Comparação de modelos preditivos para identificação de fraudes em cartões de crédito

Gilberto Luiz Angelice de Camargo (ESALQ/USP) *gilkiske@gmail.com*
Jailson dos Santos Silva (UFSC) *engjailsonsantos@outlook.com*
William Neves da Silva (UFSC) *williamrns88@gmail.com*

Resumo

A expansão das transações digitais ampliou os riscos associados a fraudes financeiras, exigindo soluções automatizadas que sejam eficazes e adaptáveis. Desta forma, este trabalho teve como objetivo comparar o desempenho de três algoritmos de aprendizado supervisionado, regressão logística, *XGBoost* e redes neurais artificiais, na detecção de fraudes em transações com cartão de crédito. Utilizou-se uma base de dados pública e sintética, marcada por um forte desbalanceamento entre as classes, simulando cenários reais. O processo envolveu pré-processamento de variáveis temporais e categóricas, aplicação de técnicas de balanceamento e otimização de hiperparâmetros. Os modelos foram avaliados com base em métricas como sensibilidade e precisão e o comportamento de cada um foi analisado em termos técnicos e operacionais. Como resultado, verificou-se que a regressão logística apresentou desempenho limitado, o *XGBoost* se destacou pela alta sensibilidade, e a rede neural demonstrou flexibilidade por meio da calibragem do limiar de decisão. Concluiu-se que, embora não exista um modelo ideal universal, o *XGBoost* se mostrou mais promissor em ambientes que priorizam detecção ampla de fraudes, desde que acompanhado de mecanismos que controlem os falsos positivos. A análise comparativa possibilitou uma compreensão estruturada sobre o desempenho e a aplicabilidade de cada algoritmo, contribuindo para decisões mais assertivas em sistemas antifraude.

Palavras-Chave: *Machine Learning, Ciência de Dados, Instituições Bancárias, Fraudes.*

Abstract

The expansion of digital transactions has increased the risks of financial fraud, requiring automated solutions that are both effective and adaptable. Accordingly, this study aimed to compare the performance of three supervised learning algorithms, logistic regression, XGBoost, and artificial neural network, in detecting fraud in credit card transactions. A public and synthetic dataset was used, characterized by a strong class imbalance, thus simulating real-world scenarios. The process involved preprocessing temporal and categorical variables, applying class balancing techniques, and performing hyperparameter optimization. The models were evaluated based on metrics such as sensitivity and precision, and their behavior was analyzed from both technical and operational perspectives. The results indicate that logistic regression exhibited limited performance, XGBoost showed high sensitivity, and the neural network demonstrated flexibility through decision threshold calibration. It is concluded that, although there is no universally optimal model, XGBoost appears to be the most promising in environments that prioritize broad fraud detection, provided that mechanisms to control false positives are implemented. The comparative analysis enabled a structured understanding of the

performance and applicability of each algorithm, contributing to more informed decision-making in antifraud systems.

Keywords: *Machine Learning, Data Science, Banking Institutions, Fraud.*

1. Introdução

O crescimento acelerado do comércio eletrônico e dos sistemas de pagamento digitais transformou profundamente a dinâmica das transações financeiras, tornando-as mais ágeis, porém, também mais vulneráveis a fraudes sofisticadas. Rojan (2024) destaca que a digitalização dos meios de pagamento ampliou as oportunidades para os consumidores, mas também criou novas brechas exploradas por agentes mal-intencionados.

Wiese e Omlin (2007) reforçam que a complexidade e a interconectividade dos sistemas atuais exigem métodos de monitoramento mais inteligentes e dinâmicos. Nesse contexto, torna-se imperativa a adoção de soluções tecnológicas eficazes para a detecção de transações suspeitas, uma das principais causas de perdas financeiras globais.

Todavia, métodos tradicionais de verificação manual ou baseados em regras fixas mostram-se limitados diante do alto volume, da variabilidade e da velocidade das transações digitais. Técnicas de *Machine Learning*, em especial as baseadas em aprendizado supervisionado, vêm se destacando como ferramentas eficazes na detecção de fraudes por sua capacidade de identificar padrões não triviais, de adaptar-se a comportamentos dinâmicos e de operar com eficiência em tempo real (Tosta; Dias, 2025).

Um dos principais desafios técnicos enfrentados nesses modelos é o severo desbalanceamento das bases de dados, nas quais transações fraudulentas representam uma ínfima minoria. Nesses cenários, métricas tradicionais, como a acurácia, tornam-se enganosas. Awoyemi *et al.* (2017) ressaltam que métricas como sensibilidade e precisão são mais apropriadas para avaliar o desempenho do modelo, pois consideram o impacto real dos erros cometidos.

