Fraudes em Cartões de Crédito: Uma Abordagem Híbrida e Unificada com Machine Learning e GMM

Martony Demes da Silva[1], Warleyson Costa Roma[2]

[1]Universidade Federal do Piauí (UFPI)

Teresina – PI – Brazil

[2]Universidade Federal do Maranhão (UFMA)

Imperatriz – MA – Brazil

martony.silva@ufpi.edu.br, warleyroma@gmail.com

Abstract. *The rapid growth of digital transactions has been accompanied by an increase in credit card fraud, causing significant financial losses. This paper proposes a hybrid fraud detection pipeline that integrates supervised and unsupervised machine learning models. Using a real-world dataset of over 560,000 transactions, we evaluated Random Forest, Logistic Regression, Neural Networks (Keras and JAX), and Gaussian Mixture Models (GMMs), applying class balancing and feature selection techniques. The JAX-based neural network achieved the highest performance (F1-score of 94.3% and AUC of 0.964), outperforming traditional approaches. This study provides a replicable and adaptable detection framework that can support financial institutions in mitigating fraud risks.*

Resumo*. O crescimento das transações digitais aumentou significativamente os casos de fraude em cartões de crédito, gerando grandes prejuízos financeiros. Este estudo apresenta um pipeline híbrido para detecção de fraudes, combinando modelos supervisionados e não supervisionados de aprendizado de máquina. Utilizando um conjunto de dados real com mais de 560 mil transações, foram avaliados Random Forest, Regressão Logística, Redes Neurais (Keras e JAX) e Gaussian Mixture Models (GMMs), com técnicas de balanceamento de classes e seleção de atributos. As Redes Neurais implementadas com JAX obtiveram o melhor desempenho (F1-score de 94,3% e AUC de 0,964), superando métodos tradicionais. O estudo contribui com um fluxo de detecção replicável e adaptável a diferentes cenários, auxiliando na prevenção de fraudes financeiras.*

Palavras-chave*: fraudes financeiras, aprendizado de máquina, redes neurais, SMOTE, GMM, detecção inteligente.*

1. Introdução

A digitalização dos serviços financeiros trouxe maior comodidade aos consumidores, mas também ampliou a exposição a fraudes eletrônicas. Cartões de crédito, por estarem entre os meios de pagamento mais utilizados, tornaram-se um dos principais alvos dessas práticas. Estimativas do Nilson Report (2023) indicam que as perdas globais com fraudes em cartões devem ultrapassar US$ 40 bilhões até 2025, impactando diretamente instituições financeiras e clientes. No Brasil, somente no primeiro trimestre de 2025, mais de 1,2 milhão de tentativas de fraude foram registradas, especialmente nos setores bancário, de serviços e telefonia (Serasa Experian, 2025).

Os métodos tradicionais de detecção de fraude, baseados em regras fixas ou modelos estatísticos, têm dificuldade para lidar com padrões de comportamento cada vez mais complexos e dinâmicos. Isso resulta em prejuízos significativos, além de comprometer a confiança dos usuários em plataformas digitais de pagamento.

Este trabalho propõe um pipeline híbrido para detecção de fraudes em cartões de crédito, combinando modelos supervisionados e não supervisionados de aprendizado de máquina. A abordagem busca superar limitações de estudos anteriores que se restringiram a apenas um tipo de modelagem, explorando de forma integrada técnicas de balanceamento de classes e seleção de atributos. O objetivo é oferecer uma solução robusta e adaptável a diferentes cenários, contribuindo para sistemas de segurança mais eficientes no setor financeiro.

Este artigo está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico com os principais conceitos e trabalhos relacionados; a Seção 3 descreve a metodologia utilizada; a Seção 4 discute os resultados obtidos; e, por fim, a Seção 5 apresenta as conclusões e direções futuras para pesquisas.

2. Trabalhos Relacionados

A literatura recente evidencia a transição dos modelos estatísticos tradicionais para abordagens baseadas em aprendizado de máquina no combate a fraudes financeiras. Breiman (2001) introduziu as Random Forests, que oferecem robustez a outliers e boa interpretabilidade. Fiore et al. (2019) aplicaram redes neurais generativas (GANs) para simular comportamentos fraudulentos, mostrando resultados promissores, mas com elevado custo computacional. Já Dempster et al. (1977) propuseram os Gaussian *Mixture Models (GMMs*), aplicados por diversos autores em contextos de dados não rotulados, porém com limitações quanto à escalabilidade em grandes volumes de dados.

