

Inspere - Instituto de Ensino e Pesquisa
Mestrado Profissional em Administração
Ricardo Ferreira Lima

Machine Learning Baseado na Combinação de Fatores Heurísticos e Racionais no Auxílio da Prevenção à Lavagem de Dinheiro

São Paulo
2025

Ricardo Ferreira Lima

Machine Learning Baseado na Combinação de Fatores Heurísticos e Racionais no Auxílio da Prevenção à Lavagem de Dinheiro

Dissertação apresentada ao Programa de Mestrado Profissional em Administração do Insper Instituto de Ensino e Pesquisa, como requisito parcial para a obtenção do título de Mestre em Administração.

Orientadora: Profa. Adriana Bruscato Bortoluzzo
Coorientadora: Profa. Carla Sofia Dias Moreira Ramos

São Paulo

2025

Ricardo Ferreira Lima

Lima, Ricardo Ferreira

Machine learning baseado na combinação de fatores heurísticos e racionais no auxílio da prevenção à lavagem de dinheiro

Ricardo Ferreira Lima. – São Paulo, 2025.
52 f.

Dissertação (Mestrado Profissional em Administração) – Insper Instituto de Ensino e Pesquisa, 2025

Orientadora: Profa. Dra. Adriana Bruscatto Bortoluzzo
Coorientadora: Profa. Dra. Carla Sofia Dias Moreira Ramos

1. Prevenção de lavagem de dinheiro. 2. Machine Learning. 3. Fatores heurísticos. 4. Fatores racionais. I. Lima, Ricardo Ferreira. II. Machine learning baseado na combinação de fatores heurísticos e racionais no auxílio da prevenção à lavagem de dinheiro

Machine Learning Baseado na Combinação de Fatores Heurísticos e Racionais no Auxílio da Prevenção à Lavagem de Dinheiro

Dissertação apresentada ao Programa de Mestrado Profissional em Administração do Insper Instituto de Ensino e Pesquisa, como requisito parcial para a obtenção do título de Mestre em Administração.

Orientadora: Profa. Adriana Bruscato Bortoluzzo
Coorientadora: Profa. Carla Sofia Dias Moreira Ramos

Banca Examinadora

Orientadora: Profa. Dra. Adriana Bruscato Bortoluzzo

Coorientadora: Profa. Dra. Carla Sofia Dias Moreira Ramos

Dr. Gerson Luis Romantini

Banco Central do Brasil - Departamento de Supervisão de Conduta (DECON)

Prof. Dr. Rinaldo Artes

Insper

São Paulo

2025

AGRADECIMENTOS

Em primeiro lugar, agradeço à minha esposa Priscilla, por sua paciência, compreensão e apoio incondicional em todos os momentos desta jornada. Sua presença constante e encorajadora foi essencial para que eu mantivesse o foco e a motivação até a conclusão dessa dissertação.

À minha orientadora, Profa. Dra. Adriana Bruscato Bortoluzzo, expressei minha profunda gratidão pela orientação firme, generosa e sempre inspiradora. Seu conhecimento, dedicação e olhar crítico foram determinantes para a qualidade deste trabalho.

À minha coorientadora, Profa. Dra. Carla Sofia Dias Moreira Ramos, agradeço pela parceria, pelas contribuições valiosas e pelo apoio ao longo de todo o desenvolvimento desta dissertação. Sua experiência e sensibilidade ampliaram as fronteiras deste estudo.

Aos membros da banca, Dr. Gerson Luis Romantini, do Banco Central do Brasil – Departamento de Supervisão de Conduta (DECON), e Prof. Dr. Rinaldo Artes, do Insper, agradeço pelas observações pertinentes, pelo olhar técnico e pelo rigor acadêmico, que enriqueceram significativamente esta dissertação.

Estendo também meus agradecimentos aos professores, colegas do mestrado, especialistas entrevistados, à equipe do Insper e a todos que, de alguma forma, contribuíram para que este projeto se tornasse realidade.

RESUMO

A lavagem de dinheiro representa desafios significativos para a estabilidade econômica global, exigindo que instituições financeiras adotem métodos mais eficazes de detecção desses atos ilícitos. Este estudo propõe uma abordagem com uso de modelo de Machine Learning (ML) *Random Forest* que combina fatores heurísticos (regras de bolso dos especialistas) e fatores racionais (regras duras das instituições financeiras) para aprimorar a identificação de transações suspeitas e reduzir o índice de falsos positivos. Utilizando um método misto sequencial, a fase qualitativa identificou fatores heurísticos por meio de entrevistas com especialistas em PLD, enquanto a fase quantitativa integrou esses fatores a modelos de ML, combinando-os com regras duras previamente testadas. Os resultados de aplicação de modelos demonstraram que a abordagem híbrida foi eficaz em detectar os casos de falsos positivos, o que reduziu em 24% a quantidade de falso positivo e melhorou o processo operacional de detecção de risco. Dessa forma, o estudo contribui tanto para a teoria, ao integrar abordagens qualitativas e quantitativas na detecção de crimes financeiros, quanto para a prática, ao oferecer um modelo mais eficiente para instituições financeiras, permitindo maior precisão na geração de alertas e otimização dos processos de monitoramento.

Palavras-chave: Prevenção de lavagem de dinheiro (PLD), Machine learning, Fatores Heurísticos, Fatores Racionais.

ABSTRACT

Money laundering poses significant challenges to global economic stability, requiring financial institutions to adopt more effective detection methods. This study proposes a Machine Learning (ML) approach using the *Random Forest* algorithm, which integrates heuristic factors (experts' tacit rules) and rational factors (institutional hard rules) to enhance the identification of suspicious transactions and reduce false positives. Adopting a sequential mixed-methods design, the qualitative phase identified heuristic elements through interviews with AML specialists, while the quantitative phase incorporated these elements into ML models alongside validated hard rules. The results demonstrated that the hybrid approach was effective in reducing false positives by 24% and improving the operational risk detection process. Thus, this study contributes to theory by integrating qualitative and quantitative strategies in financial crime detection, and to practice by offering a more efficient model for financial institutions, enhancing alert precision and optimizing monitoring procedures.

Keywords: Anti-money laundering (AML), Machine Learning, Heuristic Factors, Rational Factors.

SUMÁRIO EXECUTIVO

A lavagem de dinheiro é um desafio persistente para o sistema financeiro global. Esse processo permite que recursos oriundos de atividades ilícitas sejam inseridos no mercado formal, dificultando a identificação de crimes como corrupção, tráfico e fraudes. No Brasil, o volume de transações financeiras classificadas como suspeitas cresce ano após ano, pressionando as instituições financeiras a aprimorarem seus mecanismos de controle e detecção. A legislação brasileira estabelece que os bancos devem reportar operações suspeitas ao Conselho de Controle de Atividades Financeiras (COAF), o que resulta em milhões de alertas automáticos por ano. Contudo, a grande maioria desses alertas são falsos positivos, ou seja, transações que não configuram, de fato, indícios de crime, o que gera retrabalho, desperdício de recursos e atrasos nas investigações mais relevantes.

Atualmente esse processo são baseadas em regras duras — critérios objetivos e padronizados definidos pelas instituições financeiras. Essas regras são rígidas e, sozinhas, têm se mostrado pouco eficazes para lidar com a complexidade e diversidade de comportamentos financeiros atípicos. Em contrapartida, analistas costumam usar também “”, ou fatores heurísticos, ou seja, critérios informais baseados em experiência prática, que ajudam a filtrar casos com maior potencial de risco. Essas heurísticas, embora eficazes na prática, raramente são incorporadas de forma estruturada aos sistemas automatizados.

