APLICAÇÕES DE INTELIGÊNCIA ARTIFICIAL NA PREVENÇÃO DE FRAUDES BANCÁRIAS: UMA ABORDAGEM BIBLIOGRÁFICA

Tiago Gregorutti de Carvalho[1]

Jaqueline Brigladori Pugliesi[2]

Resumo

Este artigo analisa o uso de algoritmos de Inteligência Artificial (IA) na detecção e prevenção de fraudes bancárias, focando em técnicas supervisionadas, *deep learning* e IA generativa. A motivação está no aumento das transações digitais e na sofisticação dos ataques, que exigem soluções mais precisas e adaptativas para identificar fraudes em tempo real com baixo índice de falsos positivos. A pesquisa qualitativa, bibliográfica e documental busca identificar os métodos mais eficazes e os desafios técnicos e éticos envolvidos. A literatura destaca algoritmos como *Random Forest*, SVM e LSTM por seu desempenho na detecção de padrões anômalos, especialmente quando aplicados com técnicas de balanceamento de dados. Além disso, IA generativa, como ChatGPT e LLaMA, mostra potencial para monitores antifraude baseados em linguagem natural e geolocalização. Contudo, questões de interpretabilidade, decisões automatizadas equivocadas e necessidade de validação contínua ainda limitam sua adoção. O estudo conclui que a combinação de diferentes abordagens de IA com dados contextuais, como biometria e comportamento do usuário, constitui a estratégia mais promissora para soluções antifraude eficientes e transparentes.

Palavras-chave: Biometria. Detecção de Fraudes. Geolocalização. Inteligência Artificial. Machine Learning. Segurança Bancária.

Abstract

This article analyzes the use of Artificial Intelligence (AI) algorithms in detecting and preventing banking fraud, focusing on supervised techniques, deep learning, and generative AI. The motivation lies in the increase in digital transactions and the sophistication of attacks, which demand more precise and adaptive solutions to identify fraud in real time with low false positive rates. This qualitative, bibliographic, and documentary research seeks to identify the most effective methods and the technical and ethical challenges involved. The literature highlights algorithms such as Random Forest, SVM, and LSTM for their performance in anomaly detection, especially when applied with data balancing techniques. Furthermore, generative AI, such as ChatGPT and LLaMA, shows potential for fraud monitoring systems based on natural language and geolocation. However, issues such as interpretability, erroneous automated decisions, and the need for continuous validation still limit its widespread adoption. The study concludes that combining different AI approaches with contextual data, such as

[1] Graduando em Análise e Desenvolvimento de Sistemas pela Fatec Dr Thomaz Novelino – Franca/SP. Endereço eletrônico: tiagogregc@gmail.com.

[2] Doutora em Ciências da Computação pela USP – São Carlos/SP. Endereço eletrônico: jbpugliesi@gmail.com.

biometrics and user behavior constitutes the most promising strategy for efficient and transparent fraud prevention solutions.

Keywords: *Artificial Intelligence. Banking Security. Biometrics. Fraud Detection. Geolocation. Machine Learning.*

1 Introdução

O avanço das tecnologias digitais e o crescimento das transações bancárias eletrônicas transformaram significativamente o cenário financeiro mundial. Junto com os benefícios trazidos pela digitalização, surgiram também novos desafios relacionados à segurança da informação, com destaque para as fraudes bancárias. Esse tipo de crime, caracterizado pela apropriação indevida de valores por meio de artifícios como engenharia social, *phishing* ou manipulação de sistemas, representa uma ameaça constante para instituições financeiras e consumidores.

Nesse contexto, a aplicação da Inteligência Artificial (IA) tem se mostrado uma alternativa promissora para fortalecer os mecanismos de prevenção e detecção de fraudes. Com a capacidade de identificar padrões, analisar grandes volumes de dados em tempo real e adaptar-se a novas estratégias de ataque, a IA tem sido progressivamente incorporada aos sistemas de segurança bancária. Algoritmos supervisionados, como *Random Forest* e Regressão Logística, técnicas de *Deep Learning*, como Redes Neurais LSTM (*Long Short Term Memory*) e CNN (*Convolutional Neural Network*), e abordagens mais recentes, como a Inteligência Artificial Generativa, compõem o arsenal tecnológico atualmente em uso ou em desenvolvimento.

Ao mesmo tempo, surgem novas questões éticas, operacionais e jurídicas que desafiam a implementação dessas soluções, como a explicabilidade dos modelos, o risco de decisões automatizadas equivocadas e a proteção dos dados pessoais. O uso de IA em ambientes altamente regulados, como o financeiro, exige não apenas desempenho técnico, mas também conformidade normativa, responsabilidade e transparência.

Diante disso, este artigo tem como objetivo analisar o papel da IA na prevenção de fraudes bancárias, por meio de uma abordagem bibliográfica e qualitativa, mapeando os principais algoritmos utilizados, os contextos de aplicação e os desafios enfrentados. Busca-se, assim, contribuir para o entendimento das potencialidades e

limitações da IA nesse campo estratégico, apontando caminhos para o desenvolvimento de soluções mais seguras, eficientes e confiáveis.

2 Referencial teórico

2.1 Inteligência Artificial e *Machine Learning*

A compreensão dos mecanismos de prevenção a fraudes bancárias por meio de Inteligência Artificial exige o domínio de conceitos fundamentais que sustentam o desenvolvimento e a aplicação desses sistemas. Neste contexto, destaca-se inicialmente a própria definição de IA e suas vertentes mais recentes, como a Inteligência Artificial Generativa, bem como os princípios do Aprendizado de Máquina que possibilitam a modelagem de padrões suspeitos em ambientes financeiros.