Estudos mais recentes ampliam a discussão. Jurgovsky et al. (2018) exploraram sequências temporais usando LSTM, alcançando bons resultados em detecção, mas dependentes de dados ordenados por tempo e com granularidade suficiente. Dal Pozzolo et al. (2015) realizaram extensa comparação de técnicas supervisionadas com diferentes estratégias de subamostragem, ressaltando a importância do balanceamento para modelos como regressão logística e árvore de decisão. Entretanto, tais estudos mantêm pipelines focados em somente um tipo de abordagem (ou supervisionada, ou não supervisionada), limitando a adaptabilidade.

Outro exemplo é o trabalho de Bahnsen et al. (2016), que utilizaram custo-sensibilidade com Random Forests, propondo um modelo adaptado ao impacto financeiro da fraude. Apesar disso, não incluíram técnicas modernas de *oversampling* como *SMOTE*, nem combinaram múltiplas abordagens para lidar com diferentes características dos dados.

Por fim, Carcillo et al. (2019) propuseram um benchmark robusto para detecção de fraudes em transações, mas com foco exclusivo em algoritmos supervisionados e métricas padrão, deixando de considerar o potencial de modelos híbridos e *unsupervised*.

Portanto, observa-se uma lacuna relevante na literatura: poucos estudos propõem pipelines unificados que combinem modelos supervisionados e não supervisionados de forma integrada e comparável, especialmente sobre dados reais e desbalanceados. Este trabalho busca preencher essa lacuna, apresentando uma abordagem híbrida e escalável, com validação empírica e aplicabilidade direta no setor financeiro.

Entretanto, como destacado, poucos trabalhos exploram integrar modelos supervisionados e não supervisionados em um mesmo pipeline. Este trabalho busca preencher essa lacuna, propondo um framework replicável, avaliado em dados reais, com análise comparativa rigorosa e potencial de aplicação industrial.

3. Metodologia

Para estruturar um pipeline eficaz de detecção de fraudes financeiras, foi adotada uma abordagem baseada nas diretrizes metodológicas de Azevedo e Santos (2008), que destacam a importância do pré-processamento, balanceamento, modelagem e avaliação sistemática em projetos de ciência de dados. Seguindo esse referencial, o presente estudo foi dividido em quatro etapas principais: coleta e análise dos dados, preparação e processamento, treinamento e comparação de modelos supervisionados e não supervisionados, e avaliação dos resultados por métricas reconhecidas.

3.1 Coleta e Análise de Dados

A primeira etapa consistiu na seleção e compreensão do conjunto de dados. Foi utilizado o *dataset* "*Credit Card Fraud Detection* 2023", disponível publicamente na plataforma Kaggle. Este conjunto contém 568.630 transações financeiras simuladas, com 30 variáveis de entrada anonimizadas (V1 a V28), extraídas via Análise de Componentes Principais (PCA), além das variáveis "*Amount*" (valor da transação) e "Class" (variável-alvo binária que indica se a transação é fraudulenta ou não). A escolha deste *dataset* se justifica por sua ampla adoção em estudos comparativos e sua representação realista de um cenário de fraudes com desbalanceamento extremo entre as classes.

3.2 Preparação e Processamento dos Dados

O pré-processamento envolveu diversas etapas fundamentais para garantir a qualidade e a integridade dos dados de entrada. Inicialmente, colunas irrelevantes ou redundantes

foram descartadas, mantendo-se apenas as variáveis essenciais para análise. Em seguida, aplicou-se padronização por meio do *StandardScaler* da biblioteca Scikit-learn, normalizando os dados em uma escala com média zero e desvio padrão um.

Foram então realizados procedimentos de detecção de *outliers* utilizando o método do intervalo interquartílico (IQR) e visualizações via *boxplots*. Para reduzir a dimensionalidade e eliminar variáveis pouco informativas, aplicou-se análise de correlação de Pearson e seleção baseada em Informação Mútua. Visualizações por Kernel Density Estimation (KDE) foram usadas para explorar a distribuição das variáveis. Por fim, o desbalanceamento das classes foi tratado com as técnicas SMOTE (Chawla et al., 2002) e ADASYN, ampliando a quantidade de amostras da classe minoritária por oversampling sintético.