Essa dissertação propõe combinar regras racionais com fatores heurísticos dentro de um modelo de Machine Learning. O objetivo é tornar a geração de alertas mais eficiente e reduzir os falsos positivos, permitindo que os especialistas foquem nos casos mais críticos. Para isso, foi utilizado um algoritmo de aprendizado de máquina (*Random Forest*), ao qual foram incorporadas tanto as variáveis tradicionais (as regras duras) quanto variáveis obtidas a partir de entrevistas com especialistas em Prevenção à Lavagem de Dinheiro (PLD), que indicaram os padrões informais usados no dia a dia.

O modelo foi testado com dados de uma instituição financeira. As análises demonstraram que, ao incorporar as heurísticas dos especialistas, o algoritmo se tornou significativamente mais preciso. A principal evidência disso foi uma redução de 24% no número de falsos positivos. Esse resultado não apenas melhora a eficácia do processo, mas representa também economia de tempo e recursos para as instituições, além de potencialmente acelerar a detecção de crimes financeiros de fato.

Do ponto de vista prático, a principal recomendação do trabalho é que instituições financeiras invistam na integração da expertise humana com sistemas de machine learning, em vez de tratar esses dois universos como separados. Para isso, é necessário estruturar melhor o conhecimento tácito dos especialistas e traduzi-lo em dados que possam ser utilizados por modelos automatizados. Como cuidado, é importante garantir que o processo de coleta desse conhecimento seja contínuo e colaborativo, respeitando as peculiaridades de cada instituição e mantendo alinhamento com as normativas regulatórias.

A abordagem híbrida testada neste estudo oferece um caminho promissor para aprimorar os sistemas de combate à lavagem de dinheiro, equilibrando inteligência humana e artificial. O modelo pode ser replicado por outras instituições e adaptado para diferentes contextos, desde que respeitadas as especificidades das operações e os aspectos éticos e regulatórios envolvidos. Trata-se, portanto, de uma contribuição relevante tanto para a academia quanto para o setor financeiro.

Lista de Ilustrações e Tabelas

Figura 1 - Fluxo de monitoramento de transações, geração de alertas e tratamento.....	16
Quadro 1 - Perfil dos entrevistados.....	23
Quadro 2 - Achados das entrevistas.....	30
Tabela 1 - Relação do Parecer dos Analistas e Regras Duras.....	34
Tabela 2 - Relação do Parecer dos Analistas e Fatores identificados na fase qualitativa .	35
Tabela 3 - Relação do Parecer dos Analistas e variáveis contínuas.....	37
Tabela 4 – Estatística de Comparação do Modelos.....	38
Tabela 5 – Tabela de importância das variáveis por modelo.....	39

Sumário

1. INTRODUÇÃO	12
2. PREVENÇÃO À LAVAGEM DE DINHEIRO	15
2.1 PROCESSO DE MONITORAMENTO	15
2.2 UTILIZAÇÃO DE MODELOS DE <i>MACHINE LEARNING</i>	18
2.3 COMBINAÇÃO DE FATORES RACIONAIS E HEURÍSTICOS	20
3. METODOLOGIA E DESIGN DE PESQUISA	22
3.1 FASE 1 - MÉTODO QUALITATIVO	22
3.1.1 SELEÇÃO DE PARTICIPANTES	23
3.1.2 TÉCNICA DE COLETA DE DADOS	24
3.1.3 TÉCNICA DE ANÁLISE DE DADOS	25
3.2 FASE 2 - MÉTODO QUANTITATIVO	26
3.2.1 AMOSTRAGEM	26
3.2.2 TÉCNICA DE COLETA DE DADOS	27
3.2.3 TÉCNICA DE ANÁLISE DE DADOS	27
4. RESULTADOS	29
4.1 FASE 1- PESQUISA QUALITATIVA	29
4.2 FASE 2 – PESQUISA QUANTITATIVA	33
4.2.1 PRIMEIRA ETAPA: IDENTIFICAÇÃO DE FATORES RACIONAIS MAIS EFICAZES	33
4.2.2 SEGUNDA ETAPA: INTEGRAÇÃO DE FATORES HEURÍSTICOS E RACIONAIS EM ML <i>RANDOM FOREST</i>	35
4.3 AVALIAÇÃO DOS MODELOS APLICADOS NA GERAÇÃO DE ALERTAS	37
5. DISCUSSÃO	40
6. CONCLUSÕES	41
REFERÊNCIAS	44
GLOSSÁRIO	50
ANEXO 1 - TERMO DE CONSENTIMENTO PARA PARTICIPAÇÃO EM PESQUISA	51
ANEXO 2 - CENÁRIOS PARA AS ENTREVISTAS E ROTEIRO	52

1. INTRODUÇÃO

O aumento dos crimes financeiros representa uma ameaça significativa à estabilidade global, comprometendo a integridade dos sistemas financeiros (Jullum et al., 2020). Atividades ilícitas como fraude e tráfico de drogas impulsionam a lavagem de dinheiro, permitindo a movimentação e integração de recursos de origem criminosa no sistema financeiro formal (Alnasser, 2020). No Brasil, em 2023, o volume de operações suspeitas de lavagem de dinheiro permaneceu elevado, refletindo-se no crescimento das Comunicações de Operações Suspeitas (COS)¹ enviadas ao Conselho de Controle de Atividades Financeiras (COAF). Entre 2018 e 2022, foram registradas 4,8 milhões de comunicações, com uma média anual de 971.383, representando 0,00193% do total de 249,24 bilhões de transações financeiras. Além disso, o volume de transações apresentou um crescimento expressivo, com um aumento de 248% no primeiro trimestre de 2023 em comparação ao mesmo período de 2019, segundo o Banco Central do Brasil. Esse cenário contribuiu diretamente para a elevação do número de COS, que atingiu 1,1 milhão em 2023.

Adicionalmente, crimes como corrupção e organização criminosa seguem impulsionando a lavagem de dinheiro no país. Em 2022, foram registrados 873 casos relacionados a essas práticas, correspondendo a 8,06% de todas as investigações conduzidas pela Polícia Federal. Embora esse número possa parecer reduzido, sua relevância se evidencia ao considerar seu impacto na efetividade das investigações formais e na capacidade de mitigar crimes financeiros (FATF/OECD - GAFILAT, 2023).

Para cumprir os requisitos regulatórios, as instituições financeiras devem monitorar continuamente seus clientes, abrangendo não apenas as transações de entrada e saída de recursos em suas contas, mas também dados cadastrais e comportamentais, como propostas e questionamentos realizados. Este estudo tem como foco específico o monitoramento com regras transacionais de conta corrente, considerando sua relevância no contexto da prevenção à lavagem de dinheiro (PLD). A Lei nº 9.613/1998 estabelece diretrizes para PLD, exigindo que as instituições financeiras adotem políticas e controles internos eficazes. Complementando essa legislação, a Circular nº 3.978/2020 e a Carta Circular nº 4.001/2020, ambas do Banco Central, detalham exigências específicas para o cumprimento dessas diretrizes. Essas normas determinam a necessidade de

¹ Conforme definição do COAF, Comunicação de Operação Suspeita (COS) é a comunicação feita por instituições obrigadas sempre que forem identificadas operações que possam configurar indícios de lavagem de dinheiro ou financiamento ao terrorismo. Essa obrigação está prevista no art. 11, inciso II, alínea "b" da Lei nº 9.613, de 3 de março de 1998, que dispõe sobre os crimes de lavagem de dinheiro, e é regulamentada por normativos específicos conforme o setor regulado.

monitoramento contínuo e a comunicação obrigatória de operações suspeitas ao COAF, garantindo maior transparência e integridade ao sistema financeiro.

