a verificar suas operações.
- Isso causou uma má imagem nas redes sociais, conselho do consumidor, causando danos à imagem da marca, redução nas taxas de adesão por vários meses, avaliações públicas desfavoráveis e um aumento nas reclamações individuais dos clientes, vistas por centenas de milhares de pessoas nas redes sociais.



Olivia Wilson

"Deixo minha indignação, pois tentei efetuar minha compra mais de 5x e tive meu cartão pessoal bloqueado, com a mensagem erro na transação, não comprem nessa empresa!!"



Mastercard

"Gostaríamos de notificar a empresa devido ao alto número de transações fraudulentas efetuadas com cartões mastercard no respectivo website"

As informações que devemos obter dos stakeholders são:

- Quantos dias/meses os usuários devem relatar fraudes em seu nome?
- Foi realizada uma extração de banco de dados após esse período máximo de relatório de fraude (para garantir que não houvesse transações fraudulentas que não tenham sido classificadas adequadamente, ou seja, devidamente maduras)?
- Todas as variáveis anonimizadas dispostas na base de dados estão à disposição no momento da transação para que sejam utilizadas por um modelo semelhante ao apresentado, para uma eventual tomada de decisão para aprovar ou negar a transação em tempo real?
- Compras esporádicas de alto valor.
- , Valor médio de compras, análise de crédito como forma de evitar fraudes, horários em que as transações são efetuadas.
- Tempo entre diferentes transições, ip do usuário, classe da atividade, quais os acessos feitos, bandeiras dos cartões.



O custo da detecção de fraudes não se resume apenas aos clientes que sofreram o golpe, mas também com a confiabilidade menor do nosso sistema financeiro muitas pessoas deixam de usar cartão de crédito para efetuar as compras, o que reduz também o número de vendas. Perde-se US\$ 300 por cada incidente de classificação incorreta de fraude e bloqueio de cartão. Perde-se clientes legítimos por acusá-los de fraude. O valor vitalício de cada cliente (LTV, na sigla em inglês) é de US\$ 3.500

Porquê Random Forest?

- O Random Forest é um algoritmo de aprendizado de máquina supervisionado amplamente usado em problemas de classificação e regressão.
- O aprendizado supervisionado nesse caso superaria os métodos sem supervisão pois a partir do etiquetas confiáveis para o aprendizado do modelo em uma tarefa de detecção de anomalias.

Critérios de sucesso do modelo

- Redução nas transações fraudulentas, com menos de 0,01% de todas as previsões de fraude sendo falsos positivos.
- Aumento da sensibilidade em identificação da identificação de fraudes (em comparação com o atual modelo).