A Inteligência Artificial refere-se à capacidade de sistemas computacionais de simular aspectos do raciocínio e comportamento humano, com vistas à resolução de problemas e tomada de decisão. Segundo Sichman (2021), não há uma definição universalmente aceita, mas, em termos práticos, a IA pode ser compreendida como um campo multidisciplinar que integra lógica, estatística, ciência da computação e cognição humana para desenvolver sistemas capazes de perceber, aprender, raciocinar e agir. Faceli *et al.* (2021) complementa essa visão ao afirmar que a IA se distingue por sua capacidade de aprender a partir de dados, moldando-se constantemente para aprimorar suas respostas e decisões, o que a torna especialmente útil em contextos dinâmicos como a prevenção de fraudes bancárias.

Desde suas origens, a IA foi estruturada em abordagens distintas quanto à forma como sistemas inteligentes devem se comportar ou pensar. Segundo Faceli *et al.* (2021), essas abordagens podem ser classificadas em quatro categorias principais: sistemas que pensam como seres humanos, sistemas que agem como seres humanos, sistemas que pensam racionalmente e sistemas que agem racionalmente. Essas diretrizes nortearam o desenvolvimento dos primeiros algoritmos voltados à tomada de decisão automatizada, bem como das ferramentas de classificação e predição amplamente utilizadas no contexto bancário.

No campo da detecção de fraudes, a IA é frequentemente associada ao Aprendizado de Máquina (*Machine Learning*), especialmente em sua vertente supervisionada, que visa construir modelos capazes de relacionar entradas – como dados transacionais, comportamentais e geográficos – com saídas esperadas, como

a classificação de uma transação como legítima ou fraudulenta. Segundo Faceli *et al*. (2021), os algoritmos supervisionados aprendem padrões a partir de dados históricos rotulados e, por isso, são amplamente utilizados em contextos de classificação binária, como no caso de sistemas antifraude bancários. Essa abordagem permite que os modelos melhorem progressivamente sua acurácia à medida que novos dados são incorporados ao processo de treinamento.

Uma evolução recente nesse campo é a Inteligência Artificial Generativa (GenIA), que se baseia na combinação de IA tradicional com modelos de linguagem (LLMs – *Large Language Models*), os quais são capazes de gerar textos, padrões e respostas contextualizadas. Segundo Carle (2023), essa tecnologia tem o poder de criar novas informações a partir de conceitos já ensinados, utilizando o raciocínio da IA e modelos de linguagem para soar natural em seus resultados. Sua aplicação na área de segurança bancária é promissora, como apresentado no uso de GenIA para monitorar padrões comportamentais e geográficos de clientes, identificando anomalias em tempo real e acionando bloqueios preventivos automaticamente (Venturini et al., 2024).

Essas capacidades estão na base da transformação das soluções antifraude: os sistemas modernos não apenas analisam transações, mas também aprendem padrões, detectam anomalias, adaptam-se a novos cenários e comunicam-se com operadores humanos de forma contextualizada. Dessa forma, a IA não é mais apenas um mecanismo de suporte, mas um agente autônomo de decisão e prevenção de riscos, especialmente quando integrada a bases de dados ricas e diversas fontes comportamentais.

2.2 Técnicas de detecção de fraudes bancárias

A fraude bancária pode ser compreendida sob duas óticas complementares: jurídica e técnica. No campo jurídico, o artigo 171 do Código Penal Brasileiro tipifica o estelionato como crime, caracterizando-o como obtenção de vantagem ilícita em prejuízo alheio mediante artifício ou ardil (Brasil, 1940). Já sob o ponto de vista técnico, fraudes envolvem acessos não autorizados, manipulações de transações e uso indevido de credenciais em sistemas financeiros.

De acordo com Vilela, Ueda e Gava (2023), os principais vetores de ataque são:

- Engenharia social: que explora o comportamento humano para obter informações confidenciais.
- *Phishing*: com envio de mensagens falsas para capturar dados bancários.
- Vazamento de dados: resultante de falhas de segurança ou ataques cibernéticos.

Devido à sofisticação crescente desses ataques, a identificação em tempo real tornou-se essencial. Técnicas baseadas em IA têm se mostrado eficazes, principalmente pela capacidade de identificar padrões anômalos em grandes volumes de dados. Segundo Lima (2022), algoritmos de aprendizado de máquina aplicados a dados bancários são capazes de bloquear automaticamente transações suspeitas com alta precisão. Além disso, sistemas baseados em comportamento e geolocalização permitem detectar desvios de conduta digital, como horários atípicos de acesso ou transações fora do padrão habitual (Souza; Bordin Jr., 2023).

3 Tecnologias e ferramentas aplicáveis

A detecção de fraudes bancárias exige técnicas capazes de lidar com grandes volumes de dados, identificar padrões sutis de anomalia e responder em tempo real. Diversos algoritmos tradicionais e modelos recentes de Inteligência Artificial têm sido utilizados com esse propósito, cada um com suas vantagens e limitações.

3.1 Algoritmos tradicionais de Inteligência Artificial

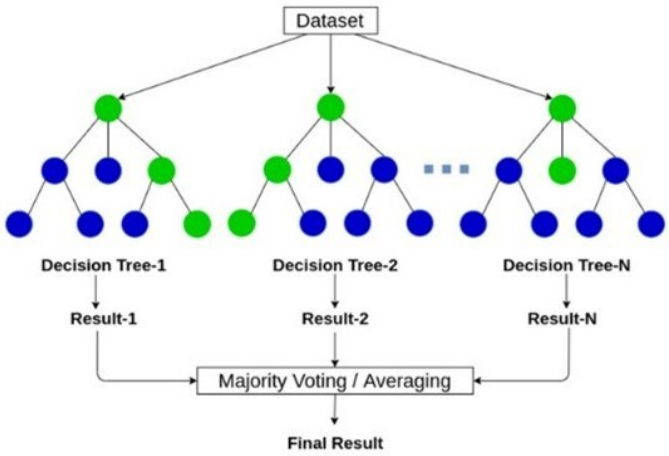
Os algoritmos tradicionais de Inteligência Artificial representam a base dos sistemas automatizados de detecção de fraudes, sendo amplamente utilizados por sua eficiência, simplicidade e robustez. Técnicas como *Random Forest*, Regressão Logística e SVM são reconhecidas pela capacidade de lidar com grandes volumes de dados e fornecer classificações precisas, mesmo em contextos desbalanceados. A seguir, são descritas as principais características desses métodos e suas aplicações no setor bancário.