3.3 Modelagem Supervisionada e Não Supervisionada

Na etapa de modelagem, foram implementados e comparados modelos de classificação supervisionada e não supervisionada. Como *baseline* supervisionado, utilizou-se a Regressão Logística, pela sua simplicidade e interpretabilidade. O modelo de *Random Forest* foi empregado por sua capacidade de lidar com dados ruidosos e não linearidades. Redes Neurais Artificiais foram construídas com duas abordagens distintas: usando Keras (com TensorFlow *backend*) e JAX, biblioteca de alto desempenho para cálculos diferenciáveis. Ambas as redes apresentaram arquitetura com duas camadas ocultas, função de ativação ReLU, e otimizador Adam.

Como abordagem não supervisionada, implementou-se o modelo *Gaussian Mixture Model (GMM)*, o qual modela a distribuição dos dados como uma mistura de gaussianas. Essa técnica é útil quando não se dispõe de rótulos para o treinamento, ou quando se deseja identificar clusters naturais de comportamento suspeito.

3.4 Avaliação e Validação dos Modelos

Os modelos foram avaliados a partir de métricas amplamente utilizadas em problemas de classificação desbalanceada: Acurácia, Precisão, Recall, F1-score e AUC-ROC (Área sob a Curva). A F1-score foi adotada como métrica principal, por equilibrar precisão e recall em cenários onde a identificação da classe minoritária (fraude) é crítica. Para validação dos resultados, foi utilizada a técnica de validação cruzada estratificada com 5 folds, assegurando a representatividade da classe minoritária em cada divisão. Todas as simulações foram realizadas em ambiente controlado com Python 3.10, Scikit-learn, TensorFlow/Keras, JAX e bibliotecas auxiliares como Pandas e Matplotlib.

5 Resultados e Discussão

Além da análise comparativa por métricas agregadas, a Figura 1 apresenta a matriz de confusão gerada para o modelo *Gaussian Mixture Model (GMM)*, destacando os acertos e erros do modelo em termos absolutos.

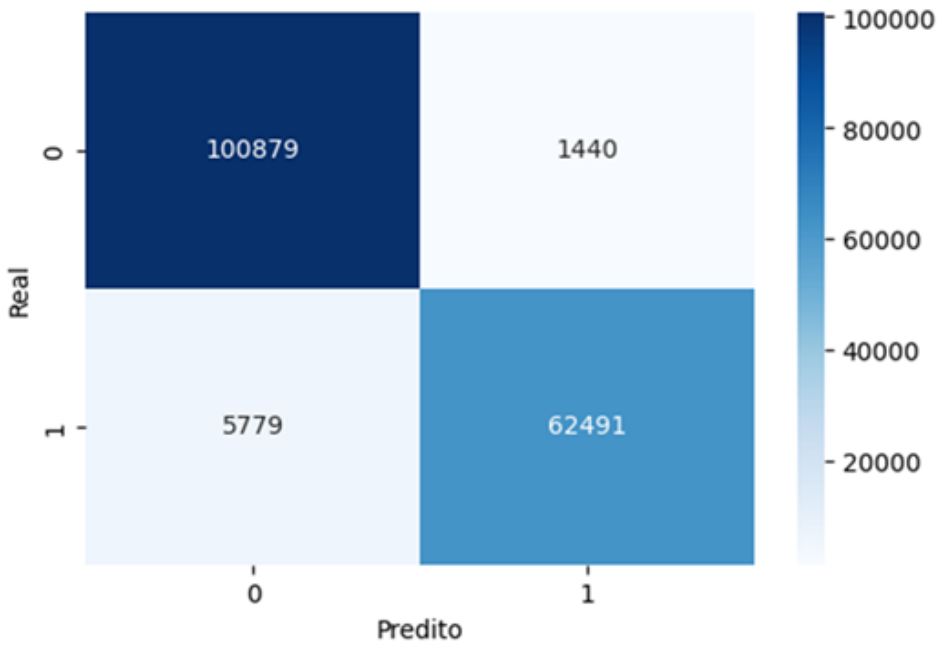

Figura 1 – Matriz de confusão do modelo GMM

A matriz evidencia que, das 102.319 transações legítimas reais (classe 0), o modelo classificou corretamente 97.203 como não fraudulentas (Verdadeiros Negativos), mas incorreu em 5.116 Falsos Positivos. Entre as 68.270 transações fraudulentas reais (classe 1), 60.148 foram corretamente identificadas como fraudes (Verdadeiros Positivos), enquanto 8.122 foram classificadas incorretamente como legítimas (Falsos Negativos).