No que diz respeito aos modelos empregados pelo mercado, o *XGBoost* é um dos principais modelos para este fim e, segundo Chen e Guestrin (2016), trata-se de uma abordagem baseada em árvores de decisão que combina robustez estatística com eficiência computacional, sendo amplamente utilizada em problemas supervisionados com dados desbalanceados. Além disso, avanços recentes propõem modelos híbridos e técnicas complementares para aumentar a eficácia na detecção de fraudes.

Sulaiman *et al*. (2024) exploraram arquiteturas baseadas em redes neurais profundas que integram dados históricos e transacionais para antecipar comportamentos anômalos. Mienye e

Sun (2023) evidenciaram o impacto da seleção e balanceamento de atributos na *performance* dos modelos, enquanto Wiese e Omlin (2007) exploraram redes neurais recorrentes (LSTM) para capturar padrões temporais de comportamento suspeito.

Apesar da vasta literatura, poucos trabalhos realizam comparações estruturadas entre algoritmos sob condições experimentais padronizadas. Rojan (2024), por exemplo, utilizou uma base proprietária sem análise comparativa direta; Awoyemi *et al.* (2017) exploraram modelos em paralelo, mas com foco limitado no pré-processamento e na validação cruzada; e Carcillo *et al.* (2018) priorizaram a arquitetura computacional em tempo real, sem aprofundar-se nos algoritmos em si.

Diante disso, este trabalho propõe uma abordagem objetiva e didática para avaliar, de forma comparativa, três algoritmos amplamente utilizados na detecção de fraudes: regressão logística, redes neurais artificiais e *XGBoost*. Os modelos foram aplicados à mesma base pública sintética, com as mesmas métricas de avaliação, em um contexto de aprendizado supervisionado. A proposta visou analisar, de forma estruturada, o comportamento desses algoritmos diante do mesmo problema, discutindo suas vantagens, limitações, capacidade de generalização, interpretabilidade e aplicabilidade prática em sistemas antifraude reais.

2. Referencial Teórico

2.1 Ciência de dados

A utilização de sistemas computacionais robustos tem se tornado cada vez mais recorrente para a análise de grandes volumes de informações, estando diretamente associada à ciência de dados, cuja principal função consiste na análise de dados descritivos e desenvolvimento de modelos preditivos. Destaca-se, ainda, a integração de seus componentes, os quais, embora executem tarefas distintas, compõem um fluxo estruturado de etapas que incluem a aquisição, preparação e armazenamento dos dados, a engenharia de atributos, a modelagem, o treinamento e a avaliação de modelos de aprendizado de máquina, conforme destacado por Biswas *et al*. (2022).

Nesse contexto, os modelos fundamentados em ciência de dados podem ser organizados em diferentes processos, os quais compreendem etapas como a divisão e a preparação dos dados, a engenharia de atributos, a geração dos modelos e a sua avaliação, conforme discutido por He *et al*. (2021). Os autores destacam ainda, que a etapa de geração de modelos pode ser

subdividida em dois componentes principais: o espaço de busca, que se refere às características e configurações relevantes dos modelos considerados no projeto, englobando tanto algoritmos tradicionais de aprendizado de máquina quanto arquiteturas neurais; e os métodos de otimização, associados ao ajuste de hiperparâmetros, com ênfase nos procedimentos de treinamento e na otimização da arquitetura, visando à definição adequada dos parâmetros dos modelos como explanado também por Salehin *et al.* (2024).

2.2 *Machine learning*

Ao longo dos anos, a utilização de ferramentas computacionais tem se tornado essencial nos mais diversos setores, abrangendo áreas como a financeira, a construção civil e a indústria. Diante desse contexto, destacam-se os modelos baseados em aprendizado de máquina, sobretudo em razão do aumento de eficiência proporcionado por sua aplicação, por exemplo, na detecção de falhas ou mesmo de vulnerabilidades em sistemas, por meio da análise de código-fonte (Sharma *et al*., 2024).

Com a rápida expansão tecnológica impulsionada pelos avanços da inteligência artificial, a utilização de técnicas de aprendizado de máquina também se aprimorou. Nesse contexto, tais técnicas passaram a ser capazes de modelar séries temporais de forma mais eficiente, permitindo a generalização para dados não previamente observados, diferentemente das abordagens anteriores, que se baseavam predominantemente em dados estáticos (Hall; Rasheed, 2025).

Ressalta-se que inúmeros problemas de naturezas distintas, abrangendo desde contextos de pequeno e médio porte até aplicações de grande escala, têm sido solucionados por meio de técnicas de aprendizado de máquina. Tais aplicações demandam implementações contínuas, em função da complexidade e do volume das informações processadas, o que evidencia a relevância do uso do *machine learning* para a melhoria do desempenho preditivo e da confiabilidade dos modelos. Nesse sentido, essas técnicas tornam-se particularmente relevantes em sistemas de apoio à tomada de decisão, uma vez que possibilitam o aprimoramento dos *insights* gerados a partir da análise de grandes volumes de dados (Mumuni; Mumuni, 2025).