Random Forest é um método de aprendizado supervisionado baseado em conjuntos de árvores de decisão (*ensemble learning*). Ele opera construindo múltiplas árvores durante o treinamento e retorna a classe que é a moda das previsões de cada árvore individual. Essa abordagem é particularmente eficaz em contextos de fraude

devido à sua robustez a dados ruidosos, capacidade de lidar com dados desbalanceados e baixo risco de *overfitting* (ajuste excessivo do modelo ao treinamento). De acordo com Bhattacharyya et al. (2011), *Random Forest* apresentou resultados superiores a outros classificadores tradicionais em bases de dados bancárias com alta desproporção entre transações legítimas e fraudulentas.

Conforme representado na Figura 1, o algoritmo *Random Forest* opera como um conjunto de árvores de decisão, onde cada árvore fornece uma previsão e o modelo final retorna a classe com maior número de votos. Essa abordagem reduz o risco de *overfitting* e melhora a acurácia em problemas de classificação binária, como a detecção de fraudes.

Figura 1 – Funcionamento do Algoritmo Random Forest

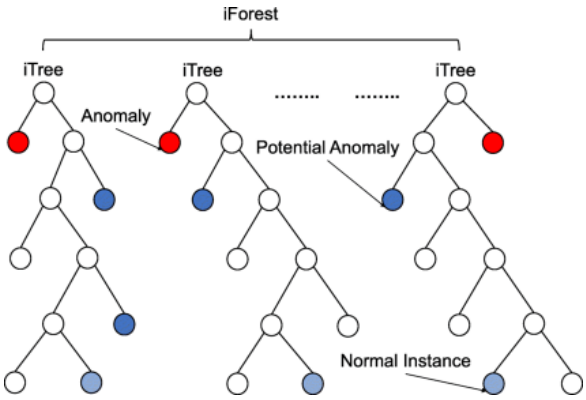


Fonte: Adaptado de Zhang; Shao; Li, 2017, p. 179.

Isolation Forest é um algoritmo não supervisionado voltado especificamente à detecção de anomalias. Em vez de aprender padrões normais de comportamento, ela busca isolar pontos que se distanciam da maioria dos dados. O princípio central é que anomalias são mais fáceis de isolar em poucas divisões. Por isso, elas tendem a ter caminhos mais curtos nas árvores construídas pelo algoritmo. Conforme Liu, Ting e Zhou (2008), a técnica é eficiente em contextos com poucas ocorrências de fraude, pois não exige rótulos para treinar o modelo.

O *Isolation Forest*, conforme apresentado na Figura 2, detecta anomalias isolando dados com base em particionamentos aleatórios. As instâncias anômalas tendem a ser isoladas rapidamente, gerando caminhos curtos nas árvores. Essa técnica é eficiente para grandes volumes de dados e não requer rótulos pré-definidos.

Figura 2 – Estrutura de Árvores no *Isolation Forest*

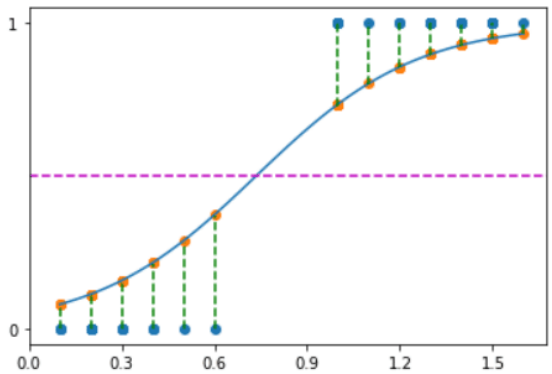


Fonte: Adaptado Liu; Ting; Zhou, 2008, p. 2.

Regressão Logística associada ao SVM (*Support Vector Machine*) une a simplicidade interpretativa da regressão com o poder de generalização do SVM, que busca maximizar os limites de separação entre as classes. Em bases com muitos atributos e dados desbalanceados, como em fraudes bancárias, o SVM pode ser ajustado com *kernels* não-lineares para capturar padrões complexos. Segundo Dal Pozzolo et al. (2014), a utilização conjunta de SVM com técnicas de balanceamento, como o SMOTE (*Synthetic Minority Over-sampling Technique*), melhora significativamente a detecção de fraudes raras, com bons índices de precisão e recall.

A regressão logística modela a probabilidade de ocorrência de um evento binário. Na Figura 3, observa-se que a curva sigmoide transforma a saída linear em uma escala de 0 a 1, permitindo interpretar a saída como uma probabilidade de fraude. É simples, interpretável e amplamente usada como *baseline* em sistemas antifraude.

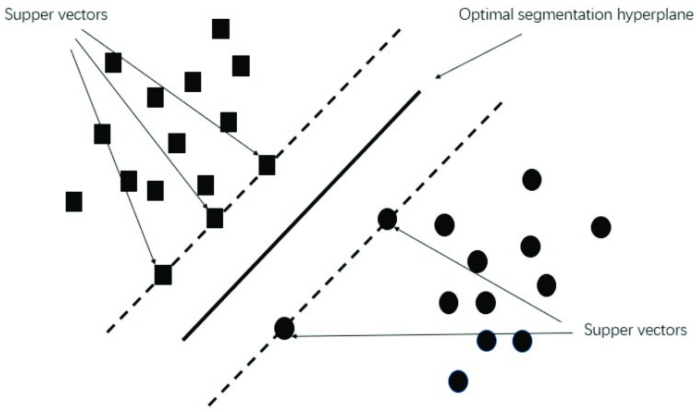
Figura 3 – Curva Sigmoide da Regressão Logística



Fonte: Adaptado de Hosmer; Lemeshow; Sturdivant, 2013, p. 78.

O *Support Vector Machine*, conforme apresentado na Figura 4, identifica o hiperplano ótimo que separa as classes com a maior margem possível. É eficaz em contextos com alta dimensionalidade e é especialmente útil quando há uma clara separação entre dados legítimos e fraudulentos.

