

Centro Universitário Euro Americano (Unieuro)

Sistemas de Informação

Luan Medrado Moreira

Lucas Rego Da Silva

**Apresentação de análise de dados para detecção**

**de fraude bancária**

Brasília

2024

Luan Medrado Moreira

Lucas Rego Da Silva

**Apresentação de análise de dados para detecção**

**de fraude bancária**

Trabalho da disciplina de Projeto Integrador de Segurança da Informação

Orientador: Prof. Msc. Edward Lima M. de Melo

Brasília

2024

**Resumo**

A fraude bancária, caracterizada por transações não autorizadas pelos titulares dos cartões, representa um grande desafio para as instituições financeiras e seus clientes (Bolton & Hand, 2002). Com o aumento da digitalização dos serviços bancários, os casos de fraudes têm se tornado cada vez mais frequentes e complexos, o que requer técnicas avançadas de detecção e mitigação (Whitrow et al., 2009).

Neste contexto, a análise de dados surge como uma ferramenta essencial para a identificação e prevenção de fraudes (Kou et al., 2014). Esse projeto tem como objetivo a utilização de um modelo de machine learning para detectar fraudes bancárias em uma base de dados escolhida, que gera um relatório que será analisado, com a hipótese de que o modelo treinado pode identificar, de forma certeira, todas as ocorrências de fraude presentes. O estudo abrange etapas de higienização de dados, treinamento e teste do modelo, e a análise dos índices de fraudes encontradas. Por fim, é esperado que os resultados contribuam para a elaboração de estratégias melhores de diminuição de fraudes, gerando assim maior segurança e confiabilidade para os consumidores e instituições.

**Palavras-chave**: Fraude bancária, Machine Learning, Análise de dados, Identificação de fraudes, Higienização dos dados.

Introdução...................................................................................... 1

1.1 Contextualização do Problema........................................................ 2

1.2 Hipótese........................................................................................... 3

1.3 Objetivo Geral.................................................................................. 4

1.4 Objetivos Específicos...................................................................... 5

**Justificativa**.......................................................................................... 6

2.1 Relevância do Tema........................................................................ 7

2.2 Impacto da Fraude Bancária nas Instituições Financeiras.............. 8

2.3 Importância da Detecção e Prevenção de Fraudes........................ 9

**Revisão da Literatura**......................................................................... 10

3.1 Conceitos de Fraude Bancária....................................................... 11

3.2 Tipos de Fraude em Transações Eletrônicas................................. 12

3.3 Machine Learning para Detecção de Fraudes................................ 13

**Metodologia**......................................................................................... 14

4.1 Coleta de Dados............................................................................. 15

4.1.1 Fontes de Dados....................................................................... 16

4.1.2 Seleção de Variáveis................................................................ 17

4.2 Higienização de Dados.................................................................. 18

4.3 Modelagem e Treinamento............................................................ 19

4.3.1 Escolha do Modelo de Machine Learning................................. 20

4.3.2 Divisão da Base de Dados (Treinamento e Teste) .................. 21

4.4 Avaliação do Modelo...................................................................... 22

4.4.1 Métricas de Avaliação............................................................... 23

4.4.2 Validação.................................................................................. 24

**Análise de Resultados**...................................................................... 25

5.1 Identificação das Fraudes............................................................. 26

5.2 Interpretação dos Índices de Fraude............................................ 27

5.3 Apresentação de Dashboard......................................................... 28

**Conclusão**.......................................................................................... 29

6.1 Síntese dos Resultados................................................................. 30

6.2 Sugestões para Trabalhos Futuros................................................ 31

6.3 Contribuições para o Combate à Fraude Bancária........................ 32

# INTRODUÇÃO

## **1.1. Contextualização do Problema**

A fraude bancária é uma prática criminosa que afeta tanto instituições financeiras quanto consumidores, resultando em perdas financeiras significativas e prejudicando a confiança dos clientes nos sistemas bancários (Bolton & Hand, 2002). Com a crescente digitalização dos sistemas financeiros e o uso em massa de cartões de crédito, a ocorrência de transações fraudulentas tornou-se mais comum e sofisticada, desafiando os métodos padrões de detecção e prevenção. Os criminosos se utilizam de diversas técnicas para realizar transações fraudulentas, desde roubo de identidade até o uso de algoritmos e métodos avançados de ataque (Delamaire et al., 2009).