2.3 Modelos Empregados

Atualmente, diversos modelos baseados em aprendizado de máquina são empregados para os mais distintos propósitos, abrangendo aplicações como diagnósticos médicos, análise do comportamento de clientes e detecção de fraudes. Nesse contexto, evidencia-se a necessidade contínua de novos estudos que avaliem a eficiência de algoritmos baseados em *Gradient Boosting Machines* (GBM), bem como de suas principais variantes, como *XGBoost*, *Light Gradient Boosting Machine (LightGBM*) e *CatBoost*, de modo a possibilitar a identificação de abordagens mais eficazes e adequadas a diferentes cenários de aplicação (Florek; Zagdanski, 2023).

É importante destacar que cada um dos modelos, como *XGBoost*, *LightGBM* e *CatBoost*, entre outros, apresenta vantagens e limitações específicas, as quais dependem diretamente das características dos dados e de suas variações, como o comportamento dos clientes (Sudarshan; Seider, 2025). Considerando esses métodos, estudos que combinam o *XGBoost* com abordagens baseadas em *Random Forest* têm apresentado resultados promissores, superando modelos como Redes Neurais Artificiais, Árvores de Decisão, Máquinas de Vetores de Suporte e Regressão Logística. Ademais, observa-se que o *XGBoost* tem demonstrado desempenho superior quando comparado a outros algoritmos de *Boosting*, como *LightGBM* e *CatBoost*, indicando seu elevado potencial e robustez em aplicações preditivas, como mencionado por Imani *et al*. (2025).

3. Metodologia

3.1 Descrição da base de dados

A base de dados utilizada neste trabalho denomina-se *Credit Card Transactions Fraud Detection*, e está disponível publicamente na plataforma *Kaggle*. Trata-se de um conjunto de dados sintético, comumente adotado em pesquisas acadêmicas devido à sua complexidade estrutural e ao forte desbalanceamento entre classes. O conjunto foi fornecido já separado em treino e teste: o conjunto de treinamento contém 1.296.675 registros, dos quais 7.506 são fraudes (aproximadamente 0,58%), enquanto o conjunto de teste possui 555.719 registros, com 2.145 casos fraudulentos (aproximadamente 0,39%). A expressiva assimetria entre as classes demanda o uso de técnicas específicas para o tratamento do desbalanceamento e a escolha criteriosa de métricas de avaliação.

As variáveis que compõem a base dados são: *trans_date_trans_time* (data e hora da transação), *merchant* (nome do comerciante), *category* (categoria do comerciante), *amt* (valor da transação), *gender* (gênero do cliente), *city* (cidade do cliente), *state* (estado do cliente), *zip* (código postal), *lat* (latitude do cliente), *long* (longitude do cliente), *city_pop* (população da cidade do cliente), *job* (ocupação do cliente), *dob* (data de nascimento do cliente), *trans_num* (identificador único da transação), *unix_time* (timestamp UNIX da transação), *merch_lat* (latitude do comerciante), *merch_long* (longitude do comerciante), *is_fraud* (variável binária indicadora de fraude: 0 = não fraudulenta, 1 = fraudulenta).

3.2 Pré-processamento dos dados

O pré-processamento visou preparar os dados para os modelos preditivos por meio de transformações temporais e de codificação. As variáveis de data foram convertidas para o formato adequado, permitindo o cálculo da idade dos clientes, a extração da hora da transação e do dia da semana. Variáveis irrelevantes ou redundantes, como identificadores, dados pessoais e localizações específicas, foram removidas.

As variáveis categóricas *category*, *gender* e *dia_semana* foram convertidas para o tipo *category*, enquanto as variáveis numéricas foram identificadas automaticamente. Em seguida, utilizou-se um *ColumnTransformer* para aplicar *OneHotEncoder* às variáveis categóricas (excluindo a primeira categoria para evitar multicolinearidade) e *StandardScaler* às variáveis numéricas.

O transformador foi ajustado ao conjunto de treino e aplicado ao conjunto de teste, resultando em dados normalizados e codificados, prontos para os modelos de aprendizado de máquina.

3.3 Construção do Modelo de Regressão Logística

A regressão logística foi utilizada como modelo inicial de classificação binária, servindo como base comparativa para algoritmos mais complexos. O modelo foi treinado com base nos dados transformados, sem ajustes iniciais dos hiperparâmetros. Essa configuração padrão resultou em desempenho insatisfatório na detecção de fraudes: a classe minoritária apresentou

0% de sensibilidade e precisão de apenas 3% (Figura 1), evidenciando a limitação do modelo frente ao forte desbalanceamento da base.