Figura 4 – Separação de Classes com SVM.



Fonte: Adaptado de Cortes e Vapnik, 1995, p.275.

Em que pese a eficácia dos algoritmos tradicionais na identificação de padrões anômalos, o aumento da complexidade e da variabilidade dos ataques tem exigido abordagens mais sofisticadas e capazes de extrair relações não lineares em grandes volumes de dados. Nesse sentido, as técnicas de *Deep Learning* surgem como uma evolução natural, oferecendo modelos mais profundos e adaptativos para lidar com o comportamento dinâmico dos usuários e com transações cada vez mais sutis e complexas.

3.2 *Deep Learning*

O *Deep Learning* (Aprendizado Profundo) representa uma evolução das técnicas tradicionais de aprendizado de máquina, estruturando Redes Neurais Artificiais com múltiplas camadas ocultas capazes de extrair padrões complexos e não lineares de grandes volumes de dados. Essa capacidade o torna altamente eficaz na detecção de fraudes bancárias, onde desvios comportamentais nem sempre seguem padrões lineares ou previamente rotulados.

Segundo Mondragón e Yarango (2025), as redes neurais profundas – especialmente as do tipo LSTM (*Long Short-Term Memory*) – têm se destacado por

sua capacidade de identificar padrões temporais em séries de transações, o que permite detectar alterações sutis no comportamento de clientes que possam indicar tentativas de fraude. As LSTM são particularmente úteis em cenários onde o histórico de eventos influencia diretamente a confiabilidade de uma nova transação.

As Redes Neurais Convolucionais (CNNs), embora mais comumente associadas ao reconhecimento de imagens, têm sido adaptadas com sucesso para detecção de fraudes por sua habilidade em capturar padrões locais de forma automática (Mondragón; Yarango, 2025).

Além disso, técnicas como *Autoencoders* têm sido aplicadas com foco na reconstrução de transações típicas, permitindo identificar desvios a partir da diferença entre a transação real e sua reconstrução. Já as Redes Adversariais Generativas (GANs) são utilizadas para a geração sintética de fraudes, o que auxilia na equalização de bases de dados desbalanceadas, conforme destacado por Mata (2023).

Apesar dos benefícios, o uso de *Deep Learning* traz desafios. Sichman (2021) alerta que, devido à complexidade estrutural das redes, esses modelos frequentemente funcionam como “caixas-pretas”, dificultando a explicação dos resultados – aspecto sensível em ambientes com alta regulação, como é o setor financeiro. Além disso, Mondragón e Yarango (2025) ressaltam que a necessidade de grandes volumes de dados e recursos computacionais robustos ainda é uma barreira para a aplicação massiva dessas técnicas em instituições de menor porte.

Em síntese, o *Deep Learning* oferece soluções promissoras para a detecção de fraudes bancárias em tempo real, sobretudo quando integrado a bases de dados transacionais contínuas e a técnicas de monitoramento comportamental. Contudo, seu uso requer equilíbrio entre desempenho, interpretabilidade e viabilidade operacional.

3.3 Inteligência Artificial Generativa

A Inteligência Artificial Generativa representa uma evolução significativa no campo da IA, caracterizada pela capacidade de gerar novos conteúdos – como textos, imagens e padrões – a partir de grandes volumes de dados previamente processados. Conforme Ertel (2018), esses sistemas são capazes de aprender a partir de grandes volumes de dados e gerar novos conteúdos que se assemelham aos dados de treinamento, demonstrando uma forma de criatividade computacional. Diferentemente dos modelos discriminativos, os modelos generativos são projetados para criar saídas

plausíveis com base em padrões aprendidos, sendo cada vez mais explorados no desenvolvimento de sistemas antifraude adaptativos.

Venturini et al. (2024) apresentam a aplicabilidade prática da GenIA por meio da idealização de um monitor de fraudes bancárias baseado em análise de geolocalização, padrões de acesso e comportamento do usuário. A ferramenta descrita pelos autores é alimentada com registros de IP – *Internet Protocol*, horários e atividades suspeitas, permitindo que a IA detecte acessos fora do padrão e execute bloqueios automáticos, notificando os usuários em tempo real.

Além disso, modelos generativos como os LLMs (*Large Language Models*) – exemplificados pelo ChatGPT (OpenAI) e LLaMA (Meta AI) – têm sido testados como ferramentas auxiliares em ambientes financeiros. Eles são capazes de compreender linguagem natural e interpretar contextos complexos, possibilitando sua integração em *chatbots*, triagem de denúncias e geração de explicações sobre atividades incomuns (Mondragón; Yarango, 2025).

Por serem treinados com base em vastas quantidades de dados, esses modelos oferecem flexibilidade e adaptabilidade, mas também exigem cuidados quanto à validação de suas respostas. Sichman (2021) discorre sobre as “alucinações de IA” – respostas incorretas produzidas por modelos generativos – o que provoca a necessidade desses sistemas passarem por fases ainda mais robustas de teste e homologação antes da aplicação real.

A plataforma Vertex AI AutoML, utilizada no protótipo de monitoramento de fraudes bancárias de Venturini et al. (2024), automatiza o treinamento de modelos supervisionados e permite que a IA seja ajustada de forma contínua. Esse processo envolve ingestão de dados comportamentais, definição de atributos e validação cruzada, resultando em modelos que aprendem e evoluem com base no histórico de transações de cada cliente.

Em suma, a IA generativa se posiciona como uma tecnologia promissora no combate às fraudes bancárias, não apenas por sua capacidade analítica, mas por sua adaptabilidade e integração com outras fontes, como biometria e geolocalização. Contudo, seu uso requer acompanhamento técnico rigoroso e atenção às limitações éticas e regulatórias.