O monitoramento ocorre formalmente com base em regras duras, estabelecidas por normas das autoridades reguladoras, e não diretamente pela legislação. Essas diretrizes orientam os "agentes econômicos obrigados" na implementação de mecanismos para identificar e comunicar operações suspeitas ao COAF. Com base nessas comunicações, as autoridades competentes podem conduzir investigações para verificar a existência de atos ilícitos. Após a aplicação dessas regras duras, os casos identificados nos critérios estabelecidos são classificados como "alertas" e encaminhados a um grupo de especialistas para uma avaliação detalhada, a fim de validar quais casos apresentam indícios suficientes de suspeição e quais são falsos positivos. Os casos classificados como suspeitos devem ser comunicados ao COAF, que dará prosseguimento ao tratamento e poderá gerar um Relatório de Investigação Financeira (RIF) para o Ministério Público e Polícia Federal, órgãos responsáveis pela persecução criminal.

Diante desse contexto, é crucial avaliar formas mais eficazes para determinar quais alertas têm maior probabilidade de possuir os elementos de suspeição suficientes para embasar uma comunicação ao COAF, com o objetivo de otimizar o processo operacional e reduzir o índice de falsos positivos, sem comprometer a gestão de risco. Alnasser (2020) sugere que os bancos têm um papel crucial nessa prevenção, enquanto Canhoto (2020) e Ferwerda et al. (2020) exploram os benefícios potenciais modelos de Machine Learning (ML) na identificação e prevenção de operações ilícitas. Para aumentar a eficiência no processo de detecção de atipicidades, o uso de regras de bolso – ou fatores heurísticos – por especialistas de PLD emerge como uma abordagem complementar. Esses fatores heurísticos aplicados informalmente pelos especialistas de PLD permitem priorizar casos com maior potencial de risco, concentrando os esforços naqueles que realmente necessitam de mais atenção. Conforme discutido por Slovic et al. (2004) e Damasio (1994), regras de bolso funcionam como mecanismos intuitivos baseados em experiência e associações emocionais, otimizando a tomada de decisão em cenários complexos ou incertos. Essa metodologia heurística não só acelera a identificação de situações críticas, mas também contribui para a melhoria do tempo de resposta e da eficácia na avaliação.

Esta dissertação tem por objetivo avaliar se o uso de modelos de Machine Learning (ML) *Random Forest* que combina regras de bolso (i.e., fatores heurísticos definidos pelos especialistas de PLD) e regras duras (i.e., fatores racionais definidos pelas autoridades reguladoras) aprimora a detecção de lavagem de dinheiro. Para responder à questão de

pesquisa, foi adotado o método misto sequencial: na primeira fase, foi aplicado o método qualitativo com base em entrevistas com especialistas em Prevenção à Lavagem de Dinheiro (PLD), para identificar os fatores heurísticos utilizados por esses especialistas no processo de avaliação de risco, e que pudessem ser mensurados quantitativamente. Em seguida, a fase quantitativa identificou em uma primeira etapa, com base na análise de uma base de dados de uma instituição financeira quais os fatores racionais ou regras duras usadas na identificação de casos suspeitos. Em uma segunda etapa da pesquisa quantitativa, foi aplicado o modelo *Random Forest* a dados dessa instituição financeira, integrando os fatores heurísticos e os fatores racionais identificados como mais eficazes, com vista a otimizar a identificação de atividades suspeitas.

Esse estudo contribui com a teoria enriquecendo a literatura acadêmica sobre o uso de ML na prevenção a crimes financeiros (Canhoto, 2020; Ferwerda et al., 2020). A inclusão dessas tecnologias no estudo de lavagem de dinheiro oferece uma nova perspectiva e aprofunda o entendimento da eficácia dessas ferramentas (Jullum et al., 2020). Adicionalmente, esta dissertação explora a combinação em modelos de aprendizado de máquina de fatores heurísticos, como a intuição dos especialistas que simplificam decisões em cenários complexos (Tversky; Kahneman, 1974), com fatores racionais, destacando como essa integração pode otimizar a tomada de decisão na detecção de atividades suspeitas. Para a prática, o estudo visa aprimorar os mecanismos de detecção de lavagem de dinheiro em instituições financeiras, através da combinação de metodologias qualitativas e quantitativas para identificar fatores críticos. Isso pode resultar em uma redução significativa de falsos positivos e no aumento da precisão na identificação de atividades suspeitas, como destacado por Chen et al. (2018) e Singh & Best (2019). Além do mais, a aplicação dessas metodologias pode conduzir ao desenvolvimento de procedimentos mais robustos e ágeis, conforme sugerido por Jullum et al. (2020).

Para a prática, o estudo visa aprimorar os mecanismos de detecção de lavagem de dinheiro em instituições financeiras, através da combinação de metodologias qualitativas e quantitativas para identificar fatores críticos. Isso pode resultar em uma redução significativa de falsos positivos e no aumento da precisão na identificação de atividades suspeitas, como destacado por Chen et al. (2018) e Singh & Best (2019). Além do mais, a aplicação dessas metodologias pode conduzir ao desenvolvimento de procedimentos mais robustos e ágeis, conforme sugerido por Jullum et al. (2020).

Este trabalho está organizado da seguinte forma. Após a introdução apresentada neste capítulo, o próximo capítulo trata da revisão da literatura, iniciando com uma análise

do processo de monitoramento e sua relevância na PLD. Em seguida, são discutidas as aplicações de técnicas de *Machine Learning* no setor, culminando com a discussão sobre a integração de fatores heurísticos e racionais nesse contexto. Nesse trecho final, também é apresentada a hipótese de pesquisa, fundamentada nos conceitos explorados ao longo da revisão. A metodologia é detalhada com base em uma abordagem de métodos mistos, abrangendo fases qualitativa e quantitativa. Por fim, os resultados obtidos são apresentados, seguidos pelas principais contribuições esperadas para a teoria e prática, e pelas limitações deste trabalho e sugestões de pesquisas futuras.

2. Prevenção à Lavagem de Dinheiro

2.1 Processo de monitoramento

Os crimes financeiros representam uma ameaça à estabilidade econômica global, comprometendo a integridade dos sistemas financeiros e a segurança das nações (FATF/OECD - GAFILAT, 2023). Dentre esses crimes, a lavagem de dinheiro se destaca por permitir a movimentação e integração de recursos provenientes de atividades ilícitas no sistema financeiro formal. Globalmente, a lavagem de dinheiro são um problema massivo, estimado em aproximadamente 2% a 5% do PIB mundial, o que equivale a valores entre US\$ 800 bilhões e US\$ 2 trilhões anualmente (UNODC, 2011; Walker, 1999). O impacto desse crime é devastador, pois compromete a estabilidade das instituições financeiras, facilita o financiamento de atividades criminosas e terroristas, e mina a confiança pública nos sistemas econômicos. Além disso, a circulação de grandes somas de dinheiro sujo distorce os mercados, cria uma competição desleal e dificulta a execução de políticas econômicas eficazes (Schott, 2006).