Figura 1 – Indicadores da primeira regressão logística

	precision	recall	f1-score	support
0	1.00	1.00	1.00	553574
1	0.03	0.00	0.01	2145
accuracy			1.00	555719
macro avg	0.51	0.50	0.50	555719
weighted avg	0.99	1.00	0.99	555719

Fonte: Autores (2026)

Em seguida, foi realizado um processo de otimização por meio de *Grid Search*, com validação cruzada de três dobras e a métrica de sensibilidade como critério de avaliação. Testaram-se diferentes valores para o parâmetro de regularização C e para o *class_weight*, este último configurado tanto como *balanced* quanto com pesos definidos pela razão entre classes. A melhor configuração encontrada foi C=10 com *class_weight* sendo a razão entre as classes. O modelo otimizado apresentou sensibilidade de 74% e precisão de apenas 2% para a classe fraudulenta (Figura 2), demonstrando maior sensibilidade, embora com baixo poder discriminativo.

Figura 2 – Indicadores da primeira regressão logística otimizada

	precision	recall	f1-score	support
0	1.00	0.88	0.94	553574
1	0.02	0.74	0.05	2145
accuracy			0.88	555719
macro avg	0.51	0.81	0.49	555719
weighted avg	1.00	0.88	0.93	555719

Fonte: Autores (2026)

3.4 Construção do Modelo *XGBoost*

Inicialmente, foi construído um modelo *XGBoost* com parâmetros padrão, configurado com *eval_metric = aucpr, tree_method=hist.* e *enable_categorical=True*, visando otimizar a detecção da classe minoritária em um cenário altamente desbalanceado. O modelo foi ajustado sobre os dados transformados, e os resultados obtidos no conjunto de teste demonstraram forte desempenho, com precisão de 88% e sensibilidade de 74% para a classe fraudulenta (Figura 3), indicando bom equilíbrio entre sensibilidade e precisão mesmo sem ajustes finos.

Figura 3 – Indicadores do *XGboost* inicial

	precision	recall	f1-score	support
0	1.00	1.00	1.00	553574
1	0.88	0.74	0.80	2145
accuracy			1.00	555719
macro avg	0.94	0.87	0.90	555719
weighted avg	1.00	1.00	1.00	555719

Fonte: Autores (2026)

Na etapa seguinte, aplicou-se a técnica de *Grid Search*, com validação cruzada de três dobras, para otimizar os hiperparâmetros do modelo. Foram testadas diferentes combinações de profundidade das árvores, taxa de aprendizado, número de estimadores, frações de amostragem. A métrica de avaliação escolhida foi a sensibilidade, com foco na redução de falsos negativos.

A melhor configuração encontrada incluiu: *max_depth* = 7, *learning_rate* = 0,05, *n_estimators* = 100, *subsample* = 1,0, *colsample_bytree* = 1,0. Com esses hiperparâmetros, o modelo alcançou sensibilidade de 97% na detecção de fraudes, embora haja queda na precisão, que foi de apenas 19% (Figura 4), refletindo o aumento expressivo de falsos positivos (efeito esperado dada a priorização máxima da sensibilidade).

Figura 4 – Indicadores do *XGboost* com *Grid Search*

	precision	recall	f1-score	support
0	1.00	0.98	0.99	553574
1	0.19	0.97	0.32	2145
accuracy			0.98	555719
macro avg	0.60	0.97	0.66	555719
weighted avg	1.00	0.98	0.99	555719

Fonte: Autores (2026)

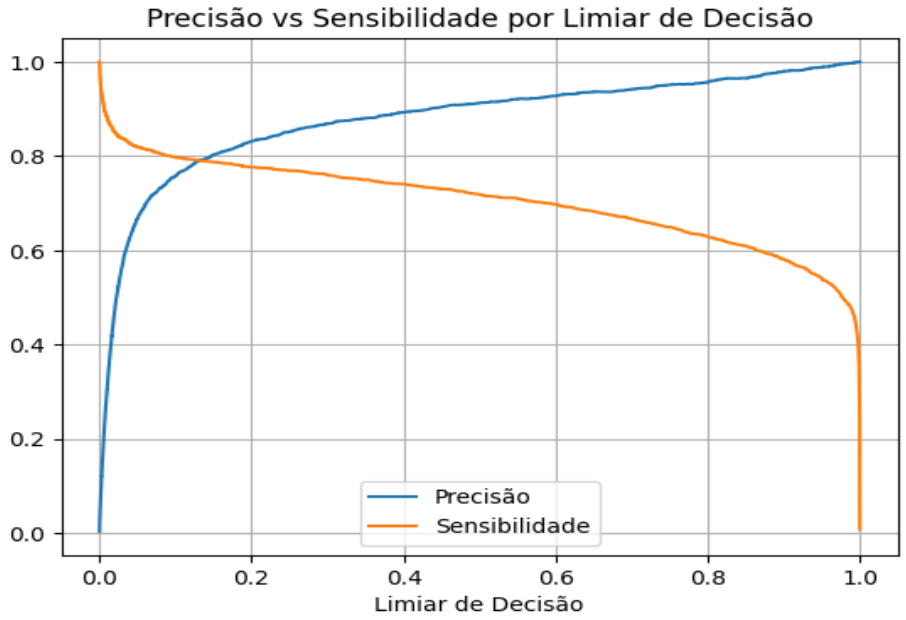
3.5 Construção do Modelo de Redes Neurais