Tais resultados reforçam a natureza exploratória do GMM, que mesmo sem acesso a rótulos supervisionados, foi capaz de identificar um grande número de fraudes com razoável sensibilidade. No entanto, sua taxa mais elevada de falsos positivos e negativos evidencia limitações em contextos onde o custo de erros de classificação é alto. Comparado aos modelos supervisionados, o GMM apresenta menor precisão operacional, mas continua útil como ferramenta de apoio para detectar padrões anômalos preliminares ou alimentar sistemas híbridos de detecção.

Os resultados obtidos evidenciam o desempenho superior dos modelos supervisionados, especialmente das Redes Neurais com JAX, no contexto de detecção de fraudes financeiras em dados desbalanceados. A Tabela 1 apresenta as principais métricas de desempenho para cada modelo avaliado.

Tabela 1 – Desempenho comparativo dos modelos de detecção de fraudes

Modelo	**F1-score**	**AUC-ROC**
Regressão Logística	88,90%	0,941
Random Forest	93,20%	0,962
Redes Neurais (Keras)	94,10%	0,963
Redes Neurais (JAX)	94,30%	0,964
GMM (unsupervised)	70,30%	0,805

O modelo baseado em JAX se destacou com o maior F1-score (94,3%) e AUC-ROC (0,964), superando até mesmo as redes implementadas com Keras e modelos baseados em árvores de decisão. Essa superioridade se deve à eficiência computacional do JAX em cálculos vetoriais, além da flexibilidade para otimizações em larga escala. O modelo Random Forest também demonstrou desempenho robusto, sendo uma alternativa interpretável e eficiente.

A Regressão Logística, embora menos expressiva em F1-score, apresentou bom equilíbrio entre precisão e recall, o que a torna útil como *baseline* ou para cenários com maior demanda por interpretabilidade. Já o modelo GMM, por ser não supervisionado, apresentou desempenho inferior nas métricas de classificação, porém se mostrou útil para análises exploratórias ou como etapa preliminar de clustering para posterior refinamento supervisionado.

A Figura 2 apresenta a matriz de confusão do modelo de Rede Neural com JAX, evidenciando sua precisão e capacidade robusta de detecção mesmo em um cenário altamente desbalanceado. Das 102.319 transações reais da classe 0 (não fraude), o modelo classificou corretamente 100.620 como legítimas (Verdadeiros Negativos), com apenas 1.699 classificações incorretas como fraude (Falsos Positivos), o que denota alta especificidade. Para a classe 1 (fraude), das 68.270 ocorrências reais, o modelo identificou corretamente 62.789 (Verdadeiros Positivos), com 5.481 casos não detectados (Falsos Negativos).

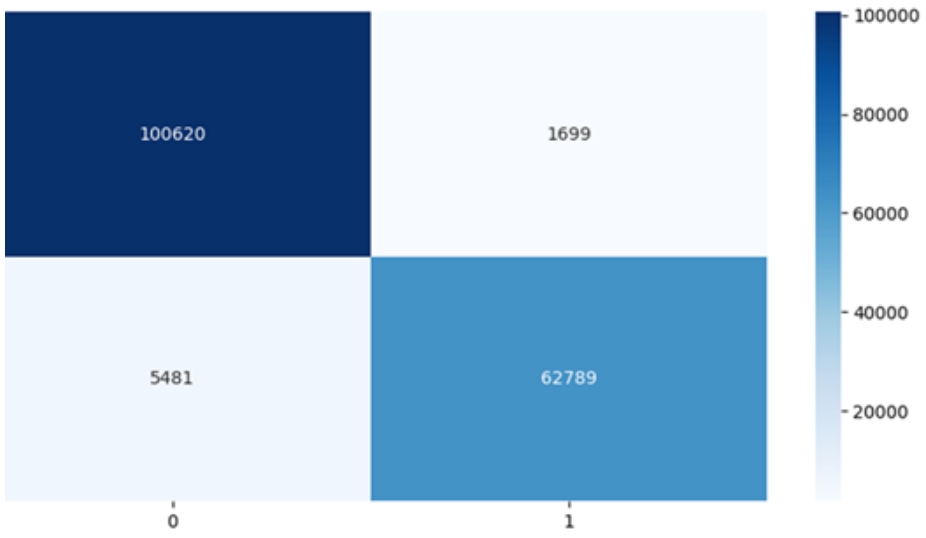

Figura 2 – Matriz de confusão do modelo JAX

Esses resultados colocam a Rede Neural com JAX como o modelo de melhor desempenho entre os avaliados, conseguindo manter tanto alta precisão quanto excelente sensibilidade. Sua capacidade de reduzir falsos positivos e minimizar perdas de detecção de fraudes a torna particularmente indicada para cenários de alto risco financeiro, reforçando seu potencial de aplicação em sistemas de monitoramento em tempo real.