O processo de lavagem de dinheiro ocorre tipicamente em três etapas distintas, nomeadamente: (i) colocação, onde os recursos ilícitos são inseridos no sistema financeiro; (ii) ocultação, fase em que os recursos são movimentados através de múltiplas transações para dificultar o rastreamento de sua origem; e (iii) integração, momento em que os recursos são formalmente incorporados ao sistema financeiro, disfarçados como fundos legítimos (Walker, 1999). A prevenção desses crimes envolve uma complexa rede de monitoramento, que começa com a detecção inicial de transações suspeitas pelas instituições financeiras e outras entidades obrigadas conforme determinando na legislação vigente e se estende até a investigação aprofundada pelas autoridades competentes, incluindo agências reguladoras e órgãos de aplicação da lei (Schott, 2006).

A Figura 1 ilustra o fluxo de monitoramento de transações suspeitas desde o momento em que a transação é realizada pelo cliente, passando pela aplicação de regras de detecção (fatores racionais ou regras duras). Essas regras podem ser baseadas em padrões comportamentais ou em limites específicos, estabelecidos para identificar anomalias em transações, seguindo as regras estabelecidas nas normas vigentes, mas não se limitando a esses cenários. Esses limites podem incluir grandes somas de dinheiro movimentadas repentinamente ou transações que não condizem com o perfil econômico do cliente (Chen et al., 2018).

Figura 1 – Fluxo de monitoramento de transações, geração de alertas e tratamento



Fonte: Elaborado pelo autor.

Conforme definido no artigo 1º da Carta Circular Nº 4.001, de 29 de janeiro de 2020, as operações ou situações que exemplificam indícios de suspeita incluem, entre outras, operações que aparentem não ter fundamento econômico ou legal, transações que demonstrem um comportamento atípico em relação ao histórico conhecido do cliente, e aquelas que envolvam países ou territórios considerados de risco elevado, como os incluídos em listas de países não cooperantes ou com deficiências estratégicas em regimes de prevenção à lavagem de dinheiro. Além disso, essas regras devem ser constantemente revisadas e atualizadas para acompanhar as novas estratégias utilizadas por criminosos na lavagem de dinheiro, garantindo que o sistema permaneça eficaz e relevante (Banco Central do Brasil, 2020).

Esse fluxo de monitoramento, com a captura inicial de transações e a aplicação de regras duras para a detecção de anomalias, é amplamente utilizado e reconhecido em sistemas de PLD. Segundo Singh e Best (2019), o uso de regras predefinidas baseadas em padrões de comportamento e limites financeiros é um dos pilares para a identificação de atividades suspeitas em ambientes financeiros, proporcionando uma primeira linha de defesa contra o crime financeiro. Após a aplicação dessas regras, é gerada uma base com a seleção de casos suspeitos (alertas), que são então analisados por especialistas em PLD.

Este processo serve para garantir que as instituições financeiras não facilitem, inadvertidamente, a lavagem de dinheiro (Demetis, 2018). Esta análise humana é crucial, pois envolve a revisão manual por especialistas em lavagem de dinheiro, que utilizam ferramentas avançadas de visualização de dados que servem de apoio à decisão para avaliar a legitimidade das transações. Na última etapa do fluxo, é realizada a conclusão da avaliação, com a deliberação sobre os casos, classificando-os como: (i) não suspeito (falso positivo) ou (ii) suspeito (verdadeiro positivo).

No contexto de detecção de crimes financeiros, especialmente na lavagem de dinheiro, a distinção entre falsos positivos e verdadeiros positivos é crucial para a eficiência do sistema de monitoramento. Um falso positivo ocorre quando uma transação legítima é incorretamente selecionada como um alerta, resultando em análises desnecessárias. Isso não apenas consome recursos, mas também pode levar a uma "fadiga de alerta", onde a eficácia do monitoramento diminui devido ao grande volume de falsos alarmes (Jullum et al., 2020). Estudos indicam que, em média, as instituições financeiras lidam com uma taxa de falsos positivos que pode superar 90%, o que evidencia a necessidade de métodos mais refinados de análise (Canhoto, 2020). Por outro lado, um falso negativo acontece quando uma transação ilícita não é objeto de alerta, permitindo que atividades criminosas continuem sem o devido tratamento. Este cenário é particularmente perigoso, pois compromete a integridade do sistema financeiro e a eficácia das medidas de PLD (Canhoto, 2020).

No Brasil, a aplicação de sistemas de monitoramento de PLD tem avançado, especialmente com o uso de tecnologias como modelos de aprendizado de máquina (*Machine Learning* - ML) e a inteligência artificial (IA). Essas tecnologias apresentam um potencial significativo para transformar a maneira como as instituições financeiras identificam e previnem a lavagem de dinheiro, aplicando técnicas como redes neurais e algoritmos de aprendizado profundo para detectar padrões anômalos de forma mais eficaz (Chen et al., 2018). Quando bem implementados, esses sistemas conseguem melhorar os procedimentos de geração de alertas, reduzindo tanto os falsos positivos quanto os falsos negativos (Chen et al., 2018).

Apesar dos avanços, as instituições financeiras ainda enfrentam desafios significativos, como a necessidade constante de atualizar regras e algoritmos para acompanhar as novas técnicas empregadas por criminosos (Demetis, 2018). Para garantir a eficiência e sustentabilidade do sistema, a literatura recomenda a implementação de um sistema de avaliação contínua de desempenho, onde métricas como precisão, *recall*, e a taxa de falsos positivos/negativos sejam monitoradas (Jullum et al., 2020). Essa prática

permite que ajustes sejam feitos nas regras conforme necessário, assegurando que as instituições financeiras continuem a aprimorar sua capacidade de detecção e resposta a transações ilícitas, especialmente em um ambiente financeiro cada vez mais dinâmico e complexo (Ferwerda et al., 2020). Essas necessidades reforçam a importância de avaliar o uso de técnicas avançadas como o ML no aprimoramento dos processos de monitoramento. Como apontado por Canhoto (2020), a implementação dessas tecnologias nas instituições financeiras, é essencial para enfrentar os desafios de um ambiente financeiro em constante evolução.

2.2 Utilização de Modelos de *Machine Learning*

Conforme dados do Banco Central do Brasil, o volume de transações financeiras aumentou em 248% no primeiro trimestre de 2023 em comparação ao mesmo período de 2019. Esse crescimento elevou significativamente a geração de alertas, ressaltando a necessidade de modelos preditivos baseados em *Machine Learning* (ML) para gerenciar o desafio de falsos positivos e melhorar a detecção de indícios de lavagem de dinheiro. Esses modelos têm se mostrado eficazes ao identificar transações suspeitas com maior precisão, permitindo que os recursos de monitoramento sejam direcionados às transações que realmente demandam maior escrutínio (Jullum et al., 2020).

Modelos de ML têm demonstrado eficiência na detecção de lavagem de dinheiro, especialmente na redução de falsos positivos, que representam um desafio significativo para o setor financeiro. Estudos como o de Bakry et al. (2024) destacam que a aplicação de frameworks baseados em ML, como *Random Forest* tem potencial para equilibrar a identificação de atividades suspeitas com a redução de alertas indevidos, aumentando a confiabilidade dos processos de monitoramento. Além disso, Kute et al. (2021) enfatizam que essas tecnologias tornam os sistemas mais precisos e robustos, mas destacam que ferramentas explicativas são essenciais para garantir transparência e aceitação por especialistas e reguladores.

Outro ponto relevante é a integração entre os modelos de ML e o conhecimento prático dos especialistas em PLD. Essa combinação não apenas enriquece os sistemas de monitoramento, como também fortalece sua adaptabilidade diante do crescente volume de transações financeiras. Jullum et al. (2020) evidenciam a eficácia de métodos supervisionados aplicados a grandes bases de dados, destacando sua contribuição para a redução de falsos positivos e a identificação de padrões anômalos com maior precisão.

