

# Universidade de Pernambuco

## Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

### Proposta de Dissertação de Mestrado

**Área: Computação Inteligente**

**Título: Desenvolvimento de Uma Metodologia Baseada em Teoria dos Grafos para Garantia da Qualidade de Dados em Transações Bancárias Utilizadas na Análise de Tipologias de Lavagem de Dinheiro**

**Orientador(a): Prof. Dr. Diego Marconi Pinheiro Ferreira Silva**  
([dmpfs@ecom.poli.br](mailto:dmpfs@ecom.poli.br))

**Coorientador: Prof. Dr. Dimas Cassimiro do Nascimento Filho**  
([dimas.cassimiro@ufape.edu.br](mailto:dimas.cassimiro@ufape.edu.br))

#### Descrição:

A lavagem de dinheiro é a ocultação da propriedade de bens, direitos ou valores provenientes de infração penal (BRASIL, 1998). Segundo estimativas, anualmente cerca de 1-2 trilhões de dólares são fruto da lavagem de dinheiro. Isso corresponde a 2-5% do PIB global (Jensen; Iosifidis, 2023). O Brasil aparece como um dos líderes nos rankings dos países que mais lavam dinheiro no mundo (Henrique S. Assumpção *et al.*, 2022), e recentemente escândalos bilionários envolvendo o crime organizado e empresas legítimas e de fachada fortalecem o status quo. Sempre que existem indícios de lavagem de dinheiro nas movimentações bancárias, as instituições financeiras devem por lei encaminhar ao Conselho de Controle de Atividades Financeiras (COAF), que em seguida aciona as autoridades de segurança pública. A Polícia Civil do Estado de Pernambuco é uma dessas autoridades, e recebe os casos com esses indícios. Após a aprovação da quebra de sigilo financeiro pelo poder judiciário, o Laboratório de Tecnologia Contra a Lavagem de Dinheiro (LABLD) da PCPE recebe uma grande volume de transações bancárias dos suspeitos investigados no período do caso investigado. O LABLD realiza uma análise minuciosa dos dados e caracteriza as tipologias de lavagem de dinheiro presentes nas transações (BACEN, 2020), para, por fim, elaborar um Relatório de Análise de Dados Bancários (RADB).

O RADB é um dos elementos que fundamentam as operações contra o crime organizado, e grande parte do tempo é dedicado à análise dos dados das transações bancárias pelo LABLD. Essa análise é realizada manualmente pelos analistas a partir de um enorme volume de dados e leva de 1 a 2 meses para ser concluída. Em razão da heterogeneidade dos dados oriundos de múltiplas instituições financeiras, antes da identificação de tipologias, o LABLD ainda realiza o processo de quarentena, no qual são avaliadas as transações para identificar inconsistências de valores e saldos, bem como transações faltantes e questões de ordenação. Permanecem como principais desafios para a PCPE a automação da identificação das tipologias de lavagem de dinheiro e a validação dos dados das transações bancárias. A identificação de tipologias de lavagem de dinheiro envolve o reconhecimento de padrões em séries temporais de transações bancárias. As séries temporais são sequências de observações coletadas ao longo do tempo. Atualmente, diferentes abordagens automatizadas de Redes Neurais Recorrentes (RNN) são utilizadas para reconhecer os padrões em séries temporais de transações bancárias (Xavier *et al.*, 2024; Girish; Bhowmik, 2024; Anaroch; Tosakulvong; Wattanapongsakorn, 2025; Liu, 2024; Venkata Lakshmi Narayana; Panigrahi, 2025). Dessa forma, o tempo necessário para realizar as análises pode ser reduzido de meses para poucos dias. No entanto, essas abordagens exigem que, dentro de uma janela temporal, as transações bancárias estejam ordenadas, dado que as RNNs são altamente sensíveis à ordem de entrada dos dados.

A **justificativa** desta proposta é contribuir tanto para o processo de quarentena do LABLD da PCPE, assegurando a qualidade da ordenação dos dados de transações bancárias, quanto para o aprimoramento dos modelos preditivos baseados em séries temporais utilizados na identificação de tipologias de lavagem de dinheiro. Trabalhos que compõem o **estado da arte** já evidenciam a importância da qualidade dos dados e seu impacto no desempenho de modelos estatísticos preditivos (Wong; Wong, 2021; Nugroho, 2023). A lacuna reside no fato de que essas abordagens concentram-se predominantemente na análise individual das variáveis, negligenciando a **qualidade das relações entre os dados**. Nesse contexto, surge a seguinte **pergunta de pesquisa**: como caracterizar a qualidade relacional dos dados em séries temporais multivariadas e de que forma a qualidade relacional impacta o desempenho de modelos preditivos baseados em redes neurais recorrentes?

