# Aprendizado de Máquina na Identificação de Transações Fraudulentas: Comparação entre LightGBM e LightGBM+SMOTE

JOÃO HENRIQUE DIAS DOS SANTOS[1];MURILO SALEM[2]
ANDERSON PRIEBE FERRUGEM[3]-DANIEL HENRIQUE BARRETOS[4]

[1]UniversidadeFederaldePelotas – joao.dias@ufpel.edu.br
[2]UniversidadeFederaldePelotas – ferrugem@inf.ufpel.edu.br
[3]UniversidadeFederaldePelotas – mcsalem@inf.ufpel.edu.br
[4]UniversidadeFederaldePelotas-dhspbarretos@inf.ufpel.edu.br

## 1. INTRODUÇÃO

As fraudes financeiras e fiscais representam um problema crítico em escala global, caracterizando-se por ações intencionais de ocultar, manipular ou falsificar informações com o objetivo de obter vantagem econômica ilícita. No contexto bancário, essas fraudes se manifestam em transações suspeitas que podem envolver desde o uso de identidades falsas até movimentações atípicas de valores, gerando prejuízos significativos tanto para instituições quanto para a sociedade, ao comprometer recursos que poderiam ser destinados a serviços públicos essenciais.
A detecção desse tipo de fraude é desafiadora, principalmente devido ao forte desbalanceamento dos dados: enquanto a grande maioria das transações é legítima, apenas uma fração mínima corresponde a fraudes. Esse cenário torna os modelos de aprendizado de máquina suscetíveis a enviesamento em favor da classe majoritária, prejudicando a identificação da classe de maior interesse. O aprendizado de máquina, pode ser entendido como pipeline de classificação (descreve o processo de treino do modelo), começando com a entrada de dados, seja imagem, vídeos ou no caso, um dataset de informações financeiras anonimizadas, representação obtida por (PCA), que comprimem e preservam padrões estatísticos sem revelar os dados originais e finalizando com a classificação em que o modelo aprende a relacionar os padrões das features com as classes (transações legítimas ou fraudulentas). Esse aprendizado ocorre matematicamente por meio de uma função de soma ponderada, onde cada variável de entrada recebe um peso que indica sua importância. O resultado dessa soma é transformado por uma função de ativação, gerando uma probabilidade associada a cada classe, além dos pesos aplicados às entradas, existe também o bias (viés), que atua como um parâmetro adicional permitindo deslocar a função para melhor ajuste aos dados. Em termos geométricos, o bias funciona como o intercepto de uma reta, evitando que o modelo seja limitado a passar pela origem. Isso amplia a flexibilidade do classificador e reduz a chance de enviesamento na decisão. Assim, o modelo consegue predizer. Neste experimento, foram empregados dois métodos complementares: o LightGBM (Light Gradient Boosting Machine), um algoritmo de classificação supervisionada eficiente para lidar com grandes volumes de dados, e a técnica de SMOTE (Synthetic Minority Oversampling Technique), aplicada para gerar amostras sintéticas da classe minoritária e mitigar o impacto do desbalanceamento. O objetivo central é avaliar o desempenho dos modelos na classificação de transações fraudulentas, por meio de métricas como AUC, precisão, recall e F1-score, além de discutir a relevância das variáveis mais importantes na tomada de decisão.

## 2. METODOLOGIA

A metodologia utilizada foi um estudo de caso, através de experimento de treino e comparação de modelos de aprendizado de máquinas. O dataset utilizado contém 284.807 transações, das quais apenas 0,17% são fraudulentas, evidenciando o

forte desbalanceamento. Por ter passado por PCA, não foi necessário aplicar limpeza adicional ou padronização, pois os dados já estavam preparados para o treinamento o dataset foi dividido em três subconjuntos: 64% para treino, 16% para validação — utilizada para ajuste de hiperparâmetros e aplicação do *early stopping* — e 20% para teste, destinado à avaliação final do desempenho. Essa divisão garante que o modelo não apenas memorize os dados, mas aprenda padrões generalizáveis, reduzindo o risco de *overfitting*..Inicialmente, realizou-se uma análise exploratória do dataset, permitindo compreender suas dimensões, quantidade de variáveis e distribuição das classes. Observou-se um desbalanceamento extremo: a classe 0, correspondente a transações legítimas, representava aproximadamente 99,83% dos dados, enquanto a classe 1, de transações fraudulentas, correspondia a apenas 0,17%. Em seguida, os dados foram organizados separando-se as variáveis preditoras (*features*) da variável alvo (*target*), que indica se uma transação é fraude ou não.