Adicionalmente, técnicas de aprendizado contínuo (*continuous learning*) permitem que os modelos de ML sejam constantemente atualizados, ajustando-se a novos padrões de comportamento sem necessidade de reestruturações completas. Essa capacidade adaptativa mantém os sistemas responsivos e preparados para lidar com ameaças emergentes (Parisi et al., 2019). Além disso, a incorporação de múltiplas fontes de dados, como registros de inteligência financeira, redes sociais e perfis de clientes (*Know Your Customer - KYC*), amplia o contexto analítico dos modelos, aprimorando a identificação de atividades suspeitas e reduzindo os falsos positivos (Chen et al., 2018).

Um aspecto importante de se observar é que garantir que os modelos de ML sejam transparentes e auditáveis é crucial para que se tenha clareza e entendimento sobre como os resultados são alcançados, promovendo maior robustez e confiança tanto para os reguladores quanto para as instituições financeiras que adotam esses sistemas. Técnicas como as descritas por Doshi-Velez e Kim (2017), incluindo métodos de *Explainable Artificial Intelligence (XAI)* como *Local Interpretable Model-Agnostic Explanations (LIME)* e *SHapley Additive exPlanations (SHAP)*, têm ajudado a interpretar algoritmos considerados "caixas-pretas". Essas abordagens não apenas aumentam a conformidade regulatória, mas também oferecem maior clareza para a tomada de decisões automatizadas.

Apesar das vantagens oferecidas por modelos de ML, desafios significativos persistem. A qualidade e a disponibilidade dos dados continuam sendo questões críticas para o desenvolvimento de modelos robustos. Bakry et al. (2024) sugerem que o uso de dados reais combinados com abordagens de aprendizado não supervisionado pode aprimorar a identificação de atividades suspeitas, especialmente no tratamento do desbalanceamento de classes, um problema recorrente em sistemas de detecção. Além disso, é necessário assegurar que os modelos preditivos sejam suficientemente robustos para minimizar tanto os falsos positivos quanto os falsos negativos, garantindo assim sua eficácia operacional.

Por fim, a implementação de modelos de ML pode enfrentar barreiras organizacionais significativas. Essa adoção não envolve apenas aspectos tecnológicos, mas também exige reestruturações operacionais, treinamento contínuo das equipes e aprimoramento contínuo dos modelos. Brock e Von Wangenheim (2019) destacam que a resistência à mudança e a necessidade de alinhar novas tecnologias aos sistemas legados podem retardar sua adoção. Portanto, estratégias bem estruturadas de integração e treinamento são essenciais para superar esses desafios e maximizar os benefícios dos modelos de ML em PLD.

2.3 Combinação de Fatores Racionais e Heurísticos

A teoria da tomada de decisão, amplamente discutida por autores como Kahneman e Tversky (1974), destaca que os métodos heurísticos são fundamentados em análises sistemáticas de dados, permitindo decisões detalhadas e baseadas em evidências. Em contrapartida, as heurísticas, como descrito por Gigerenzer e Gaissmaier (2011), são estratégias cognitivas que simplificam o processo decisório ao focar em elementos-chave, ignorando informações secundárias. Essas abordagens complementares integram agilidade intuitiva e rigor analítico, otimizando a tomada de decisões em ambientes de alta complexidade.

Assim, com base na literatura e no contexto da PLD, espera-se que a combinação de fatores heurísticos e fatores racionais possibilitem que instituições financeiras equilibrem a rapidez da intuição (i.e., fatores heurísticos) com a profundidade analítica oferecida pelos métodos racionais. Enquanto os métodos racionais realizam análises quantitativas detalhadas com base em grandes volumes de dados, os métodos heurísticos são aplicados informalmente, oferecem agilidade para identificar padrões suspeitos e permitem que especialistas priorizem casos críticos e direcionem seu tempo de forma mais eficiente (Jullum et al., 2020). Espera-se que esta abordagem híbrida enriqueça a detecção de atividades suspeitas, otimizando os processos de monitoramento e aumentando sua eficácia na prevenção a crimes financeiros.

As heurísticas podem ser ajustadas para contextos específicos, aumentando sua eficiência em ambientes dinâmicos como é um caso do setor financeiro, onde os padrões de comportamento criminoso estão em constante evolução. Gigerenzer e Gaissmaier (2011) descrevem métodos heurísticos como estratégias cognitivas que simplificam e agilizam a tomada de decisões ao ignorar partes da informação, sendo particularmente úteis em cenários de incerteza e alta complexidade. Kahneman e Tversky (1974) apontam que julgamentos sob incerteza frequentemente utilizam fatores heurísticos, como a da representatividade, que permite aos especialistas comparar padrões percebidos com experiências anteriores. A integração dessas abordagens heurísticas com métodos racionais fortalece a capacidade de tomada de decisão, combinando rapidez com precisão analítica, o que é essencial para melhorar os sistemas de detecção de lavagem de dinheiro. Gigerenzer (2008) argumenta que métodos heurísticos simplificam a complexidade inicial, enquanto métodos racionais fornecem a profundidade necessária para validar suspeitas, criando um equilíbrio entre agilidade e robustez.

No contexto específico de PLD, a aplicação prática dessa integração pode ser observada no uso de métodos heurísticos pelos especialistas para reduzir a complexidade inicial, filtrando transações que demandam maior atenção. Posteriormente, métodos racionais são empregados para aprofundar a análise das transações selecionadas, validando suspeitas com base em evidências quantitativas. Essa sinergia é particularmente útil em contextos de alto volume transacional, como o financeiro, onde ameaças emergentes requerem soluções ágeis e adaptativas (Alnasser, 2020).

Além disso, fatores heurísticos podem evoluir sucessivamente por meio de aprendizado contínuo, o que permite que os sistemas ajustem seus critérios de detecção com base em *feedbacks* de especialistas e novos padrões de atividade criminosa. Parisi et al. (2019) destacam que o aprendizado contínuo aprimora os modelos de detecção ao incorporar *feedbacks* iterativos e dados atualizados, aumentando sua adaptabilidade. Essa abordagem é essencial, especialmente frente à evolução constante das práticas de lavagem de dinheiro, como também reforçado por Alnasser (2020). Assim, métodos heurísticos não devem ser vistos como estáticos, mas como componentes dinâmicos que acompanham mudanças no cenário criminoso.