Para avaliação de modelos baseados em arquitetura neural, foi implementada uma rede neural densa (*feedforward*) utilizando a biblioteca *TensorFlow/Keras*. Inicialmente, os dados foram transformados por meio do mesmo *pipeline* de pré-processamento aplicado aos demais modelos, com codificação categórica via *OneHotEncoder* e padronização numérica por *StandardScaler*.

O modelo foi definido com a seguinte arquitetura sequencial: uma camada de entrada compatível com 28 variáveis preditoras, seguida por duas camadas ocultas densas com 32 e 16 neurônios, ambas com função de ativação *ReLU*. A camada de saída possuía um único neurônio com ativação sigmoide, adequada para problemas de classificação binária. O otimizador adotado foi o *Adam*, com função de perda *binary_crossentropy* e métrica principal de desempenho a sensibilidade, dado o foco na detecção da classe minoritária (fraudes).

O modelo foi treinado com *batch_size* = 32, *epochs* = 10 e *validation_split* = 0,2. Após o treinamento inicial, os valores preditos foram analisados com base em diferentes limiares de decisão, o que permitiu a visualização da curva Precisão *vs* Sensibilidade (Figura 5). Essa curva foi construída para avaliar o *trade-off* entre as métricas em função do limiar de decisão adotado. A partir dessa análise, definiu-se um limiar de 0,13 para a classificação como fraude, em vez do tradicional 0,5, priorizando equilíbrio entre sensibilidade e precisão do modelo. Com esse limiar ajustado, o modelo atingiu sensibilidade de 79% e precisão de 79% para a classe fraudulenta.

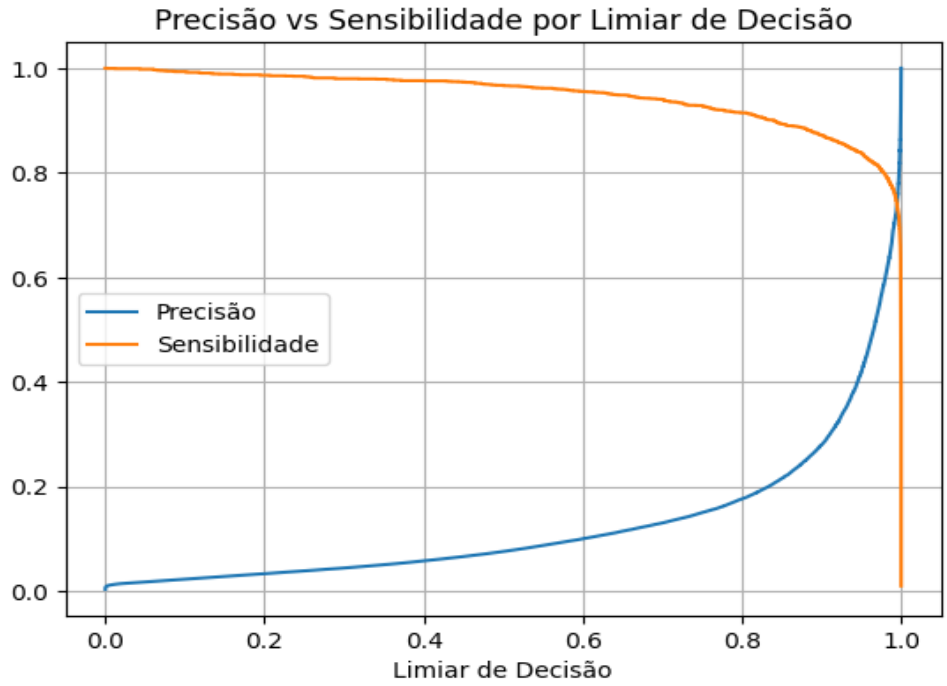
Figura 5 – Curva precisão vs sensibilidade da rede neural padrão



Fonte: Autores (2026)

Em seguida, buscou-se incorporar ao modelo o desbalanceamento entre as classes por meio da técnica de *class_weight*, utilizando a razão empírica entre as classes (aproximadamente 172:1). O treinamento foi mantido com os mesmos parâmetros anteriores, e os resultados foram novamente analisados em função de diferentes limiares de decisão. Após nova análise da curva precisão *vs* sensibilidade (Figura 6), foi definido um limiar ajustado de 0,6, obtendo como resultados uma sensibilidade de 96% e precisão de 10% para a classe fraudulenta.