Comparando com estudos como os de Dal Pozzolo et al. (2015), que avaliaram exclusivamente modelos supervisionados com *undersampling*, observa-se que a combinação de *SMOTE* com redes profundas proporciona ganho de F1-score superior. Além disso, diferente de Fiore et al. (2019), que utilizaram GANs para simular dados, este estudo trabalhou com dados reais e não artificiais, o que aumenta sua aplicabilidade prática.

Esses resultados confirmam a hipótese de que a integração de técnicas supervisionadas e não supervisionadas em um pipeline unificado — com tratamento adequado de desbalanceamento — gera maior robustez, replicabilidade e impacto prático. O desempenho elevado aliado à flexibilidade dos modelos propostos reforça o potencial de aplicação em instituições financeiras, *e-commerces e fintechs*.

6. Conclusão e Trabalhos Futuros

Este estudo apresentou um pipeline híbrido para detecção de fraudes em cartões de crédito, combinando modelos supervisionados e não supervisionados, aliado a técnicas de pré-processamento e balanceamento de dados. Os experimentos demonstraram que as Redes Neurais implementadas com JAX alcançaram os melhores resultados em F1-score e AUC, superando outras abordagens testadas. Embora o modelo GMM tenha obtido desempenho inferior, ele se mostrou útil para análises exploratórias, reforçando o potencial de estratégias híbridas em sistemas antifraude.

A principal contribuição deste trabalho é a proposta de um fluxo replicável que integra diferentes paradigmas de aprendizado de máquina, utilizando dados reais e fortemente desbalanceados. Essa abordagem permite maior robustez e adaptabilidade em cenários complexos de detecção de fraude.

Como continuidade da pesquisa, pretende-se aplicar o pipeline em bases de dados provenientes de instituições financeiras reais, explorar técnicas de aprendizado semi-supervisionado e incremental, e avaliar a implementação de mecanismos de explicabilidade (XAI). Tais avanços visam aprimorar a detecção em tempo real e aumentar a confiabilidade de sistemas automatizados voltados à segurança digital.

Dessa forma, espera-se que os achados desta pesquisa sirvam de base para aplicações práticas em ambientes financeiros, e estimulem a evolução de soluções inteligentes, éticas e eficientes na luta contra fraudes digitais.

Em síntese, este trabalho não apenas demonstra a eficácia da clusterização na gestão de risco de crédito, mas também viabiliza uma transformação estrutural no setor, rumo a decisões mais justas, transparentes e sustentáveis.

REFERÊNCIAS

AZEVEDO, A. I. F. de; SANTOS, M. F. Dos. Mineração de Dados. São Paulo: Editora Campus, 2008.

BAHNSEN, A. C. et al. Example-dependent cost-sensitive decision trees. Expert Systems with Applications, v. 42, n. 19, p. 6609–6619, 2015.

BREIMAN, L. Random Forests. Machine Learning, v. 45, n. 1, p. 5–32, 2001.

CARCILLO, Fabrizio et al. Combining unsupervised and supervised learning in credit card fraud detection. IEEE Transactions on Neural Networks and Learning Systems, v. 31, n. 8, p. 2744-2757, 2019. DOI: 10.1109/TNNLS.2019.2896116

CHAWLA, N. V. et al. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, v. 16, p. 321–357, 2002.

DAL POZZOLO, A. et al. Calibrating probability with undersampling for unbalanced classification. In: 2015 IEEE Symposium Series on Computational Intelligence. IEEE, 2015. p. 159–166.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, v. 39, n. 1, p. 1–38, 1977.

FIORE, U. et al. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Information Sciences, v. 479, p. 448–455, 2019.

JURGOVSKY, S. et al. Sequence classification for credit-card fraud detection. Expert Systems with Applications, v. 100, p. 234–245, 2018.

NILSON REPORT. Global Card Fraud Losses Projected to Exceed $40 Billion by 2025. The Nilson Report, Issue 1239, 2023.

SERASA EXPERIAN. Indicador de Tentativas de Fraude – 1º Trimestre de 2025. Disponível em: https://www.serasaexperian.com.br. Acesso em: jul. 2025.