Estratégia

Estudo do sistema de pagamentos: Comecei por entender o funcionamento do sistema de pagamentos, incluindo os diferentes participantes e os tipos de transações envolvidas.

Análise estatística: Investi tempo em análises estatísticas para avaliar os diferentes modelos disponíveis e determinar qual seria mais adequado para o problema em questão. A escolha da regressão logística foi baseada em sua simplicidade e na compatibilidade da saída com o objetivo definido.

Decisão pela linguagem de programação: Optei por Python devido à sua simplicidade e ao vasto conjunto de ferramentas disponíveis para análise de dados. Apesar do prazo apertado, acreditei que essa escolha proporcionaria vantagens em termos de eficiência e flexibilidade.

Implementação do modelo de machine learning: Após a escolha da regressão logística, dediquei tempo para aprender e implementar o modelo de machine learning. Essa etapa foi crucial para transformar o conhecimento teórico em uma solução prática e eficaz.

Tarefas

2.1.a: Os principais participantes na indústria de pagamentos são:
Comprador, Emissor, Bandeira do cartão, Adquirente e Comerciante;

- Comprador: Aquele que realiza a compra ou pagamento pelo serviço;
- Emissor: A instituição financeira que emitiu o cartão;
- Bandeira do cartão: A rede que processa o pagamento do cartão, facilitando a comunicação dos envolvidos da transação(Exemplos de bandeiras de cartão incluem Visa, Mastercard, American Express, entre outros.);
- Adquirente: A instituição financeira que processa os pagamentos para o Comerciante, fornecendo a ele os equipamentos e serviços para receber os pagamentos;
- Comerciante: A pessoa ou empresa que vende produtos ou fornece serviços.

Fluxos de pagamento:

- Pull: No pagamento pull, a iniciativa para iniciar a transação parte do receptor e o comprador deve aceitar a solicitação para ter os fundos retirados de sua conta(Exemplos: Cartão de Crédito, Cartão de Débito e Pix).
- Push: No caso do push, a iniciativa para iniciar a transação parte do pagador, ele já inicia a transação autorizando os fundos de serem retirados de sua conta(Exemplos: Transferência bancária e Pix).

2.1.b: As diferenças entre Adquirente, Sub-Adquirente e Gateway de Pagamento:

- Sub-Adquirente: É um intermediador de pagamentos que atua entre o Comerciante e o Adquirente, eles oferecem serviço de processamento de pagamentos trabalhando junto do Adquirente, simplificando o processo e deixando menos burocrático para o Comerciante;
- Gateway de Pagamentos:É um serviço que facilita a autorização e processamento de pagamentos, coletando informações sensíveis(Como número de cartão de crédito) e transmitindo ao Adquirente para autorizar, além de possuir sistemas antifraude e facilidade para reportar problemas.

As mudanças que eles trazem ao fluxo de pagamento:

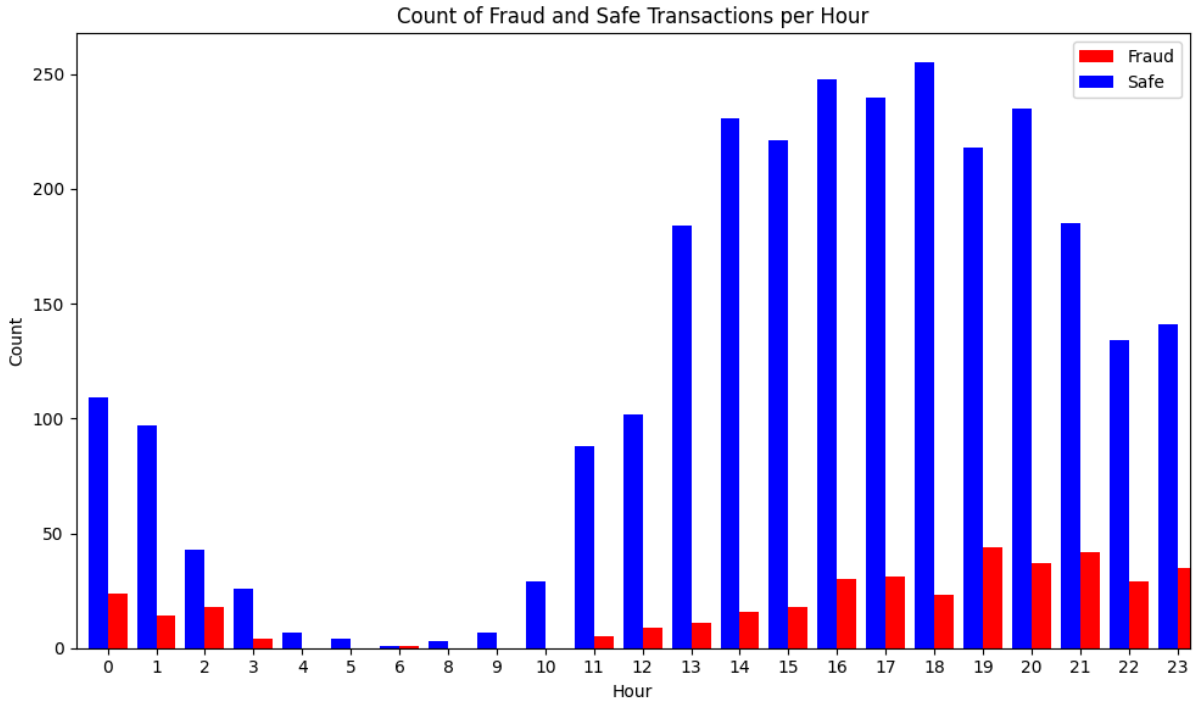
- Para o Adquirente não há mudanças significativas no processo, ele continua recebendo as requisições do Comerciante e passando para Bandeira do cartão;
- Para os Sub-Adquirente eles passam a fazer a ponte entre o Comerciante e o Adquirente e acabam sujeitos a regulamentação local de onde estão atuando;
- O Gateway de Pagamentos fica entre o Comerciante e o Adquirente, formando uma camada no processo, especialmente garantindo a segurança dos dados das transações.

2.1.c: As diferenças entre Chargeback e Cancelamento:

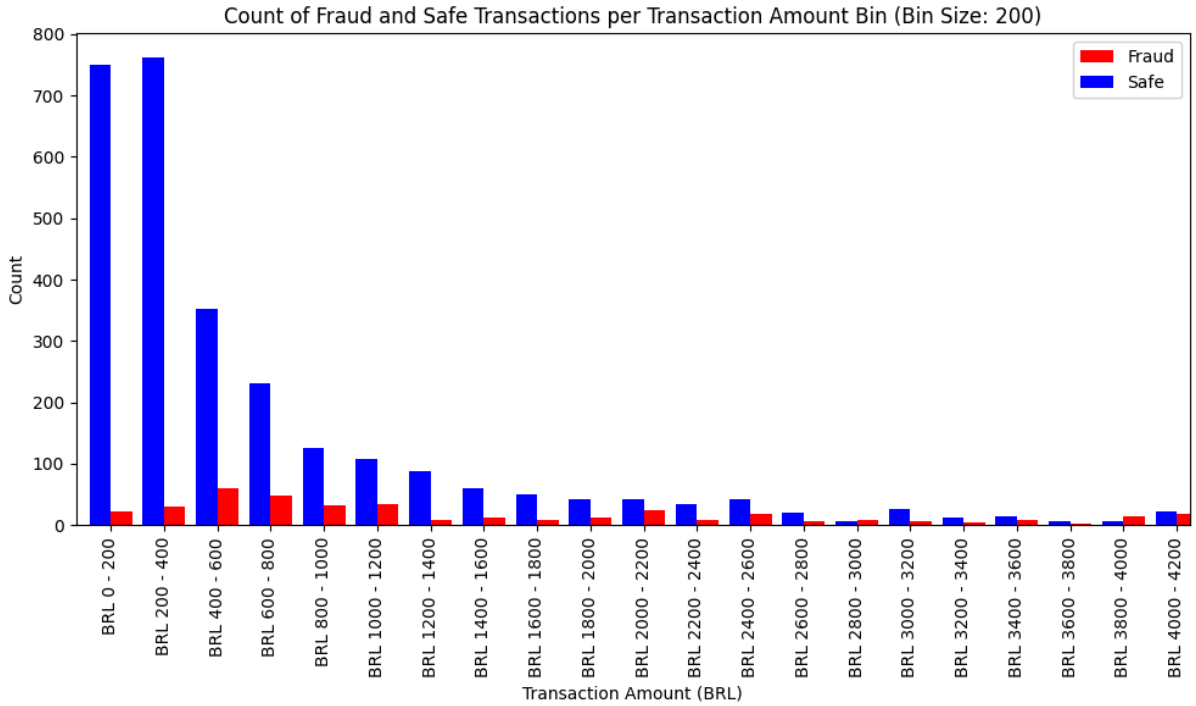
- O Chargeback é iniciado pelo titular do cartão com o Emissor do cartão e pode ocorrer por motivos de fraude, tanto do cliente(Tentar ficar com o produto sem pagar), quanto do comerciante(Recebeu mas não entregou o produto ou serviço), o processo em si envolve todas as etapas do fluxo de pagamentos e envolve investigação de documentos tanto do Comprador quanto do Comerciante, além disso Chargebacks influenciam nas métricas e geram custos para todos os envolvidos;
- Já o cancelamento ocorre diretamente entre o Comprador e Comerciante, sendo uma decisão unilateral por parte do comprador, seja ela por motivos de insatisfação com o produto ou serviço ou mesmo por ter mudado de ideia, nesse caso o processo não envolve toda a cadeia de pagamento, portanto não gera custos adicionais.