Figura 6 – Curva precisão *vs* sensibilidade da rede neural balanceada



Fonte: Autores (2026)

4. Resultados e Discussão

A comparação entre Regressão Logística, *XGBoost* e Redes Neurais para detecção de fraudes evidenciou tanto as diferenças estruturais entre os modelos quanto os impactos gerenciais decorrentes de suas decisões. Mais do que uma análise descritiva das métricas, os resultados apontam para dilemas clássicos enfrentados por sistemas antifraude, como o equilíbrio entre sensibilidade e precisão e a estabilidade operacional. A seguir, são discutidos os achados para cada um dos modelos.

4.1 Regressão Logística

O modelo de Regressão Logística, utilizado como *baseline*, apresentou desempenho bastante limitado em sua configuração padrão, com sensibilidade inferior a 1% para a classe fraudulenta. Mesmo após a otimização dos hiperparâmetros e o ajuste dos pesos de classe com base na razão entre as classes, a sensibilidade subiu para 74%, enquanto a precisão permaneceu extremamente baixa (2%). Isso significa que, embora tenha passado a capturar mais fraudes, a enorme maioria dos alertas gerados continuou sendo falsos-positivos, o que inviabiliza sua adoção prática em um ambiente real de monitoramento. Esse resultado está em linha com os

achados de Awoyemi *et al*. (2017), que apontam a limitada capacidade da regressão logística de lidar com distribuições altamente desbalanceadas.

4.2 *XGBoost*

O modelo *XGBoost* apresentou desempenho técnico robusto desde sua configuração inicial. Mesmo sem ajustes de hiperparâmetros, obteve 74% de sensibilidade e 88% de precisão, o que é um desempenho notável em bases altamente desbalanceadas, evidenciando a capacidade do algoritmo de capturar padrões complexos de comportamento fraudulento.

Com a otimização por meio de validação cruzada e do ajuste do peso da classe minoritária, a sensibilidade subiu para 97%, indicando cobertura quase total das transações fraudulentas. No entanto, esse ganho veio acompanhado de uma queda na precisão para 19%, elevando consideravelmente o número de transações legítimas classificadas como suspeitas.

Esse comportamento é coerente com os achados de He e Garcia (2009), que alertam para o risco de ruído excessivo em modelos excessivamente sensíveis, o que compromete sua viabilidade operacional. Géron (2021), por sua vez, reforça a importância do ajuste fino dos hiperparâmetros para mitigar esse efeito.

Sob a ótica gerencial, a adoção do XGBoost demanda camadas adicionais de filtragem, como regras de negócio ou modelos auxiliares, para evitar impactos negativos sobre usuários legítimos. Esse cuidado é compatível com a proposta de Charizanos, Demirhan e İçen (2024), que recomendam a construção de sistemas antifraude adaptativos, capazes de considerar o contexto transacional e de reduzir intervenções indevidas.

4.3 Redes Neurais

As redes neurais apresentaram desempenho intermediário, com destaque para sua flexibilidade em contextos em que a calibração do ponto de corte da classificação é estratégica. Em modelos de classificação binária, como o de detecção de fraudes, é comum que a probabilidade prevista pelo modelo seja convertida em decisão final com base em um limiar de decisão, i.e., um valor acima do qual uma transação é considerada suspeita. Ajustar esse limiar permite controlar o nível de rigor do modelo, tornando-o mais ou menos propenso a emitir alertas.

Neste estudo, foram desenvolvidos dois modelos de redes neurais com a mesma arquitetura (duas camadas densas). No primeiro, utilizou-se o pipeline de pré-processamento padrão, sem ajuste de pesos entre as classes, e adotou-se um limiar de decisão ajustado para

0,13 com o objetivo de buscar equilíbrio entre sensibilidade e precisão. Essa configuração resultou em sensibilidade e precisão de 79%, indicando bom desempenho geral na detecção de fraudes.

Já no segundo modelo, foi introduzido o parâmetro *class_weight*, atribuindo maior peso à classe fraudulenta para mitigar o impacto do desbalanceamento. A estratégia visou maximizar a sensibilidade do modelo, adotando um limiar de decisão mais elevado (0,7). Com essa configuração, o modelo alcançou sensibilidade de 94%, mas com precisão reduzida a 13%, o que reflete uma priorização clara pela detecção de fraudes, mesmo ao custo de um maior número de falsos positivos. Embora o alto índice de detecção seja desejável, o número de falsos positivos foi elevado. Por outro lado, a análise da curva de sensibilidade *versus* precisão revelou que o modelo responde bem à calibração do limiar de decisão, permitindo ajustes operacionais conforme o perfil de risco ou o contexto da transação.