Diante desse contexto, esta pesquisa propõe a hipótese de que a combinação em modelos de ML de fatores heurísticos, identificados junto de especialistas em PLD, e fatores racionais, pode aumentar a eficiência na distinção entre atividades legítimas e suspeitas, reduzindo a taxa de falsos positivos na detecção de lavagem e dinheiro. Essa abordagem visa alavancar o conhecimento prático subjetivo dos especialistas, somando-o à precisão e escalabilidade analítica dos modelos computacionais. Assim, a hipótese foi definida da seguinte forma:

Hipótese 1: *A combinação em modelos de ML de fatores heurísticos usados por especialistas em Prevenção à Lavagem de Dinheiro e fatores racionais usados pelas instituições financeiras, contribui para a redução da taxa atual de falsos positivos na detecção de atividades suspeitas de lavagem de dinheiro.*

3. METODOLOGIA E DESIGN DE PESQUISA

Foi adotada uma abordagem de métodos mistos em um processo sequencial, combinando o método qualitativo e quantitativo para investigar como melhorar a PLD em transações financeiras. Em uma primeira fase, foi aplicado o método qualitativo para identificar as regras de bolso mensuráveis (i.e., fatores heurísticos) mais frequentemente usados pelos analistas à PLD nas suas análises. Estas regras práticas e intuitivas têm por base o conhecimento empírico dos especialistas. A primeira fase teve por base entrevistas com analistas de diferentes instituições financeiras. Foram selecionados para passar para a 2ª fase de análise os fatores heurísticos que podem ser mensurados quantitativamente. Em uma segunda fase, foi aplicado o método quantitativo com enfoque em uma instituição financeira específica do mercado brasileiro. A fase quantitativa teve 2 etapas. Na primeira etapa, foram identificadas as regras duras (i.e., fatores racionais) usadas por essa instituição no processo de monitoramento de PLD, identificando-se quais os mais eficazes. Esses fatores racionais foram depois, em uma segunda etapa, integrados em um modelo de ML *Random Forest* com os fatores heurísticos identificados na fase qualitativa da pesquisa. O modelo foi então testado de forma a avaliar a sua eficácia na redução da taxa de falsos positivos na detecção de atividades suspeitas de lavagem de dinheiro. Estas duas fases e métodos aplicados são apresentados em maior detalhe nas seções abaixo.

3.1 Fase 1 - Método Qualitativo

A pesquisa qualitativa é fundamental para explorar fenômenos complexos, como a identificação de padrões sutis de lavagem de dinheiro, objeto desta pesquisa. Para além disso, o método qualitativo permite também integrar as percepções e experiências dos especialistas diretamente envolvidos nessas avaliações (Creswell, 2009), sendo particularmente eficazes para compreender como os indivíduos atribuem significado às suas ações e decisões em contextos específicos. Em termos concretos, nesta pesquisa o método qualitativo ofereceu uma possibilidade de identificar práticas rotineiras informais dos especialistas em lavagem de dinheiro, que refletem regras de bolso (i.e., fatores heurísticos) adotadas em função da sua experiência e conhecimento acumulado em ambientes de incerteza (Kahneman; Tversky, 1973), ou seja, na análise de casos assinalado como apresentado alto potencial de risco.

3.1.1 Seleção de Participantes

A seleção de participantes para este estudo concentrou-se em especialistas em PLD do setor financeiro em diversas instituições financeiras e bancos de diferentes portes no Brasil. Esta diversidade visou obter uma visão abrangente das práticas adotadas no mercado em geral, e não apenas restritas a uma instituição específica. Em cada instituição, foram intencionalmente selecionados especialistas em PLD, que tivessem (1) experiência significativa no setor (i.e., mínimo de cinco anos de experiência), (2) envolvimento direto na identificação de transações suspeitas, e (3) conhecimento aprofundado das regulamentações e técnicas de PLD. O objetivo da aplicação destes critérios de seleção foi o de garantir a obtenção de *insights* sobre os fatores heurísticos empregados na identificação de atividades suspeitas.

No total, foram incluídos no estudo 15 especialistas e analistas de PLD de três instituições financeiras (bancos ou IPs²), com idades, anos de experiência, escolaridade e cargos distintos (ver Quadro 1).

Quadro 1 – Perfil dos entrevistados

Entrevistado	idade	Anos de Experiencia	Escolaridade	Cargo	Empresa	Tipo (Banco ou IP)
Entrevistado_1	45	22 anos	Pós Graduação	Especialista de PLD/FT	Empresa_1	Banco
Entrevistado_2	34	12 anos	Graduação	Especialista de PLD/FT	Empresa_1	Banco
Entrevistado_3	37	9,6 anos	Pós Graduação	Especialista de PLD/FT	Empresa_2	IP
Entrevistado_4	46	15 anos	Graduação	Analista de PLD/FT	Empresa_2	IP
Entrevistado_5	41	14 anos	Pós Graduação	Especialista de PLD/FT	Empresa_3	Banco
Entrevistado_6	44	20 anos	Graduação	Analista de PLD/FT	Empresa_3	Banco
Entrevistado_7	40	9 anos	Graduação	Analista de PLD/FT	Empresa_1	Banco
Entrevistado_8	46	15 anos	Graduação	Analista de PLD/FT	Empresa_1	Banco
Entrevistado_9	44	19 anos	Graduação	Analista de PLD/FT	Empresa_3	Banco
Entrevistado_10	37	9 anos	Pós Graduação	Analista de PLD/FT	Empresa_1	Banco
Entrevistado_11	39	5 anos	Pós Graduação	Analista de PLD/FT	Empresa_1	Banco
Entrevistado_12	40	13 anos	Pós Graduação	Analista de PLD/FT	Empresa_2	IP
Entrevistado_13	38	12 anos	Pós Graduação	Especialista de PLD/FT	Empresa_2	IP
Entrevistado_14	34	14 anos	Graduação	Analista de PLD/FT	Empresa_1	Banco
Entrevistado_15	39	9 anos	Pós Graduação	Analista de PLD/FT	Empresa_1	Banco

Fonte: Elaborado pelo autor

² As Instituições de Pagamento (IPs), regulamentadas pela Lei nº 12.865/2013, são entidades autorizadas pelo Banco Central para viabilizar transações financeiras, como transferências, pagamentos e emissão de moeda eletrônica. Diferentemente dos bancos, elas não captam depósitos remunerados nem oferecem crédito.

Analisando o Quadro 1, é possível observar que a idade média dos respondentes foi de aproximadamente 40 anos (idade mínima de 34 anos e máxima de 46 anos), com uma média de aproximadamente 13 anos de experiência (mínimo de 5 anos e máximo de 22 anos). Todos os entrevistados tinham graduação ou pós-graduação, sendo 5 dos respondentes especialistas de PLD/FT (financiamento terrorista) e os restantes 1º analistas de PLD/FT.

As entrevistas foram suspensas quando se atingiu a saturação temática, ou seja, quando nas novas entrevistas conduzidas se deixou de identificar novas informações relevantes que pudessem trazer novos *insights* para a pesquisa, havendo já suporte empírico suficiente para os achados da análise de dados (Creswell, 2009). Desta forma, este número de participantes foi considerado adequado para fornecer uma visão robusta e abrangente das práticas informais ou uso de fatores heurísticos pelos especialistas na PLD, garantindo a profundidade e a qualidade dos achados.

3.1.2 Técnica de Coleta de Dados

A coleta dos dados foi realizada por meio de entrevistas, que permitiram explorar as experiências e percepções dos participantes da pesquisa na PLD de forma aprofundada (Kavle e Brinkmann, 2009). As entrevistas seguiram um roteiro previamente elaborado, mas foram conduzidas de forma semiestruturada e com questões abertas, permitindo flexibilidade para que os entrevistados pudessem abordar aspectos específicos e relevantes à sua experiência. Esse método foi escolhido considerando a natureza subjetiva e interpretativa do tema investigado, alinhando-se a diretrizes descritas por autores como Kavle e Brinkmann (2009). Essa abordagem não apenas permitiu identificar padrões e tendências, mas também proporcionou uma compreensão mais profunda das motivações e desafios enfrentados pelos participantes no contexto da PLD.

Conforme o Anexo 1, todos os entrevistados assinaram um termo de consentimento antes do início da entrevista, garantindo sua anuência e a confidencialidade das informações fornecidas. As entrevistas foram conduzidas por videoconferência com suporte da ferramenta Microsoft Teams, com duração aproximada de 30 minutos cada. Com o consentimento dos participantes, todas as sessões foram gravadas para posterior transcrição com suporte do software Transkriptor.

