



TOCANTINS
GOVERNO DO ESTADO



Alunos: Iago Batista Antunes Leobas, Francisco Raphael F. de Araujo, Gabriel Teixeira.

Disciplina: Mineração de Dados

Data: 10/09/2022

Trabalho avaliativo do processo de mineração de dados - parte 1

Reflexo da pandemia e dos novos hábitos de consumo, o comércio eletrônico mundial teve crescimento de 55,5% desde 2020. Os principais meios de pagamento dos brasileiros para as compras online são efetuadas através do cartão de crédito, porém assim como o crescimento do e-commerce, também houve crescimento em golpes e fraudes.

Conjunto Escolhido: Detecção de fraudes e anomalias em cartões de crédito.

No mundo digital de hoje, onde trilhões de transações com cartões de crédito acontecem por dia, a detecção de fraudes é um desafio. Nosso DataSet, traz um conjunto de dados originado por um instituto(não divulgado por questões de privacidade dos clientes) que passou por situações de fraudes e transações que fogem do padrão(consideradas anomalias transacionais).

O conjunto de dados presente para análise é composto pelas seguintes colunas: distance from home, distancefromlast_transaction, ratio to median purchase price, repeat_retailer, used_chip, used pin number, online_order, fraud.

Contexto: A mineração de dados é feita no contexto de casos onde são compras apenas no cartão de crédito e online ao mesmo tempo, pois a ideia é a detecção de fraudes, caso haja uma irregularidade em compras presenciais o contexto passa a ser encaixado como roubo.

Explicação de Recursos:

- **Distance_from_home** - Distância de casa em que a transação ocorreu - Quantitativa.
- **Distance_from_last_transaction** - Distância que a última transação aconteceu - Quantitativa.
- **Ratio_to_median_purchase_price** - Razão da transação do preço de compra para o preço de compra mediano - Qualitativa.
- **Repeat_retailer** - Transação que aconteceu do mesmo varejista - Qualitativa.
- **Used_chip** - É a transação através do uso do chip (cartão de crédito). - Qualitativa.
- **used_pin_number** - Transação aconteceu usando o número código de segurança - Qualitativa.
- **online_order** - A transação é um pedido online - Qualitativa
- **Fraud** - A transação é fraudulenta - Qualitativa.

O script para extrair os dados está disponível em:

<https://www.kaggle.com/datasets/dhanushnarayananr/credit-card-fraud?resource=download>

Hipótese

Nesta análise, procuramos entender todas as possíveis formas e condições existentes que cada transação pode tomar. De início, tratamos cada variável particularmente, procurando entender como encontrar as ações fraudulentas de forma mais direta e consciente.

1. O primeiro ponto foi tratar o campo **Used_chip**, observamos que cada compra utilizando o cartão físico, realiza o uso físico do chip do cartão. O que quer dizer que a transação não foi mediante a alguma fraude, pois quem utilizou do cartão fisicamente ou foi o portador do cartão, ou foi alguém que praticou furto/roubo, o que tira de nossas análises de fraudes.

Sendo assim, restaram as compras que não utilizaram o chip. Essas compras elas devem ter atenção redobrada, pois podem partir delas alguma transação fraudulenta, visto que compras utilizando o chip passam por um processo de validação físico para averiguar falhas. (vamos alterar ainda)

2. Outra tratativa que devemos separar, diz respeito ao campo, **online_order** (compras terem sido efetuadas por meio online). Dessa forma, compras que foram efetuadas de forma online são nosso foco alvo. Essas compras são feitas mediante ao vazamento de dados, utilizando dados roubados de cartões ou por meio da clonagem dos dados do cartão (compras online utilizam o CVV). (vamos alterar ainda)
3. Usamos o **Ratio_to_median_purchase_price** que nos revela a mediana do preço médio de compra. Com isso, consideramos que a grande maioria das compras devem ser até 10 vezes o valor de compra médio de cada usuário. Pois assim, conseguimos separar compras que estão distantes do normal das transações.