Para se proteger, as instituições financeiras procuram métodos tecnológicos para detectar padrões suspeitos e impedir transações ilícitas em tempo real, sem comprometer a experiência do usuário. Entre esses métodos, a análise de dados e o uso de machine learning emergem como ferramentas essenciais. Ao processar grandes volumes de dados, esses modelos são capazes de identificar padrões atípicos que podem caracterizar uma fraude, possibilitando uma resposta mais ágil e precisa na prevenção de crimes financeiros (Carcillo et al., 2018). Assim, a aplicação de técnicas avançadas de análise de dados para detecção de fraude tornou-se não apenas uma necessidade, mas uma prática indispensável para a segurança e integridade das operações financeiras (Ngai et al., 2011).

## **1.2. Hipótese**

O modelo de machine learning será capaz de identificar com precisão todas as fraudes presentes na base de dados, permitindo a detecção eficaz de padrões fraudulentos e contribuindo para uma análise confiável dos índices de fraude bancária.

## **1.3. Objetivo Geral**

Apresentar uma análise de dados de fraude bancária através de detecção por um modelo de machine Learning.

## **1.4. Objetivos Específicos**

* Selecionar uma base de dados relevante para o estudo.
* Realizar a higienização da base de dados, aplicando técnicas de limpeza, normalização e padronização.
* Aplicar o modelo de machine learning aos dados preparados.
* Treinar e ajustar o modelo de machine learning.
* Analisar os resultados obtidos.

# Justificativa

## **2.1. Relevância do Tema**

A crescente incidência de fraudes bancárias representa um desafio crítico para o setor financeiro, afetando a segurança das transações e a confiança dos clientes nas instituições financeiras (Bolton & Hand, 2002). Tendo em vista a digitalização dos serviços bancários e o crescimento exponencial do uso de transações eletrônicas, as transações fraudulentas estão evoluindo e se tornando cada vez mais sofisticadas, pedindo soluções tecnológicas que exigem maior complexidade técnica para a sua identificação e prevenção.

A análise de dados e o uso de técnicas de machine learning emergem como ferramentas estratégicas nesse contexto, pois possibilitam a identificação rápida e precisa de transações fraudulentas (Awoyemi et al., 2017).

Além de proteger as instituições financeiras contra perdas significativas, a detecção eficaz de fraudes beneficia diretamente os consumidores, promovendo um ambiente financeiro mais seguro e confiável (Bhattacharyya et al., 2011). Ao utilizar e estudar modelos de machine learning voltado à detecção de fraudes, este projeto contribui e oferece suporte prático para o desenvolvimento de métodos que reforçam a integridade do sistema bancário, atendendo a uma necessidade urgente tanto para o setor financeiro quanto para a população.

## **2.2. Impacto da Fraude Bancária nas Instituições Financeiras**

A fraude bancária impacta as instituições financeiras de forma profunda, comprometendo não apenas a saúde financeira, mas também a confiança do cliente (Bolton & Hand, 2002). As perdas financeiras associadas às fraudes são significativas e afetam diretamente os lucros, uma vez que as instituições são, muitas vezes, responsáveis por ressarcir os valores aos clientes prejudicados (Bhattacharyya et al., 2011). Esses prejuízos, junto aos custos de desenvolver sistemas de segurança mais fortes, trazem um ônus considerável. Além disso, as fraudes constantes podem gerar multas regulatórias e perdas à imagem institucional, pois a falta de segurança notada pelos consumidores diminui a confiabilidade e impacta a competitividade no mercado.