Conforme o roteiro apresentado no Anexo 2, após uma questão geral sobre PLD, foram apresentados a cada um dos entrevistados dois de cinco cenários específicos. O uso

de cenários teve por objetivo obter uma visão detalhada sobre o processo de tomada de decisão desses profissionais em situações de risco, e mais especificamente sobre as regras de bolso adotadas. O uso de cenários possibilita uma maior contextualização da situação analisada, ajudando o entrevistado a explicar o seu comportamento e escolhas (i.e., processo seguido e identificação de regras de bolso usadas) com maior profundidade e detalhe.

3.1.3 Técnica de Análise de Dados

A análise de dados qualitativos seguiu a metodologia de análise de conteúdo, conforme descrita por Bardin (2016). Este método permitiu identificar, categorizar e interpretar padrões significativos nos dados textuais. A análise foi feita com auxílio do software NVivo 14. Uma descrição mais detalhada deste processo é descrita de seguida:

Transcrição: Todas as entrevistas foram transcritas integralmente para garantir a precisão dos dados. As transcrições foram revisadas para assegurar que nenhuma nuance relevante fosse perdida durante o processo de codificação.

Codificação Inicial: Antes de avançar para a categorização, foi realizada uma codificação aberta, onde cada unidade de significado nos dados foi rotulada com códigos iniciais. Esses códigos ajudaram a organizar os dados em unidades menores e mais manejáveis, facilitando a análise subsequente.

Categorização: Os códigos foram agrupados em categorias temáticas, permitindo identificar padrões e subtemas que emergiram dos dados. As categorias incluíram "Fatores de Risco", "Técnicas de Detecção", "Desafios Operacionais" e "Melhorias Sugeridas". A categorização seguiu um processo de refinamento, onde temas iniciais foram reorganizados ou combinados com base em sua relevância e frequência nos dados. Isso permitiu uma compreensão mais profunda das interações entre os diferentes fatores capturados com os entrevistados.

Identificação de Padrões: Durante a categorização, padrões nos dados foram identificados. Isso pôde incluir padrões temporais (por exemplo, mudanças ao longo do tempo nas práticas de detecção), padrões de comportamento (como os especialistas lidavam com certos tipos de transações) e padrões contextuais (como diferentes instituições enfrentavam desafios semelhantes de formas variadas). A identificação de padrões proporcionou uma visão mais estruturada sobre os principais fatores que influenciam a eficácia dos métodos de detecção.

Interpretação: Os dados categorizados foram interpretados para compreender os principais fatores que influenciam a eficácia dos métodos de detecção de lavagem de dinheiro. As interpretações foram embasadas nas literaturas revisadas e nas respostas dos participantes. Além disso, as interpretações consideraram o contexto específico de cada instituição financeira e como os fatores qualitativos poderiam interagir com variáveis quantitativas já utilizadas nos sistemas de detecção.

3.2 Fase 2 - Método Quantitativo

O objetivo do método quantitativo consistiu em desenvolver e validar modelos preditivos baseados em ML para a detecção de atividades suspeitas de lavagem de dinheiro em instituições financeiras. Utilizando uma base de dados, composta por alertas de transações financeiras históricas, o estudo aplicou o modelo *Random Forest* para identificar padrões que indicam risco de lavagem de dinheiro. Conforme sugerido por Jullum et al. (2020), a utilização de abordagens quantitativas proporciona uma análise escalável e precisa, especialmente útil para processar grandes volumes de dados e reduzir a taxa de falsos positivos, uma das principais limitações dos sistemas tradicionais de alerta.

Além disso, a combinação de dados históricos com técnicas avançadas de ML possibilita a criação de modelos robustos capazes de capturar complexidades nas transações financeiras, que podem não ser facilmente detectáveis por métodos tradicionais ou exclusivamente qualitativos (Medar, Rajpurohit & Bachulkar, 2017). Dessa forma, essa abordagem não apenas buscou aumentar a precisão, como também otimizar a adaptabilidade dos modelos frente às contínuas evoluções nas práticas de lavagem de dinheiro, conforme destacado por Sudjianto et al. (2010), que ressaltam a importância da aplicação de métodos quantitativos dinâmicos e atualizáveis na PLD.

3.2.1 Amostragem

A amostragem quantitativa no estudo foi baseada em uma seleção de dados históricos de casos que foram identificados como “alertas” de transações suspeitas pelas regras de monitoramento de uma instituição financeira brasileira. A base de dados coletada junto dessa instituição financeira incluiu aproximadamente 45 mil alertas, provenientes da movimentação de 9,7 mil clientes no período de janeiro de 2022 a março de 2024, proporcionando a amostra necessária para a construção e validação dos modelos preditivos de ML. O estudo utilizou base de dados com alertas reais, classificados como suspeitos e não suspeitos, seguindo o método descrito por Jullum et al. (2020), no qual transações

legítimas — que não geraram comunicação de operação suspeita — foram incluídas para aprimorar a capacidade do modelo em reconhecer padrões típicos de comportamento financeiro e, assim, aumentar sua robustez na distinção entre atividades normais e potencialmente ilícitas.

Para assegurar que o modelo possa ser aplicável em diferentes contextos, a amostra incluiu transações de variados tipos, montantes, e origens, cobrindo um amplo espectro de atividades financeiras. Essa abordagem diversificada permitiu assegurar que os modelos desenvolvidos fossem generalizáveis e eficazes na identificação de padrões suspeitos em cenários reais e complexos, maximizando sua aplicabilidade.

3.2.2 Técnica de Coleta de Dados

A coleta de dados foi realizada a partir da extração dos alertas diretamente dos sistemas da instituição financeira participante. Os dados foram pré-processados para garantir a integridade e qualidade da base utilizada no treinamento dos modelos. O processo de pré-processamento incluiu, além da limpeza dos dados, a detecção e correção de valores ausentes e anômalos, a normalização das variáveis e a criação de variáveis sintéticas capazes de aumentar o poder preditivo do modelo (Lopez-Rojas & Axelsson, 2012).

3.2.3 Técnica de Análise de Dados

A análise dos dados quantitativos foi dividida em duas etapas. Na primeira etapa, foram identificados na instituição financeira objeto de estudo, todos os fatores racionais (i.e., regras duras) na detecção de transações suspeitas. Essa identificação foi feita através da extração da base de dados de alertas no período de janeiro de 2022 a março de 2024, permitindo compreender padrões associados a transações com maior potencial de risco.

Na segunda etapa da análise de dados quantitativos, foi usada técnica de ML *Random Forest*, integrando os fatores heurísticos mensuráveis identificados na fase qualitativa da pesquisa, com os fatores racionais identificados no primeiro estágio do método quantitativo. O processo de análise seguiu os seguintes passos:

Modelagem Preditiva: O primeiro passo foi realizado uma análise de associação entre variáveis categóricas, utilizando o teste Qui-quadrado. Cada alerta foi representado

por um vetor de características, incluindo histórico do cliente, detalhes da transação e padrões de alerta anteriores (Jullum et al., 2020).

Validação Cruzada: Para garantir a robustez dos modelos, foi aplicada validação cruzada k -fold, conforme sugerido por Hastie, Tibshirani e Friedman (2009). Nesse método, o conjunto de dados é dividido em k subconjuntos (ou "folds") de tamanho aproximadamente igual. O modelo é treinado k vezes, utilizando em cada iteração $k - 1$ folds para treino e o fold restante para teste. O F1-score foi utilizado como métrica principal de avaliação, pois representa a média harmônica entre precisão (proporção de alertas corretamente classificados como suspeitos) e taxa de detecção (capacidade do modelo de identificar todas as transações suspeitas presentes nos dados). Segundo Powers (2011), essa métrica evita distorções na avaliação do modelo, garantindo um melhor compromisso entre identificar corretamente transações suspeitas e minimizar alertas indevidos.