3.4 Biometria, geolocalização e IA no combate a fraudes

A evolução das tecnologias de autenticação digital ampliou o uso de dados biométricos, como impressões digitais, reconhecimento facial e voz, bem como de dados de geolocalização, como endereços IP e padrões de acesso por região, como ferramentas estratégicas na prevenção de fraudes bancárias. A Inteligência Artificial atua nos processos ao coletar, cruzar e interpretar esses dados em tempo real, identificando desvios em relação ao comportamento esperado de cada usuário. Segundo Mondragón e Yarango (2025), sistemas modernos de IA operam com múltiplas variáveis contextuais, como histórico de localização, dispositivos utilizados e biometria registrada, para construir modelos dinâmicos de risco, capazes de classificar transações suspeitas e acionar medidas preventivas automatizadas.

Cabe ainda mencionar que a biometria tem sido amplamente adotada por instituições financeiras como método de autenticação baseado em características únicas dos indivíduos. A IA, nesse contexto, atua tanto no processo de reconhecimento, por meio de algoritmos que codificam padrões biométricos, quanto na detecção de tentativas de falsificação, como ataques por máscaras faciais ou gravações de voz (Mondragón; Yarango, 2025).

Além do uso operacional, os dados biométricos também têm relevância jurídica. Em situações de contestação de operações bancárias, sistemas de IA podem auxiliar na construção de laudos periciais com base em logs biométricos. Segundo Soares e Carvalho (2023), a perícia no sistema de biometria do banco é a forma mais eficaz de esclarecer se houve ou não violação do sistema de segurança e eventual fraude na contratação eletrônica. Esses sistemas processam dados coletados no momento da autenticação e os relacionam com o perfil histórico do usuário, permitindo análises forenses com base em algoritmos treinados.

Belanda e Cavalcanti (2020) reforçam que, para além da segurança, a biometria reduz a fricção nas interações com o usuário e amplia a rastreabilidade dos contratos digitais, transformando-se também em um instrumento de prova contratual.

A geolocalização também tem ganhado destaque como mecanismo antifraude, especialmente quando associada a variáveis comportamentais e biométricas. Plataformas baseadas em IA utilizam dados como localização geográfica, tipo de dispositivo, sistema operacional e horário de acesso para estabelecer perfis de comportamento legítimos.

No estudo de Venturini et al. (2024), a IA generativa foi aplicada a um monitor antifraude capaz de cruzar dados de localização e comportamento, detectando

acessos fora do padrão habitual do usuário. Ao identificar uma tentativa de login em horário incomum ou a partir de um local não registrado, o sistema automaticamente acionava medidas preventivas, como dupla autenticação ou bloqueio temporário.

Mondragón e Yarango (2025) observam que a integração de geolocalização com modelos de decisão baseados em Aprendizado de Máquina torna possível a adaptação contínua dos sistemas antifraude. Com isso, a IA não apenas reage a eventos, mas aprende com os padrões históricos de cada cliente, antecipando riscos.

4 Trabalhos correlatos e estudos relevantes

Nos últimos anos, o número de pesquisas aplicando Inteligência Artificial à detecção de fraudes bancárias aumentou significativamente, refletindo a urgência de soluções tecnológicas capazes de lidar com ameaças complexas e em constante evolução. Esta seção apresenta os principais estudos comparativos e experimentais analisados na revisão bibliográfica, com foco em métricas de desempenho, algoritmos utilizados e desafios enfrentados.

O estudo de Mondragón e Yarango (2025) apresenta uma revisão sistemática da literatura científica sobre o uso da IA na detecção e prevenção de fraudes financeiras. Os autores destacam que os algoritmos mais recorrentes em pesquisas práticas são *Random Forest*, Regressão Logística, Redes Neurais Artificiais, SVM, LSTM e *Isolation Forest*, sendo estes frequentemente avaliados com base em métricas como acurácia, recall e AUC-ROC. Um dos principais desafios relatados nos estudos é o desbalanceamento dos dados, já que transações fraudulentas representam uma fração muito pequena do total, o que pode comprometer a eficácia de classificadores tradicionais.

Na mesma linha, o relatório da Mata (2023) traz uma análise comparativa aplicada a dados reais de operações bancárias em Portugal. O estudo mostrou que algoritmos como *Random Forest* e SVM apresentaram melhor desempenho em relação à detecção de fraudes com baixa taxa de falsos positivos, especialmente quando combinados com técnicas de balanceamento, como SMOTE. Além disso, os autores observaram que a simplicidade e interpretabilidade de modelos como a Regressão Logística continuam sendo vantajosas em ambientes com alta regulação.

A pesquisa de Venturini et al. (2024), por sua vez, traz uma abordagem inovadora ao propor o uso de Inteligência Artificial Generativa na construção de um

monitor antifraude com base em geolocalização e comportamento de acesso. O estudo destaca que modelos generativos, como os LLMs, podem complementar sistemas tradicionais ao fornecer respostas interpretáveis, apoiar operadores humanos e adaptar regras com base em linguagem natural. Contudo, os autores apontam que essas tecnologias ainda enfrentam desafios, como o risco de "alucinações" e a dificuldade de auditoria.

Por fim, Sichman (2021) chama atenção para as implicações éticas e os riscos sistêmicos da automação baseada em IA. Segundo o autor, embora as aplicações antifraude tenham alto potencial, é necessário garantir transparência, responsabilização e proteção aos usuários, principalmente em contextos em que decisões automatizadas podem gerar bloqueios indevidos ou prejuízos financeiros sem intervenção humana.

Diante do cenário analisado, observa-se que a literatura tem avançado no uso de IA tradicional e profunda para combate à fraude, mas ainda há lacunas importantes, especialmente quanto à integração entre modelos generativos, biometria e fatores contextuais (como localização e histórico de navegação). Essas lacunas fundamentam a relevância e a atualidade da proposta de estudo apresentada neste trabalho.

5 Metodologia

5.1 Tipo de pesquisa

Esta pesquisa, quanto à sua natureza, objetivo, procedimentos técnicos e abordagem metodológica, se enquadra nos critérios descritos no Quadro 1.