Além do impacto financeiro direto, as instituições também enfrentam desafios operacionais e de reputação. A necessidade de responder rapidamente a incidentes de fraude implica em esforços constantes de monitoramento e atualização tecnológica, o que demanda tempo, recursos humanos e financeiros (Ngai et al., 2011). Adicionalmente, a reputação de uma instituição pode ser gravemente prejudicada se considerada vulnerável a ataques, o que reduz o engajamento de novos clientes e pode levar à perda de clientes atuais (Delamaire et al., 2009). Assim, a detecção e prevenção de fraudes bancárias se torna uma prioridade estratégica, necessária para manter a credibilidade e a longevidade das instituições financeiras.

**2.3. Importância da Detecção e Prevenção de Fraudes**

A Identificação e prevenção de fraudes são de extrema relevância para o setor bancário, pois transações ilícitas representam não apenas prejuizos financeiros significativos para instituições e clientes, mas também impactam a confiança dos consumidores e a reputação das instituições financeiras. Com a crescente das transações digitais, a detecção eficaz de ações suspeitas virou um dos principais desafios das instituições, que necessitam reagir contra crimes financeiros sem comprometer a experiência do usuário.

Prevenir fraudes é fundamental para garantir a segurança e a integridade das operações financeiras, gerando uma diminuição nos danos econômicos e protegendo os dados e os bens dos clientes. Além disso, a imediata detecçãp de fraudes possibilita que instituições tomem medidas de correção em tempo real, evitando que fraudes se tornem problemas maiores, como o roubo de identidade e a disseminação de informações sensíveis. Assim, os investimentos em tecnologias de análise de dados e machine learning para detecção e prevenção de fraudes são cada vez mais essenciais, tornando o setor financeiro mais resiliente e confiável (Carcillo et al., 2018).

# REVISÃO DA LITERATURA

**3.1. Conceitos de Fraude Bancária**

A fraude bancária é definida como qualquer atividade ilegal que visa obter vantagens financeiras ilícitas em transações bancárias, seja através de métodos de falsificação, roubo de identidade ou transações fraudulentas (Bolton & Hand, 2002). A literatura identifica diversos tipos de fraudes bancárias, incluindo fraude em cartões de crédito, fraudes de empréstimos, e lavagem de dinheiro, cada uma com suas particularidades em termos de execução e impacto sobre as instituições financeiras e consumidores (Delamaire et al., 2009). Com a evolução das tecnologias digitais, essas fraudes tornaram-se mais complexas e sofisticadas, envolvendo, por exemplo, ataques cibernéticos e o uso de malware para roubo de informações.

**3.2. Tipos de Fraude em Transações Eletrônicas**

A fraude em transações eletrônicas abrange uma variedade de práticas ilegais que visam obter vantagens financeiras por meio de tecnologias digitais, especialmente no contexto bancário. A literatura destaca diversos tipos de fraudes em transações eletrônicas, que variam em complexidade e impacto:

1. **Fraude com Cartão de Crédito**: É uma das fraudes eletrônicas mais comuns e envolve o uso não autorizado de informações de cartão de crédito para realizar transações fraudulentas. Essas fraudes podem ocorrer tanto online quanto em lojas físicas, e são facilitadas pelo roubo de dados do cartão, seja por phishing ou malware (Delamaire et al., 2009).
2. **Phishing e Spear Phishing**: Essas práticas envolvem o envio de e-mails, mensagens de texto ou chamadas que fingem ser de instituições financeiras legítimas para enganar os usuários e obter informações confidenciais, como senhas e números de cartões. A diferença entre phishing e spear phishing é que o último é mais direcionado e personalizado, tornando-o mais difícil de detectar (Huang et al., 2020).
3. **Malware Bancário**: Este tipo de fraude ocorre quando um software malicioso é instalado no dispositivo do usuário, permitindo que criminosos capturem dados de login e outras informações bancárias sensíveis. Exemplos conhecidos incluem trojans bancários e ransomware, que podem comprometer gravemente a segurança de dados financeiros (Polla et al., 2013).
4. **Ataques Man-in-the-Middle (MitM)**: Nesse tipo de fraude, o invasor intercepta a comunicação entre o usuário e o servidor da instituição financeira, permitindo que o fraudador capture informações sensíveis ou altere a transação. Isso ocorre frequentemente em conexões Wi-Fi desprotegidas ou através de aplicativos inseguros (Bhargava & Sharma, 2013).
5. **Fraude de Identidade e Abertura de Contas Falsas**: Criminosos utilizam identidades roubadas ou falsas para abrir contas bancárias e realizar transações, frequentemente associadas a esquemas de lavagem de dinheiro. Essa prática é facilitada por brechas na autenticação de novos clientes e pode ser difícil de identificar sem um monitoramento robusto (Alzahrani et al., 2020).