2.2: Eu responderia dizendo que entendo a preocupação e que faremos o possível para resolver. Explicaria o motivo de ter sido negado e como podemos recorrer. Logo em seguida, iria buscar junto do cliente providenciar mais documentos para auxiliar na defesa, tendo em vista que o motivo da recusa foi documentação insuficiente. Por último, garantiria que assim que tiver mais informação, entrarei em contato. E deixaria claro que ele pode entrar em contato novamente para qualquer nova dúvida ou informação.

3.1: Primeiramente gostaria de apresentar 2 gráficos, um sobre os horários x transações e outro sobre os valores x transações.



Com esse gráfico podemos observar que após as 19h a proporção de fraudes por transações autênticas aumenta e esse padrão segue até às 2h.

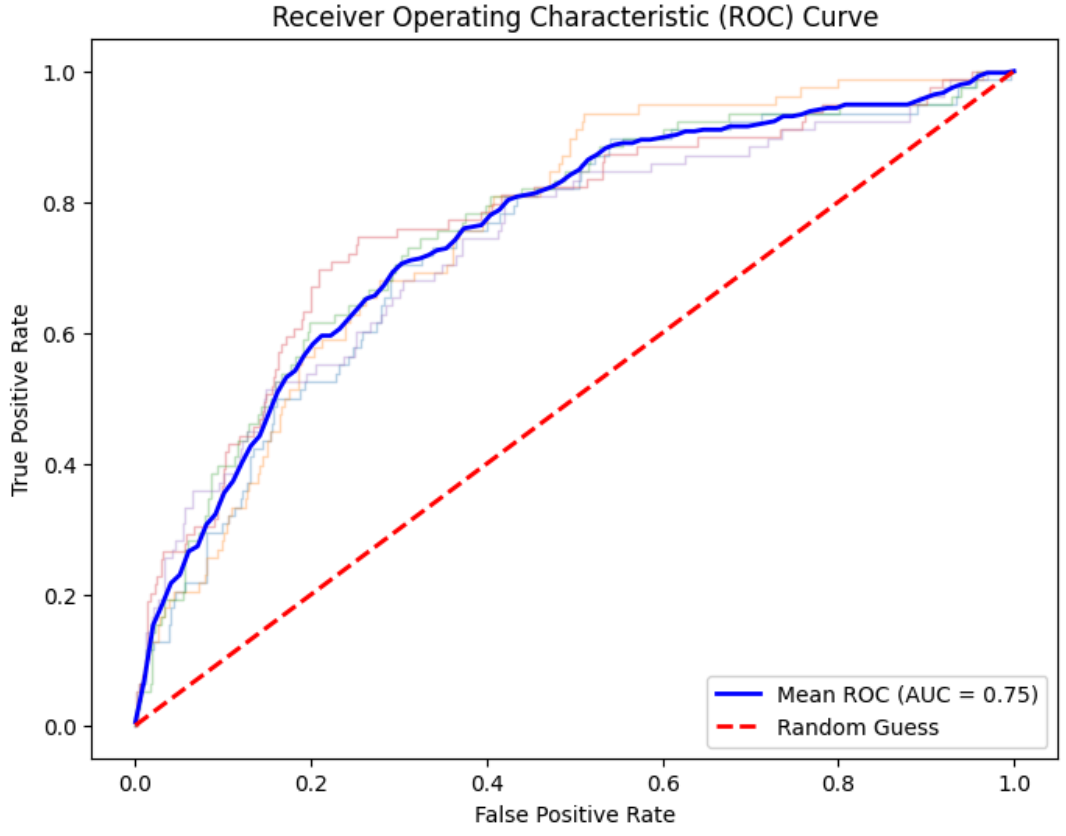


E esse segundo gráfico nos permite observar que a taxa de fraudes por transações autênticas aumenta à medida que o valor das transações aumenta, chegando em alguns pontos a quantidade de fraudes ser maior que a de transações legítimas.

3.2: Acredito que os históricos de transação do usuário ajudariam, assim como o tempo desde a criação da conta e a data das últimas atividades, essa última tendo em vista a possibilidade de roubo de contas inativas. Além disso, pelo histórico das transações antigas do cliente poderíamos ver se o horário e valores condizem com o histórico do cliente.

3.3: Medidas de autenticação de transações mais firmes(Solicitar confirmação por parte do usuário), além de manter os sistemas antifraude sempre atualizados e dentro do que há de mais avançado na indústria e instruir os usuários sobre como proteger seus dados fora do ambiente de pagamento da empresa.

3.4: Alinhado com a estratégia mencionada anteriormente, optei por desenvolver um algoritmo de machine learning simples utilizando regressão logística. Para aprimorar a precisão do modelo, ajustei a threshold com base na Curva Característica de Operação do Receptor (ROC) da validação cruzada.
Conduzi uma série de testes para determinar o tamanho ideal da amostra usada para treinar o modelo, visando garantir sua eficácia e generalização.
Pude observar é quanto maior a threshold menos falso positivo porém mais transações passavam ser serem pegas, sendo assim optei por deixar a threshhold em 0.3, pois ao mesmo tempo que pegava 24,3% das transações fraudulentas, tinha uma taxa de um acerto para cada 1,47 falso positivo.



No gráfico abaixo temos os resultados da validação cruzada, podemos observar que o modelo não apresenta variações e portanto responde bem a novas entradas.

No código o dicionário chamado payload pode ter seus valores trocados para validar transações.

```
payload = {
    "transaction_id" : 2342357,
    "merchant_id" : 29744,
    "user_id" : 97051,
    "card_number" : "434505******9116",
    "transaction_date" : "2019-11-30T23:16:32.812632",
    "transaction_amount" : 373,
    "device_id" : 285475
}
```

```
   card_number  transaction_date  transaction_amount
0    -1.717973           1.11817           -0.444061
Row 1 - Prediction: 0
{'transaction_id': 2342357, 'recommendation': 'approve'}
```

Aprovando caso ele preveja False para has_cbk.

```
   card_number  transaction_date  transaction_amount
0    -1.717973           1.11817            3.332809
Row 1 - Prediction: 1
{'transaction_id': 2342357, 'recommendation': 'deny'}
```

E negando caso ele preveja True para o has_cbk.