Quadro 1 – Procedimentos Técnicos e Abordagem Metodológica

Critério	Classificação
Natureza	Aplicada: visa gerar conhecimento com aplicação prática no combate a fraudes.
Objetivo	Exploratória e descritiva: busca compreender o tema e caracterizar suas variáveis.
Procedimentos Técnicos	Bibliográfica e documental: fundamentada em fontes secundárias e documentos oficiais.

Critério	Classificação
Abordagem Metodológica	Qualitativa: interpretação dos dados com base em significados e contextos.

Fonte: De autoria própria.

Esta pesquisa é de natureza aplicada, pois se propõe a investigar o uso de algoritmos de Inteligência Artificial no contexto das fraudes bancárias, com foco em sua aplicação na prevenção e monitoramento de transações suspeitas. Com isso, pretende-se gerar conhecimento que possa ser empregado em situações reais, especialmente na construção de soluções que empreguem IA em ambientes bancários.

5.2 Procedimentos metodológicos

Do ponto de vista técnico, esta pesquisa caracteriza-se como uma pesquisa bibliográfica, por se fundamentar na análise de materiais já publicados, como artigos científicos, dissertações, relatórios técnicos, documentos institucionais e normas regulatórias. Tais materiais oferecem a base teórica necessária para a compreensão dos conceitos fundamentais da Inteligência Artificial, das tecnologias empregadas e dos contextos de aplicação, conforme discutido nos tópicos da Seção 2.

Segundo Cervo, Silva e Bervian (2007), a bibliografia constitui uma fonte secundária de dados, reunindo as contribuições científicas acumuladas sobre determinado tema.

> [...] toda bibliografia já tornada pública em relação ao tema estudado, desde publicações avulsas, boletins, jornais, revistas, livros, pesquisas, monografias, teses, materiais cartográficos, etc. [...] e sua finalidade é colocar o pesquisador em contato direto com tudo o que foi escrito, dito ou filmado sobre determinado assunto [...] (Lakatos; Marconi, 2003, p. 183).

Adicionalmente, adota-se também a abordagem documental, considerando-se documentos oficiais, estatísticas públicas, relatórios técnicos (como os da FEBRABAN, DataSenado e Banco Central).

5.3 Objetivos e abordagem

Quanto aos objetivos, a pesquisa é de caráter exploratório, já que busca ampliar o entendimento sobre um campo emergente – o uso de IA, incluindo IA

generativa, na prevenção de fraudes bancárias – cuja delimitação temática é essencial para posterior aprofundamento.

> Muitas vezes as pesquisas exploratórias constituem a primeira etapa de uma investigação mais ampla. Quando o tema escolhido é bastante genérico, tornam-se necessários seu esclarecimento e delimitação, o que exige revisão da literatura, discussão com especialistas e outros procedimentos. O produto final deste processo passa a ser um problema mais esclarecido, passível de investigação mediante procedimentos mais sistematizados (Gil, 2008, p. 27).

Além disso, a investigação assume natureza descritiva, pois visa identificar, caracterizar e classificar os diferentes algoritmos e abordagens tecnológicas utilizados em sistemas antifraude. “São estas as que habitualmente realizam os pesquisadores sociais preocupados com a atuação prática” (Gil, 2002, p. 42). A intenção é mapear como essas soluções têm sido aplicadas na prática, conforme evidenciado pelos trabalhos analisados na Seção 3, por exemplo, os que descrevem o uso de redes neurais, algoritmos de aprendizado supervisionado, IA generativa (como ChatGPT e Vertex AI) e biometria.

Por meio dessa abordagem, pretende-se verificar a relação entre a utilização das tecnologias de Inteligência Artificial e a prevenção das fraudes bancárias. As pesquisas descritivas, segundo Gil (2002), têm por finalidade revelar possíveis associações entre variáveis relevantes ao tema estudado.

5.4 Abordagem metodológica

Quanto a abordagem, busca-se uma análise qualitativa, em que os dados coletados são predominantemente descritivos e que tem por finalidade dar significado a esses dados. Com a descrição qualitativa oportuniza-se não só demonstrar a aparência do fenômeno, mas também a sua essência, explicando sua origem, relações e mudanças, na tentativa de inferir as consequências (Triviños, 1987).

Assim, o que se busca não é quantificar relações numéricas, mas interpretar os significados e implicações das tecnologias analisadas, suas limitações, potencialidades e riscos associados, como alucinações de IA, viés algorítmico, e a capacidade de adaptação contínua dos modelos.

Essa escolha metodológica permite compreender a complexidade das soluções baseadas em IA, especialmente no contexto de segurança bancária, onde variáveis como comportamento do usuário, localização geográfica, padrões de acesso e contexto regulatório são cruciais para a eficácia dos sistemas antifraude.

6 Resultados e discussão

A análise da literatura especializada evidenciou que a aplicação de Inteligência Artificial no combate às fraudes bancárias é uma abordagem consolidada, porém ainda em evolução. Entre os principais resultados encontrados está a superioridade de algoritmos como *Random Forest*, SVM e LSTM na detecção de transações suspeitas, especialmente em cenários com grande volume de dados e alta incidência de ruído. Esses modelos destacaram-se por sua capacidade de manter baixos índices de falsos positivos e elevados níveis de acurácia (Bhattacharyya et al., 2011; Dal Pozzolo et al., 2014; Mondragón; Yarango, 2025).

Adicionalmente, verificou-se que técnicas não supervisionadas, como *Isolation Forest* e *Autoencoders*, possuem grande utilidade na identificação de anomalias em dados não rotulados, característica comum em bases de transações bancárias. Esses métodos contribuem significativamente na fase inicial de análise exploratória, sendo eficazes na identificação de padrões atípicos que indicam potenciais fraudes (Liu, Ting e Zhou, 2008; Mata, 2023).