**3.3. Machine Learning para Detecção de Fraudes**

A aplicação de machine learning para detecção de fraudes tornou-se uma abordagem indispensável na proteção de sistemas financeiros, permitindo identificar padrões suspeitos e prevenir atividades fraudulentas de forma rápida e eficiente. Modelos de machine learning, como árvores de decisão, redes neurais e algoritmos de aprendizado profundo, são amplamente utilizados devido à sua capacidade de aprender padrões complexos a partir de dados históricos e identificar anomalias em tempo real (Awoyemi et al., 2017). O uso de técnicas supervisionadas e não supervisionadas também possibilita a identificação de fraudes mesmo em cenários onde os dados estão desequilibrados, uma característica comum em conjuntos de dados de fraudes bancárias.

# METODOLOGIA

# REFERÊNCIAS

 G. J. Bolton e D. J. Hand, “Statistical Fraud Detection: A Review,” Statistical Science, vol. 17, no. 3, pp. 235–249, 2002. Acessado 26 outubro de 2024.

 N. P. Whitrow, D. J. Hand, P. Juszczak, D. Weston e N. M. Adams, “Transaction aggregation as a strategy for credit card fraud detection,” Data Mining and Knowledge Discovery, vol. 18, pp. 30–55, 2009. Acessado 26 outubro de 2024.

 Y. Kou, C. Lu, S. Sirwongwattana e Y. Huang, “Survey of fraud detection techniques,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 557–583, 2014. Acessado 26 outubro de 2024.

 L. Delamaire, H. Abdou e J. Pointon, “Credit card fraud and detection techniques: A review,” Banks and Bank Systems, vol. 4, no. 2, pp. 57–68, 2009. Acessado 26 outubro de 2024.

 F. Carcillo, Y.-A. Le Borgne, O. Caelen e G. Bontempi, “Streaming active learning strategies for real-life credit card fraud detection: Assessment and visualization,” International Journal of Data Science and Analytics, vol. 5, pp. 285–300, 2018. Acessado 27 outubro de 2024.

 O. Awoyemi, A. Adetunmbi e S. Oluwadare, “Credit card fraud detection using machine learning techniques: A comparative analysis,” IEEE International Conference on Computing Networking and Informatics (ICCNI), pp. 1–7, 2017. Acessado 27 outubro de 2024.

 D. Bhattacharyya, S. Jha, K. Tharakunnel e J. C. Westland, “Data mining for credit card fraud: A comparative study,” Decision Support Systems, vol. 50, no. 3, pp. 602–613, 2011. Acessado 27 outubro de 2024.

 E. Ngai, Y. Hu, Y. Wong, Y. Chen e X. Sun, “The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature,” Decision Support Systems, vol. 50, no. 3, pp. 559–569, 2011. Acessado 27 outubro de 2024.

 L. Huang, T. Zhao e J. Li, “Phishing detection based on deep learning and natural language processing,” IEEE Access, vol. 8, pp. 179218–179228, 2020. Acessado 2 novembro de 2024.

 M. Polla, F. Martinelli e D. Sgandurra, “A survey on security for mobile devices,” IEEE Communications Surveys & Tutorials, vol. 15, no. 1, pp. 446–471, 2013. Acessado 2 novembro de 2024.

 A. Bhargava e A. Sharma, “Decision tree analysis on J48 algorithm for data mining,” International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 6, pp. 1114–1119, 2013. Acessado 2 novembro de 2024.

 A. Alzahrani, M. Jalal e A. Firdhous, “Machine learning based fraud detection in social security using multi-layer feedforward neural networks,” Procedia Computer Science, vol. 175, pp. 45–51, 2020. Acessado 2 novembro de 2024.