Reunião 24/08

Participantes - João, William, Giovanni, Thiago (iFood), Guilherme (iFood)

Pautas tratadas e sugestões de mudanças

1. **Descrição do Problema**

O problema realmente é de que a empresa possui modelos de Machine Learning rodando em produção, os quais apresentam comportamentos inesperados em dados do mundo real, como em situações de fraudadores que tentam burlar o sistema. Entretanto, foi deixado claro pelo Thiago de que esses *bugs* nos modelos também acontecem com usuários nos quais não deveria ser detectado como fraude e acaba detectando. Ou seja, o foco não necessariamente tem que ser no de ataque, mas sim de compreender o comportamento das features que podem causar essas anomalias, uma vez que acontece tanto com atacantes quanto com clientes normais que acabam por ter uma compra não autorizada por essa detecção falha.

2. **Artigos relacionados ao tema**

Os artigos foram passados para o pessoal da empresa e eles irão analisar posteriormente à reunião.

3. **Escolha do dataset**

Os *datasets* apresentados, principalmente os dois primeiros[1], foram vistos como bons ponto de partida para o problema, uma vez que o foco está em detectar fraude e não fraude a partir dos dados. Entretanto, alguns pontos foram colocados, como escolher um dataset onde o foco não é uma caracterização binária, mas de multiclasse, pois assim teríamos uma forma diferente de validar nossa implementação. Além disso, foi recomendado que se observasse códigos já feito no kaggle destes datasets para análise de features criadas a partir dos dados que já se tem, buscando algo um pouco mais com o que se tem na realidade da empresa, a exemplo de “quanto o cliente gastou nos últimos 6 meses?” “quanto o cliente gastou na última semana?” “qual a média de gastos?”. Ou seja, não é só colocar o *dataset* do jeito que ele está para o modelo, mas também fazer um processo de *feature engineering*.

Também deixaram um alerta para ver se as features dos dados não estão normalizadas ou em PCA.

4. **Treinamento de modelos para teste**

Os modelos propostos foram vistos como uma solução plausível para nossa análise, porém foi feita a recomendação de que criássemos uma rede neural, além de trocar o modelo de árvore (Random Forest) por um MLP.

[1] *https://www.kaggle.com/datasets/ealaxi/paysim1* e *https://www.kaggle.com/datasets/johancaicedo/creditcardfraud*

5. **Ferramentas para exploração de vulnerabilidade**

Segundo o Thiago e Guilherme, essas ferramentas podem realmente ser muito úteis no nosso trabalho, porém apenas futuramente e recomendaram simplificarmos um pouco nosso método para posteriormente fazer uso das ferramentas de geração de exemplos adversariais, pois assim iremos conseguir entender melhor como elas funcionam de verdade, tendo consciência do que se esperar de resultado e não vai ser algo tão complexo. Portanto, a sugestão foi fazer o trabalho de maneira mais simples a princípio visando ao entendimento do problema, usando uma espécie de testes com determinada visão de negócios embutida. Por exemplo, a visão atribuída é a de que o comportamento da feature "número de transações" deveria ser linear, isto é, quanto maior o número de transações, maior a probabilidade de resultar em fraude. Assim, nosso papel é analisar este comportamento, ver se realmente está se comportando assim, testar como o modelo se comporta ao mudarmos o label de uma entrada, colocar um ruído etc.

Slide da Reunião -
https://docs.google.com/presentation/d/1IiavYP9_abls5FkE7Mai7d6Mb9F4jl2bD-tuVbtpzHc/edit?usp=sharing