Um dos avanços mais relevantes apontados nos estudos recentes é o uso de Inteligência Artificial Generativa. Modelos como o ChatGPT e o LLaMA têm sido testados em protótipos de monitoramento antifraude baseados em linguagem natural e geolocalização (Venturini et al., 2024), mostrando boa adaptabilidade para o suporte à tomada de decisão em tempo real. No entanto, os mesmos estudos também alertam para riscos associados, como a possibilidade de respostas incorretas, conhecidas como alucinações, e a falta de explicabilidade dos modelos, o que pode comprometer sua aceitação em ambientes regulados (Sichman, 2021; Mondragón; Yarango, 2025).

A integração entre biometria, geolocalização e IA também aparece como uma tendência promissora. Sistemas baseados em múltiplas variáveis comportamentais são capazes de gerar perfis dinâmicos de risco, ampliando a efetividade das medidas preventivas e permitindo reações automáticas a comportamentos atípicos (Soares; Carvalho, 2023; Venturini et al., 2024).

Por fim, a literatura destaca três grandes desafios recorrentes:

- O desbalanceamento das bases de dados, que requer o uso de técnicas como o SMOTE para manter o desempenho dos classificadores (Dal Pozzolo et al., 2014).

- A explicabilidade dos modelos, essencial para auditorias e conformidade regulatória (Sichman, 2021).
- A necessidade de treinamento contínuo e adaptação dos sistemas frente a novas modalidades de ataque (Mata, 2023; Mondragón; Yarango, 2025).

Diante da análise dos estudos revisados, é possível perceber que, embora a IA represente um avanço significativo na capacidade de resposta às fraudes bancárias, sua adoção prática ainda esbarra em obstáculos técnicos e normativos. A presença de algoritmos altamente eficazes, como *Random Forest*, LSTM e *Isolation Forest*, não elimina a necessidade de validação contínua e de cuidados com o viés algorítmico, sobretudo em ambientes regulados e sensíveis como o financeiro. A crítica mais evidente está na dicotomia entre performance e explicabilidade: os modelos mais sofisticados, como os de *deep learning* e GenIA, tendem a operar como "caixas-pretas", dificultando auditorias e a aceitação institucional. Isso indica que o progresso técnico precisa caminhar lado a lado com mecanismos de transparência.

Outro ponto relevante é a necessidade de integração entre abordagens. A literatura apresenta que modelos híbridos, que combinam algoritmos supervisionados e não supervisionados, técnicas de balanceamento e dados contextuais como biometria e geolocalização, oferecem maior robustez na identificação de fraudes. No entanto, a aplicação desse ecossistema tecnológico ainda depende de infraestrutura adequada, padronização dos dados e alinhamento com normas jurídicas. A crítica que se impõe, portanto, é que, apesar da eficácia das ferramentas analisadas, há um distanciamento entre o potencial técnico e a maturidade institucional para sua implementação plena. A superação dessa lacuna passa pela colaboração entre desenvolvedores, reguladores e operadores do sistema financeiro.

Considerações finais

Este estudo teve como objetivo analisar o uso da Inteligência Artificial, com ênfase em algoritmos tradicionais, técnicas de *Deep Learning* e abordagens generativas, na detecção e prevenção de fraudes bancárias. A partir de uma pesquisa de caráter qualitativo, bibliográfica e documental, foi possível identificar as principais estratégias utilizadas no setor financeiro para combater transações fraudulentas, bem como os desafios envolvidos na implementação dessas soluções.

Destaca-se que algoritmos supervisionados, como *Random Forest*, SVM e Regressão Logística continuam sendo amplamente adotados devido à sua robustez e bom desempenho em contextos com dados desbalanceados. Técnicas como LSTM, *Autoencoders* e *Isolation Forest* também demonstram grande potencial, especialmente na identificação de padrões anômalos em tempo real. Além disso, o avanço da Inteligência Artificial Generativa abre novas possibilidades para o desenvolvimento de sistemas mais flexíveis e interpretativos, como observado na utilização de modelos como ChatGPT e LLaMA.

Apesar dos avanços, o estudo evidenciou que ainda há lacunas relevantes a serem exploradas. Entre elas, destacam-se a necessidade de maior explicabilidade dos modelos, a integração efetiva entre IA, biometria e geolocalização, e o desenvolvimento de mecanismos que garantam a transparência e a segurança jurídica nas decisões automatizadas. O risco de decisões incorretas sem intervenção humana e a possibilidade de vieses algorítmicos também foram apontados como pontos críticos.

Como desdobramento futuro, sugere-se o aprofundamento da pesquisa empírica por meio de estudos de caso em instituições financeiras, bem como a construção de protótipos que integrem IA generativa com dados comportamentais e jurídicos, considerando o cenário regulatório brasileiro. A combinação entre tecnologia, regulação e ética será essencial para consolidar sistemas antifraude que sejam eficazes, confiáveis e socialmente responsáveis.

Referências

BELANDA, Douglas; CAVALCANTI, Ana Elizabeth Lapa Wanderley. **Biometria como mecanismo de formação e prova contratual: um olhar para as transações eletrônicas bancárias na sociedade da informação.** Revista dos Tribunais, São Paulo: RT, v. 1016, 2020. Disponível em: https://dspace.almg.gov.br/handle/11037/37570. Acesso em: 10 abr. 2025.

BHATTACHARYYA, Siddhartha. JHA, Sanjeev. THARAKUNNEL, Kurian. WESTLAND, J. **Data mining for credit card fraud: A comparative study.** Decision Support Systems, v. 50, p. 602-613. 2011. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0167923610001326?via%3Dihub. Acesso em: 11 abr. 2025.

BRASIL. **Decreto-Lei nº 2.848, de 7 de dezembro de 1940. Código Penal**. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm. Acesso em: 11 abr. 2025.

CARLE, Eben. **Ask a Techspert: What is generative AI?** Website, 11/mar/2023. Disponível em: https://blog.google/inside-google/googlers/ask-a-techspert/what-is-generative-ai/. Acesso em: 11 abr. 2025.

CERVO, Amado Luiz; SILVA, Roberto da; BERVIAN, Pedro A. **Metodologia científica.** 6. ed. São Paulo: Pearson, 2007. 176 p.