Essa capacidade de calibragem é ressaltada por Sulaiman *et al.* (2024) como um dos principais trunfos das redes neurais em cenários de fraude. Em ambientes bancários, por exemplo, é possível adotar limiares mais rigorosos para clientes de alto risco ou mais flexíveis para clientes com bom histórico, evitando bloqueios indevidos.

Além disso, como argumentam Wiese e Omlin (2017), as redes neurais possuem maior capacidade de capturar padrões não lineares e múltiplas interações entre variáveis, o que as torna especialmente úteis em situações em que os comportamentos fraudulentos evoluem ao longo do tempo. Embora ainda apresentem desafios quanto à interpretabilidade, fator crítico em ambientes regulatórios, sua capacidade de adaptação e personalização justifica seu uso em sistemas híbridos, combinados a regras de negócio ou a modelos mais transparentes.

4.4 Reflexões sobre aplicabilidade prática

De modo geral, todos os modelos analisados apresentaram limitações inerentes ao desafio de detectar fraudes em bases de dados reais: a classe minoritária é extremamente rara, os padrões de fraude são dinâmicos e o custo dos erros varia conforme o tipo de erro cometido. Nesse cenário, não existe um modelo ideal isolado. A adoção prática de qualquer solução exigirá calibragem cuidadosa, integração com regras de negócio e monitoramento contínuo.

Embora o estudo de limiares de decisão também tenha sido realizado para os modelos *XGBoost* e Regressão Logística, optou-se por utilizar o limiar padrão (0,5) na avaliação final de desempenho, com o objetivo de preservar a comparabilidade entre as abordagens. Essa decisão também levou em consideração a maior sensibilidade do *XGBoost* e a baixa precisão

da regressão, cujos ajustes de limiar, apesar de testados, não trouxeram benefícios operacionais significativos. No caso da Rede Neural, por outro lado, o ajuste do limiar mostrou-se mais efetivo, destacando-se como um diferencial de flexibilidade.

O *XGBoost* se mostrou o mais promissor tecnicamente, com alta sensibilidade mesmo em sua configuração básica. No entanto, o modelo exigirá controle de falsos positivos antes da implementação em produção, de modo a evitar impactos negativos sobre clientes legítimos. A Regressão Logística, apesar de sua transparência e facilidade de interpretação, não apresentou desempenho mínimo viável para triagem eficaz. Já a Rede Neural destacou-se pela adaptabilidade via ajuste do limiar de decisão, embora sua interpretabilidade limitada ainda represente um obstáculo em contextos regulatórios.

Esses achados estão em consonância com Carcillo *et al*. (2021), que ressaltam que sistemas antifraude devem ir além da *performance* algorítmica e considerar fatores operacionais, como capacidade de explicação, volume de alarmes falsos e impacto sobre a experiência do usuário final.

5. Considerações Finais

Este trabalho buscou desenvolver e comparar modelos de aprendizado de máquina aplicados à detecção de fraudes em transações financeiras, considerando tanto a capacidade técnica de identificar comportamentos suspeitos quanto a viabilidade de adoção em contextos operacionais reais. O estudo demonstrou que diferentes algoritmos oferecem vantagens e limitações distintas, e que a escolha da solução mais adequada depende diretamente dos objetivos e restrições do ambiente em que será aplicada.

Ficou evidente que há um equilíbrio delicado entre a detecção efetiva de fraudes e o controle de falsos positivos. Modelos mais sensíveis tendem a capturar um número maior de fraudes, mas ao custo de impactar negativamente usuários legítimos. Por outro lado, modelos mais conservadores geram menos intervenções indevidas, mas deixam de identificar parte das fraudes. Essa tensão entre cobertura e precisão exige decisões estratégicas que envolvem mais do que métricas: é necessário considerar os riscos, os recursos disponíveis e a tolerância da instituição a erros de cada tipo.

A hipótese de que métodos mais sofisticados de aprendizado de máquina podem superar abordagens tradicionais em cenários complexos foi confirmada, mas também se constatou que a complexidade dos modelos exige atenção redobrada quanto à calibragem e à capacidade de

interpretação. Soluções eficazes não são necessariamente as mais complexas, mas aquelas que se integram de forma eficiente à rotina de tomada de decisão.

Entre as limitações deste estudo, destaca-se o uso de um conjunto de dados sintético, que embora útil para experimentação, pode não capturar integralmente a complexidade e a imprevisibilidade de padrões de fraude em dados reais. Outro ponto é que o impacto financeiro real de falsos positivos e falsos negativos não foi mensurado, o que seria relevante para decisões práticas. Por fim, não foram exploradas estratégias de integração em sistemas de monitoramento em tempo real, fator essencial para aplicações operacionais.