O primeiro treinamento foi realizado com o algoritmo Light Gradient Boosting Machine (LightGBM), que constrói árvores sucessivas de decisão (*boosting*), avaliando a cada iteração o desempenho no conjunto de validação. Para evitar treinamento excessivo sem ganhos reais, utilizou-se o mecanismo de *early stopping*, configurado para interromper o processo caso não houvesse melhora na métrica AUC após 50 iterações consecutivas. Essa estratégia assegura que o modelo atinja seu ponto ótimo, prevenindo sobreajuste e promovendo maior capacidade de generalização para dados não vistoscompensando o desbalanceamento atribuindo mais peso aos poucos exemplos de fraude, o que resulta em métricas globais altas (como ROC AUC), mesmo sem uma boa detecção efetiva de fraudes. . Já o SMOTE é uma técnica de balanceamento de dados balanceamento, para datasets desbalanceados, ele pega um classe minoritária o multiplica sinteticamente, ou seja assim ele replica essas variaveis com padrões semelhantes assim temos um conjunto de dados maior e mais equilibrado, a ser utilizado para o treinamento do modelo, de forma mais detalhada gera instâncias sintéticas da classe 1, enriquecendo o conjunto de treino com mais padrões de fraude. Assim, é possível comparar os dois cenários — com e sem SMOTE — e avaliar, por meio de métricas como PR AUC, recall e F1-score da classe 1, qual modelo apresenta melhor capacidade de identificar fraudes reais. Ao não aplicarmos um balanceamento pode-se enviesar o modelo, causando overfitting, assim o modelo não vai decorar os dados apenas, e sim analisar os padrões.

## 3. RELATOS E IMPACTOS GERADOS

O modelo LightGBM apresentou desempenho de alto nível na detecção de fraudes em ambos os cenários, evidenciando sua capacidade de lidar com desbalanceamento extremo do dataset. No modelo original, sem balanceamento explícito, obteve-se **ROC AUC = 0,962** e **PR AUC = 0,740**, conseguindo capturar aproximadamente 90% das fraudes. O recall da classe 1 foi de 0,89, a precisão de 0,65 e o F1-score de 0,75, refletindo bom equilíbrio entre captura de fraudes e número de falsos alarmes. A matriz de confusão mostrou 87 verdadeiros positivos, 11 falsos negativos, 56.817 verdadeiros

negativos e 47 falsos positivos. As variáveis mais relevantes para a detecção foram **V14, V4 e V7**, além de *Amount* e *Time*.

O treinamento foi controlado por early stopping (interrompido se não houvesse ganho de AUC após 50 iterações), e o melhor desempenho foi obtido na 131ª árvore.
- ROC AUC: 0,962
- PR AUC: 0,740
- Recall (classe 1): 0,89
- F1-score (classe 1): 0,75

Após a aplicação do **SMOTE** e *undersampling* da classe majoritária, observou-se melhora significativa nas métricas: **ROC AUC = 0,984** e **PR AUC = 0,852**. O recall da classe 1 permaneceu alto (**0,88**), a precisão e o F1-score subiram para 0,80, evidenciando maior capacidade de identificação das fraudes reais. A matriz de confusão indicou 86 fraudes corretamente detectadas e 12 não identificadas, com 32 falsos positivos, mostrando aumento do custo em termos de alertas falsos, mas ganho na sensibilidade do modelo.

Resultados após balanceamento com SMOTE:
- ROC AUC: 0,984
- PR AUC: 0,852
- Recall (classe 1): 0,88
- F1-score (classe 1): 0,80

O primeiro modelo (LightGBM) mostrou-se eficiente ao lidar com dados desbalanceados, alcançando métricas elevadas e baixo número de falsos positivos. Contudo, sua detecção de fraudes reais ainda foi limitada pela escassez de exemplos da classe 1. Já o segundo modelo (LightGBM + SMOTE) conseguiu ampliar a capacidade de generalização, aumentando recall e F1-score, mesmo ao custo de alguns falsos positivos adicionais. Essa abordagem se mostra mais adequada em aplicações reais de detecção de fraudes, onde a prioridade é não deixar passar transações fraudulentas. A comparação entre os dois modelos evidencia que o LightGBM já lida bem com desbalanceamento extremo, mas o uso do SMOTE ampliou a capacidade do modelo em generalizar e detectar fraudes reais. Conclui-se que a combinação LightGBM + SMOTE é a abordagem mais eficaz, equilibrando alta detecção de fraudes com controle aceitável de falsos positivos.

## 4. REFERÊNCIAS BIBLIOGRÁFICAS


- BURKOV, Andriy. *The Hundred-Page Machine Learning Book*. Andriy Burkov, 2019.
- LIGHTGBM. Documentation. Disponível em: https://lightgbm.readthedocs.io/en/stable/.
- MILLER, J.; JUNGER, S. Fraud Detection in Financial Transactions. *Journal of Financial Security*, v.12, n.2, p. 45–59, 2010.
- GPREDÁ, Gabriel. *Credit Card Fraud Detection Predictive Models*. Kaggle Notebook. Disponível em: https://www.kaggle.com/code/gpreda/credit-card-fraud-detection-predictive-models/notebook
- ELGENDY, Mohamed. *Deep Learning for Vision Systems*. Manning Publications, 2020.