

# Descrição do Problema

## Contexto

No cenário atual, o volume de transações financeiras digitais cresce exponencialmente devido ao aumento do uso de serviços online, bancos digitais, cartões virtuais e e-commerces. Esse crescimento torna os sistemas mais vulneráveis a fraudes, exigindo soluções automatizadas para detectar atividades suspeitas com rapidez e eficiência.

## O Problema

A detecção de transações financeiras fraudulentas é um problema de classificação onde, dado um conjunto de informações sobre cada transação, é necessário identificar se aquela transação é legítima ou fraudulenta.

Detectar fraudes em tempo real é fundamental para impedir que transações fraudulentas sejam concluídas, minimizando prejuízos financeiros e danos à reputação da instituição.

## Desafios

- **Desequilíbrio entre classes:** Transações fraudulentas são raras em comparação com as legítimas, dificultando o aprendizado do modelo.
- **Dinâmica e adaptabilidade:** Fraudadores mudam suas estratégias constantemente, exigindo modelos que consigam se adaptar rapidamente.
- **Velocidade de decisão:** Sistemas de detecção precisam ser eficientes para analisar milhões de transações em tempo real, sem impactar a experiência do usuário.
- **Complexidade dos dados:** Os padrões de fraude são difíceis de identificar manualmente.

## Impactos

- **Financeiros:** Redução direta de perdas financeiras causadas por fraudes.
- **Confiança:** Manutenção da confiança dos clientes e parceiros e melhora na experiência do usuário.
- **Operacionais:** Otimização de recursos com sistemas automáticos, reduzindo a necessidade de análise manual.

# Instruções para Execução

Link do colab:  
<https://colab.research.google.com/drive/1hZZgAABAhU1jUKSgsSU4wcP2yulnukOU#scrollTo=O-OQZQYFmOIX>

Link do dataset:  
<https://www.kaggle.com/datasets/aryan208/financial-transactions-dataset-for-fraud-detection>

1. Subir o notebook no Google Colab.
2. Importar o dataset diretamente do Kaggle ou carregar o arquivo .csv localmente.
3. Executar as células sequencialmente:
  - Instalar dependências
  - Carregar e preparar os dados
  - Treinar o modelo
  - Avaliar os resultados

## Informações sobre o Dataset utilizado

### Introdução ao dataset: (Financial Transactions Dataset for Fraud Detection)

O dataset foi pego no kaggle, ele é um dataset público de larga escala. Ele tem uma nota de usabilidade de 10 (nota máxima no kaggle), é constantemente atualizado, com sua última atualização à dois meses atrás.

### Overview de atividades:

#### 🌟 Activity Overview



### Informações gerais:

Este conjunto de dados contém 5 milhões de transações financeiras geradas sinteticamente, projetadas para simular o comportamento do mundo real para pesquisas de detecção de fraudes e aplicações de aprendizado de máquina. Cada registro de transação inclui os seguintes campos:

**Detalhes da Transação:** ID, timestamp (carimbo de data/hora), contas do remetente/destinatário, valor, tipo (depósito, transferência etc.)

**Características Comportamentais:** tempo desde a última transação, pontuação de desvio de gastos, pontuação de velocidade, pontuação de anomalia geográfica

**Metadados:** localização, dispositivo utilizado, canal de pagamento, endereço IP, hash do dispositivo

**Indicadores de Fraude:** rótulo binário de fraude (is\_fraud) e tipo de fraude (por exemplo, lavagem de dinheiro, tomada de conta)

O conjunto de dados segue padrões realistas de fraude e anomalias comportamentais, tornando-o adequado para:

- Modelos de classificação binária e multiclasse
- Sistemas de detecção de fraudes
- Detecção de anomalias em séries temporais
- Engenharia de atributos e explicabilidade de modelos