



TOCANTINS
GOVERNO DO ESTADO



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

Disciplina: Mineração de Dados

Data: 26/09/2022

Trabalho avaliativo do processo de mineração de dados

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. Compras que ocorrem distante de casa, e distante da última transação são possivelmente fraudulentas por destoar do padrão “normal” de consumo.
2. Os usuários geralmente fazem compras online em lugares familiares, ou em casa ou no trabalho, por isso compras realizadas longe de casa ou muito distante da última compra são suspeitas.
3. O valor médio gasto nas compras que ultrapassa de até 10 vezes o **ratio_to_median_purchase_price** são possivelmente compras suspeitas, pois fogem muito do padrão de valor gasto do usuário.