CORTES, Corinna; VAPNIK, Vladimir. **Support-vector networks. Machine Learning**, v. 20, p. 273–297, 1995. Disponível em: https://scispace.com/papers/support-vector-networks-2jd9a1kl0z. Acesso em: 11 abr. 2025.

DAL POZZOLO, Andrea. CAELEN, Olivier. LE BORGNE, Yann-Aël. WATERSCHOOT, Serge. BONTEMPI, Gianluca. **Learned lessons in credit card fraud detection from a practitioner perspective.** Expert Systems with Applications, v. 41, p. 4915–4928. 2014. Disponível em: https://www.researchgate.net/publication/260837261_Learned_lessons_in_credit_card_fraud_detection_from_a_practitioner_perspective. Acesso em: 12 abr. 2025.

ERTEL, Wolfgang. **Introduction to Artificial Intelligence**. 2. ed. Cham, Switzerland: Springer, 2018. 365 p.

FACELI, Katti; VELLOSO, Tiago A.; ZANCUL, Eduardo de Souza; COSTA, João P.; ARAÚJO, Raydonal O. **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina.** 2. ed. Rio de Janeiro: LTC, 2021. 304 p.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 4ª ed. São Paulo: Atlas, 2002. 171 p.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social.** 6ª ed. São Paulo: Atlas, 2008. 200 p.

HOSMER Jr., D. W. LEMESHOW, S. STURDIVANT, R. X. **Applied Logistic Regression.** 2013. Disponível em: https://onlinelibrary.wiley.com/doi/chapter-epub/10.1002/9781118548387.fmatter. Acesso em: 11 abr. 2025.

LAKATOS, Eva Maria. MARCONI, Marina de Andrade. **Fundamentos de Metodologia Científica.** São Paulo: Atlas, 2003. 368 p.

LIMA, Jardielma Queiroz de. **Detecção de fraudes em cartões de crédito utilizando técnicas de aprendizado de máquina.** Trabalho de Conclusão de Curso - Sistema de Informação. Instituto Federal do Espírito Santo. 2022. 76 p.

LIU, Fei Tony. TING, Kai Ming. ZHOU, Zhi-Hua. **Isolation forest.** In: 2008 Eighth IEEE International Conference on Data Mining. Pisa: IEEE, 2008. p. 413–422. Disponível em: https://www.researchgate.net/publication/224384174_Isolation_Forest. Acesso em: 12 abr. 2025.

MATA, Pedro Alexandre T.F. **Análise comparativa de algoritmos de inteligência artificial na deteção de fraude transacional em contexto bancário: uma revisão sistemática de literatura.** Dissertação em Análise de Dados e Sistemas de Apoio à Decisão – Coimbra Business School, Politécnico de Coimbra, 2023. 2023. 81 p. Disponível em: https://comum.rcaap.pt/handle/10400.26/48729. Acesso em: 11 abr. 2025.

MONDRAGÓN, M. G. A.; YARANGO, L. C. V. **Revisión sistemática sobre el uso de la Inteligencia Artificial para la detección y prevención de fraudes financieros.** Cuenca: Universidad Politécnica Salesiana, 2025. Disponível em: https://repositorio.uss.edu.pe/handle/20.500.12802/14029. Acesso em: 11 abr. 2025.

SICHMAN, Jaime Simão. **Inteligência Artificial e Sociedade: Avanços e Riscos. Estudos Avançados**, São Paulo, v. 35, n. 101, p. 37-50, 2021. Disponível em: https://www.revistas.usp.br/eav/article/view/185024. Acesso em: 11 abr. 2025.

SOARES, Dennis Verbicaro; CARVALHO, Emerson Benjamin Pereira de. **Fraudes em contratos eletrônicos de empréstimos bancários: vulnerabilidade do consumidor, inteligência artificial e prova pericial em sistemas de biometria.** Revista Pensamento Jurídico, São Paulo, v. 17, n. 2, 2023. Disponível em: https://ojs.unialfa.com.br/index.php/pensamentojuridico/article/view/813. Acesso em: 11 abr. 2025.

SOUZA, D. H. M.; BORDIN JR., C. J. **Detecção de fraude de cartão de crédito por meio de algoritmos de aprendizado de máquina**. Revista Brasileira de Computação Aplicada, v. 15, p. 1–11, 2023. Disponível em https://doi.org/10.5335/rbca.v15i1.13790. Acesso em: 11 abr. 2025.

TRIVIÑOS, Augusto Nibaldo Silva. **Introdução à pesquisa em ciências sociais: A pesquisa qualitativa em educação.** O positivismo, a fenomenologia e o marxismo. São Paulo: Atlas, 1987. 175 p.

VENTURINI, Leonardo do Carmo; SILVA, Otávio Brandão da; OLIVEIRA, Thiago Necolau de; KALILI, Roberto Marcos. **Idealização de monitor de fraudes bancárias a partir da inteligência artificial generativa**. Revista Contemporânea, v. 4, p. 01-20, 2024. Disponível em: https://ojs.revistacontemporanea.com/ojs/index.php/home/article/view/6719. Acesso em: 11 abr. 2025.

VILELA, Erica; UEDA, Eduardo Takeo; GAVA, Vagner Luiz. **Phishing e engenharia social: conceitos, modalidades, técnicas de detecção e prevenção de fraudes; uma revisão sistemática da literatura.** Revista de Tecnologia e Sociedade, São Paulo, v. 19, p. 1-22, 2023. Disponível em: https://ipt.br/2023/01/27/phishing-e-engenharia-social-conceitos-modalidades-tecnicas-de-deteccao-e-prevencao-de-fraudes-uma-revisao-sistematica-da-literatura/. Acesso em: 12 abr. 2025.

ZHANG, Y.; SHAO, H.; LI, Y. **A hybrid fraud detection algorithm based on data mining. Procedia Computer Science**, v. 122, p. 689–694, 2017. Disponível em: https://www.computer.org/csdl/proceedings-article/icbaie/2020/09196322/1n90WxjuO6A. Acesso em: 12 abr. 2025.