= Comparativo dos modelos



True negative	False positive
29933 99.73%	8 0.03%

Árvore de
decisão

False negative	True positive
8 0.03%	51 0.17%

True negative	False positive
29938 99.79%	3 0.01%

Random
Forest

False negative	True positive
3 0.01%	56 0.19%

True negative	False positive
29927 99.76%	14 0.05%

SVC

False negative	True positive
27 0.09%	32 0.11%

True negative	False positive
29613 98.71%	328 1.09%

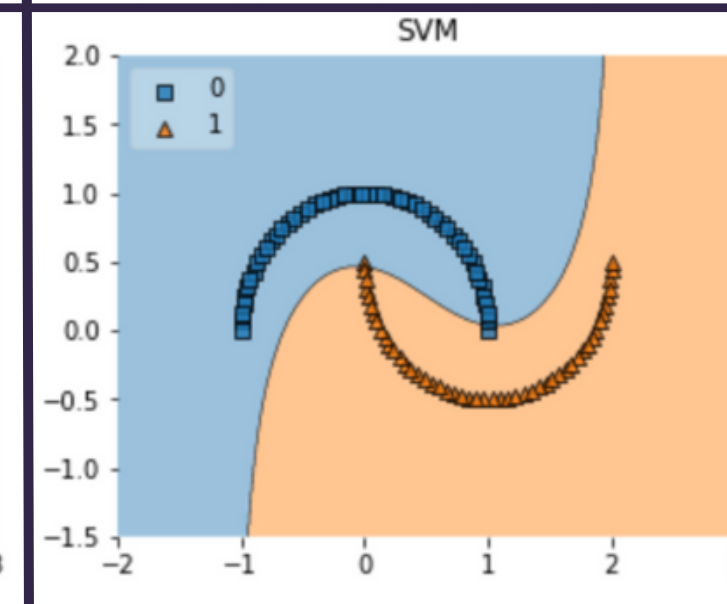
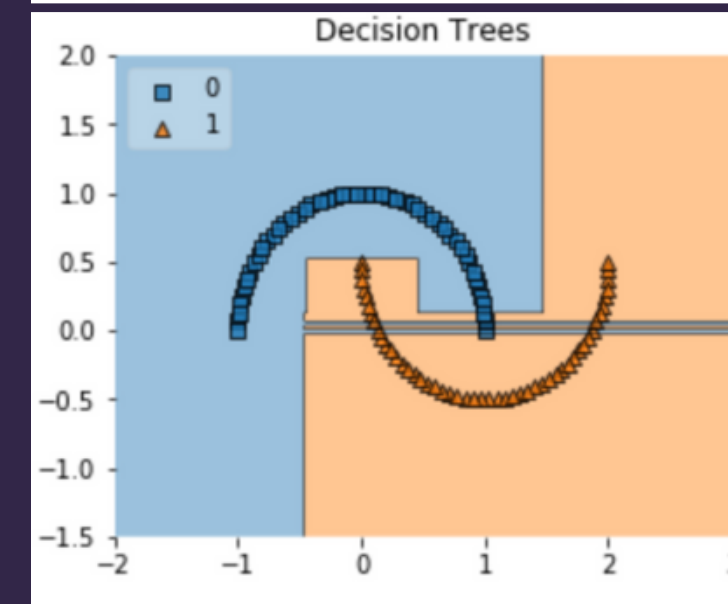
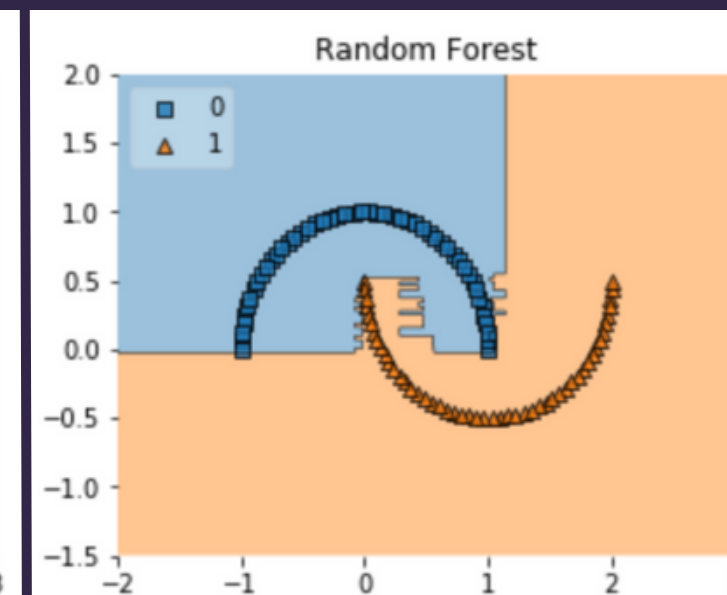
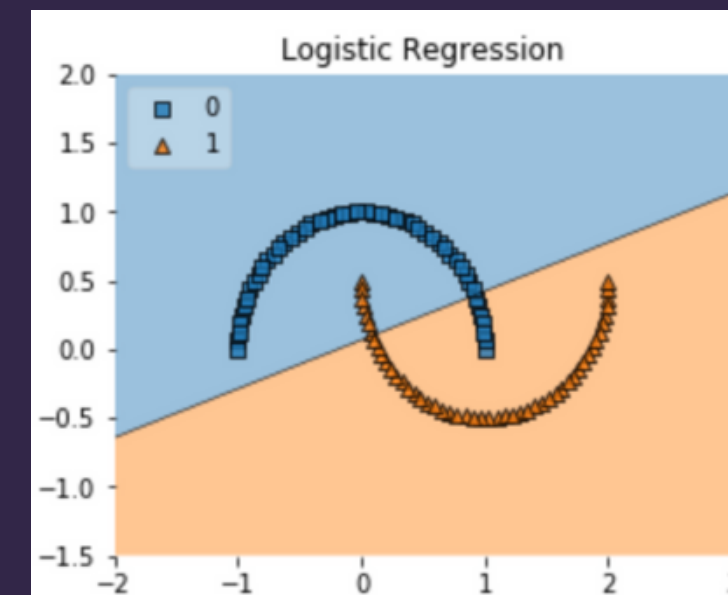
Regressão
Logística

False negative	True positive
6 0.02%	53 0.18%

True negative	False positive
29929 99.76%	12 0.04%

Regressão
Linear

False negative	True positive
21 0.07%	38 0.13%



*Nota-se que ao analisar os possíveis modelos podemos ver que o Modelo Random Forest possui maior confiabilidade do que os modelos pois possui a menor porcentagem de erros e maior sensibilidade de detecção de fraudes.

Estimativa financeira (Random forest x others)

A cada 100 mil transações de 300 dólares

Transações bem sucedidas
(True negative)

Transações fraudulentas não reconhecidas pelo sistema
(Falso negativo)

Transações identificadas como fraude mas que são autênticas
(Falsos positivos)

Reconhecimento de transações fraudulentas (True positive) Sensibilidade

	Árvore de decisão	Regressão Linear	Random forest	Regressão logística	SVC
Transações bem sucedidas (True negative)	99.730 Transações +29.933.000 Usd	99.70 Transações +29.929.000 Usd	99.790 Transações +29.938.000 Usd	98.710 Transações +29.613.000 Usd	99.760 Transações +29.927.000 Usd
Transações fraudulentas não reconhecidas pelo sistema (Falso negativo)	30 transações -8.000 Usd	70 transações -21.000 Usd	10 transações -3.000 Usd	20 transações -6.000 Usd	90 transações -27.000 Usd
Transações identificadas como fraude mas que são autênticas (Falsos positivos)	80 transações -17.000 USD LTV: -280.000 USD	40 transações -12.000 USD LTV: -140.000 USD	10 transações -3.000 USD LTV: -35.000 USD	109 transações -327.000 USD LTV: -381.500 USD	50 transações -15.000 USD LTV: -175.000 USD
Reconhecimento de transações fraudulentas (True positive) Sensibilidade	<div><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div><div></div><div></div><div></div></div>	<div><div></div><div></div><div></div><div></div><div></div></div>

Redução de apx 30% nas transações fraudulentas, com menos de 0,01% de todas as previsões de fraude sendo falsos positivos.



Caso haja um aumento no número de clientes 1.000.000 que usam este cartão de crédito os lucros podem chegar a mais de 300 milhões.



Vamos aprimorar o UX UI do site e criar novas estratégias de fidelização de clientes. Além disso iremos captar clientes perdidos.