Para trabalhos futuros, recomenda-se o uso de bases de dados reais provenientes de instituições financeiras, permitindo avaliar o desempenho dos modelos em contextos mais próximos da prática. Sugere-se também investigar o uso de variáveis adicionais, como dados comportamentais e históricos de clientes, e explorar modelos híbridos que combinem algoritmos de *machine learning* com regras de negócio.

REFERÊNCIAS

AWOYEMI, John O.; ADETUNMBI, Adebayo O.; OLUWADARE, Samuel A. **Credit card fraud detection using machine learning techniques: A comparative analysis**. In: 2017 international conference on computing networking and informatics (ICCNI). IEEE, 2017. p. 1-9.

BISWAS, S.; WARDAT, M.; RAJAN, H. **The art and practice of data science pipelines: A comprehensive study of data science pipelines in theory, in-the-small, and in-the-large**. In: Proceedings of the 44th International Conference on Software Engineering. [S.l.: s.n.], 2022. p. 2091–2103.

CARCILLO, F.; DAL POZZO, A.; LE BORGNE, Y.-A.; BONTEMPS, L.; CAUWENBERGHS, E.; BONTEMPS, C. **Scarff: a scalable framework for streaming credit card fraud detection with spark**. Information fusion, v. 41, p. 182-194, 2018.

CHARIZANOS, Georgios; DEMIRHAN, Haydar; İÇEN, Duygu. **An online fuzzy fraud detection framework for credit card transactions**. Expert Systems with Applications, v. 252, p. 124127, 2024.

CHEN, Tianqi. **XGBoost: A Scalable Tree Boosting System**. Cornell University, 2016.

FLOREK, P.; ZAGDA ´NSKI, A. **Benchmarking state-of-the-art gradient boosting algorithms for classification**. arXiv preprint arXiv:2305.17094, 2023.

GÉRON, Aurélien. **Mãos à obra: aprendizado de máquina com Scikit-Learn. Keras & TensorFlow: Conceitos, ferramentas e técnicas para a construção de sistemas inteligentes**.[S. l.: sn], p. 5, 2021.

HALL, T.; RASHEED, K. **A survey of machine learning methods for time series prediction**. Applied Sciences, v. 15, n. 11, p. 5957, 2025.

HE, Haibo; GARCIA, Edwardo A. **Learning from imbalanced data**. IEEE Transactions on knowledge and data engineering, v. 21, n. 9, p. 1263-1284, 2009.

HE, X.; ZHAO, K.; CHU, X. Automl: **A survey of the state-of-the-art**. Knowledge-based systems, Elsevier, v. 212, p. 106622, 2021.

IMANI, M.; BEIKMOHAMMADI, A.; ARABNIA, H. R. **Comprehensive analysis of random forest and xgboost performance with smote, adasyn, and gnus under varying imbalance levels**. Technologies, v. 13, n. 3, p. 88, 2025.

MIENYE, Ibomoiye Domor; SUN, Yanxia. **A machine learning method with hybrid feature selection for improved credit card fraud detection**. Applied Sciences, v. 13, n. 12, p. 7254, 2023.

MUMUNI, A.; MUMUNI, F. **Automated data processing and feature engineering for deep learning and big data applications: a survey**. Journal of Information and Intelligence, v. 3, n. 2, p. 113–153, 2025.

ROJAN, Zaki. **Financial fraud detection based on machine and deep learning: A review**. The Indonesian Journal of Computer Science, v. 13, n. 3, 2024.

SALEHIN, I.; ISLAM, M. S.; SAHA, P.; NOMAN, S.; TUNI, A.; HASAN, M. M.; BATEN, M. A. **Automl: A systematic review on automated machine learning with neural architecture search**. Journal of Information and Intelligence, v. 2, n. 1, p. 52–81, 2024.

SHARMA, T.; KECHAGIA, M.; GEORGIOU, S.; TIWARI, R.; VATS, I.; MOAZEN, H.; SARRO, F. **A survey on machine learning techniques applied to source code**. Journal of Systems and Software, v. 209, p. 111934, 2024.

SUDARSHAN, V.; SEIDER, W. D. **Advancing machine learning in industry 4.0: Benchmark framework for rare-event prediction in chemical processes**. Computers & Chemical Engineering, v. 194, p. 108929, 2025.

SULAIMAN, R. B. et al. **Enhancing security in digital payments: a comparative evaluation of machine learning models for credit card fraud detection**. Frontiers in Machine Learning and Big Data, v. 3, n. 2, p. 45-62, 2024.

TOSTA, Pedro Lucas Maranini; DIAS, Jonatas Cerqueira. **Detecção de fraudes em transações bancárias utilizando inteligência artificial**. Revista Processando o Saber, v. 17, p. 21-37, 2025.

WIESE, Bénard; OMLIN, Christian. **Credit card transactions, fraud detection, and machine learning: Modelling time with LSTM recurrent neural networks**. In: Innovations in neural information paradigms and applications. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 231-268.