O **propósito** deste projeto de pesquisa é desenvolver uma metodologia baseada em teoria dos grafos para caracterizar a qualidade relacional de dados em séries temporais multivariadas e aplicá-la às transações bancárias utilizadas pelo LABLD da PCPE na identificação de tipologias de lavagem de dinheiro. Por meio de métodos fundamentados em teoria dos grafos (Jungnickel, s.d.; Gross; Yellen; Anderson, 2018; West *et al.*, 2001), serão construídos modelos de fluxos bancários a partir dos dados de transações financeiras, com o objetivo de avaliar sua qualidade e consistência, bem como determinar a ordem correta de ocorrência dos fluxos transacionais. A partir dessas verificações, sistemas especialistas poderão apoiar os analistas do LABLD da PCPE na identificação de tipologias de lavagem de dinheiro com maior agilidade, confiabilidade e rastreabilidade analítica. Além de avançar o estado da arte em qualidade de dados e reconhecimento de padrões em séries temporais multivariadas, este projeto também contribuirá para o combate à lavagem de dinheiro e para o fortalecimento institucional do LABLD e da Polícia Civil de Pernambuco, por meio do FUNREPOL, que destina recursos oriundos de operações bem-sucedidas ao reaparelhamento e à modernização do órgão.

O **objetivo geral** desta proposta é desenvolver uma metodologia baseada em teoria dos grafos para garantia da qualidade relacional de dados e aplicá-la a transações bancárias reais utilizadas na análise de tipologias de lavagem de dinheiro pelo Laboratório de Tecnologia Contra a Lavagem de Dinheiro (LABLD) da Polícia Civil de Pernambuco (PCPE).

Os **objetivos específicos** são:

1. Desenvolver uma metodologia baseada em teoria dos grafos para caracterização da qualidade relacional de dados;
2. Aplicar e validar a metodologia em dados anonimizados de transações bancárias fornecidos pelo LABLD da PCPE;
3. Investigar o impacto da qualidade relacional dos dados no desempenho de modelos preditivos baseados em redes neurais recorrentes para análise de tipologias de lavagem de dinheiro.

#### **Referências Bibliográficas:**

ANAROCH, Chayawat; TOSAKULVONG, Waruntorn; WATTANAPONGSAKORN, Naruemon. Anti-Money Laundering Detection Using Traditional, Deep, and Hybrid Machine

Learning: A Performance Comparison. en. *In: 2025 6th International Conference on Big Data Analytics and Practices (IBDAP)*. Chiang Mai, Thailand: IEEE, ago. 2025. p. 239–244.

BACEN. **CIRCULAR N 4.001, de 29 de Janeiro de 2020**. [S. l.: s. n.], 2020.

BRASIL. **LEI N 9.613, de 3 de Março de 1998**. [S. l.: s. n.], 1998.

GIRISH, K K; BHOWMIK, Biswajit. Money Laundering Detection in Banking Transactions using RNNs and Hybrid Ensemble. en. *In: 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. Kamand, India: IEEE, jun. 2024. p. 1–7.

GROSS, Jonathan L; YELLEN, Jay; ANDERSON, Mark. **Graph theory and its applications**. [S. l.]: Chapman e Hall/CRC, 2018.

HENRIQUE S. ASSUMPÇÃO *et al.* Delator: Money laundering detection via multi-task learning on large transaction graphs. *In: 2022 IEEE International Conference on Big Data (Big Data)*. Osaka, Japan: IEEE, dez. 2022.

JENSEN, Rasmus Ingemann Tuffveson; IOSIFIDIS, Alexandros. Fighting Money Laundering With Statistics and Machine Learning. en. **IEEE Access**, v. 11, p. 8889–8903, 2023.

JUNGNICKEL, Dieter. **Graphs, networks and algorithms**. [S. l.]: Springer.

LIU, Qi. Identification of Financial Fraud in Enterprises using Mish Activation Function-Bidirectional Gated Recurrent Unit. en. *In: 2024 International Conference on Data Science and Network Security (ICDSNS)*. Tiptur, India: IEEE, jul. 2024. p. 1–4.

NUGROHO, Heru. A Review: Data Quality Problem in Predictive Analytics. en. **IJAIT (International Journal of Applied Information Technology)**, v. 7, n. 02, p. 79, ago. 2023.

VENKATA LAKSHMI NARAYANA, G; PANIGRAHI, Suvasini. Developed Residual RNN with Temporal Attention for Efficient Fraudulent Activities Detection in Financial Industry using SVM Features. en. *In: 2025 IEEE International Conference on Contemporary Computing and Communications (InC4)*. Bangalore, India: IEEE, mar. 2025. p. 1–6.

WEST, Douglas Brent *et al.* **Introduction to graph theory**. [S. l.]: Prentice hall Upper Saddle River, 2001. v. 2.

WONG, Ka Yee; WONG, Raymond K. Big data quality prediction informed by banking regulation. en. **International Journal of Data Science and Analytics**, v. 12, n. 2, p. 147–164, ago. 2021.

XAVIER, Isadora *et al.* Towards the Identification of Money Laundering in Bank Transactions using Recurrent Neural Networks. en. *In: 2024 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. Bogota D.C., Colombia: IEEE, nov. 2024. p. 1–5.