Treinamento e Avaliação dos Modelos: Com base nas variáveis identificadas, os modelos foram treinados utilizando o modelo de aprendizado supervisionado *Random Forest*, avaliando seu desempenho na classificação de alertas suspeitos. Esse método foi escolhido devido à sua robustez na detecção de padrões complexos e capacidade de lidar com dados desbalanceados, características frequentemente encontradas em cenários de detecção de fraudes financeiras (Chen & Guestrin, 2016; Breiman, 2001).

Interpretação dos Resultados: Nessa etapa validou-se se a integração das regras duras e regras de bolso melhora de fato a discriminação de alertas com maior propensão a risco, e qual percentual de descarte pode ser aplicado para melhorar o processo operacional de detecção, de forma a permitir maior precisão e uma redução substancial nas taxas de falsos positivos (Sudjianto et al., 2010). Os resultados foram analisados com base na comparação entre as variações de modelos com e sem a aplicação de regras, utilizando métricas como o F1-score.

4. RESULTADOS

4.1 Fase 1- Pesquisa Qualitativa

A partir das entrevistas conduzidas, emergiu um conjunto de fatores heurísticos, ou regras de bolso, que os especialistas em PLD empregam para identificar e priorizar alertas com maior potencial de se tornarem uma comunicação ao COAF. Cinco fatores foram identificados como mais frequentemente adotados pelos especialistas e analistas, e que foram agrupados em 2 grupos, nomeadamente: (1) incompatibilidade entre movimentação financeira e características específicas do indivíduo e (2) variáveis contextuais ou individuais. O primeiro agrupou três fatores: (1.a) idade, renda e movimentação financeira, (1.b) a profissão e o perfil de movimentação, e (1.c.) renda declarada, profissão e movimentação financeira. Já o segundo agrupou os dois fatores restantes: (2.a.) local de residência e risco geográfico e (2.b.) mídia negativo e histórico de envolvimento. No Quadro 2, são apresentadas algumas citações ilustrativas dos entrevistados para cada um destes fatores heurísticos identificados.

Quadro 2 – Achados das entrevistas

Grupo	Entrevistado	Comentário	
(1) incompatibilidade entre movimentação financeira e características específicas do indivíduo	(1.a) idade, renda e movimentação financeira,	Entrevistado_13	"Sim, a localidade dele é um fator muito importante . A idade, uma pessoa de 24 anos, geralmente pessoas de 24 anos no início de carreira geralmente ganham um pouco, a renda é um pouco mais baixa."
		Entrevistado_14	"Pela idade, o cliente com 24 anos não parece ter ainda uma renda ou patrimônio que justificaria a movimentação elevada."
		Entrevistado_15	"A idade de 24 anos e o patrimônio baixo de 20 mil reais tornam a movimentação alta em 15 dias suspeita, e caberia uma investigação mais profunda."
		Entrevistado_3	"Eu avaliaria se as contrapartes de entrada e saída seriam as mesmas ou seriam diferentes... avaliaria se as contrapartes são das mesmas regiões, porque poderia ser uma conta de laranja."
		Entrevistado_7	"Ele é jovem, não é um impeditivo receber esse valor, pode ser herança, mas pela renda baixa, chama atenção."
	(1.b) a profissão e o perfil de movimentação	Entrevistado_1	"Um ponto a avaliar seria a profissão do cliente, especialmente em casos de grandes movimentações que poderiam ser consistentes com a atividade empresarial ou comercial."
		Entrevistado_12	"A renda é um pouco baixa, mas se ele tiver uma atividade comercial por fora, pode justificar essas movimentações , então busco por qualquer pista disso."
		Entrevistado_12	"Confirmaria se ele tem alguma participação societária em empresas, isso pode ajudar a entender o volume de movimentação."
		Entrevistado_5	"Eu entenderia as contrapartes com que ele está se relacionando, se ele tem, às vezes ele é um empregado, mas ele também tem participação societária em alguma empresa ."
		Entrevistado_8	"Também verificaria se ele é sócio de alguma empresa ou possui participação societária , o que pode agregar mais informação ao caso."
	(1.c.) renda declarada, profissão e movimentação financeira	Entrevistado_10	"A incompatibilidade entre a renda de burau e a movimentação elevada, em um curto período de tempo, é algo que investigaria mais profundamente."
		Entrevistado_11	"Quando a renda declarada não bate com o montante movimentado , eu já considero isso um sinal claro de que algo está fora do normal."
		Entrevistado_12	"O montante movimentado chama atenção, considerando esses outros pontos, como a renda CLT, a idade e o local onde reside."
		Entrevistado_5	"A movimentação de 679 mil em 15 dias, para um CLT com renda de 5 mil, não faz sentido."
		Entrevistado_9	"Renda versus movimentação já olhando está incompatível. O dinheiro vem de um monte de gente e ele já repassa para outros. Isso chama atenção."
(2) variáveis contextuais ou individuais	(2.a.) local de residência e risco geográfico	Entrevistado_1	"O local de residência , Umuarama no Paraná, que é na fronteira, chama atenção também. A região em si já gera um risco adicional."
		Entrevistado_13	"O local de residência do cliente em regiões de fronteira , como Umuarama, já aumenta o risco devido à proximidade com áreas conhecidas por contrabando."
		Entrevistado_6	"A primeira coisa é olhar a região , já que ele mora em Umuarama, próximo à fronteira, isso já seria algo para considerar como fator de risco."
		Entrevistado_14	"A localidade interfere no jeito de você analisar o alerta. Dependendo da localização, já gera um pouco mais de risco."
		Entrevistado_4	"A região de fronteira eleva o risco do caso , pois essas áreas são conhecidas por alto índice de lavagem de dinheiro."
	(2.b.) mídia negativo e histórico de envolvimento	Entrevistado_4	"Nesse caso aqui, diante das informações que você me passou aqui, o que levantaria pra mim uma bandeira vermelha aqui seria, ele é perto de região fronteiriça , a capacidade financeira dele não comporta."
		Entrevistado_10	"Tentaria buscar se ele já foi alertado em períodos anteriores , ou se já teve alguma mídia negativa."
		Entrevistado_13	"Busco informações de mídia desabonadora para ver se há algo que corrobore o risco, tanto para o cliente quanto para as contrapartes."
		Entrevistado_8	"Eu iria olhar a Bolsa Família, até pela idade, ele mora num... no interior, né, iria ver se ele foi, se ele é algum beneficiário e mídias, tá, processo, se tem algum processo criminal relacionado à LD e pesquisas no Google para ver se tem alguma informação adicional."
		Entrevistado_11	"Buscaria também se ele tem algum processo ou alguma mídia que possa corroborar com a análise, especialmente se envolve crimes financeiros."
		Entrevistado_12	"Além das informações do alerta, procuraria saber se ele tem alguma mídia negativa ou processos judiciais que pudessem influenciar a análise."
		Entrevistado_13	"Procuraria o nome do cliente pra ver se tem mídia ou se ele tem algum processo criminal que possa corroborar na análise do risco."
Entrevistado_9	"Vejo se ele tem algum processo, especialmente relacionado à lavagem de dinheiro ."		

Fonte: Elaborado pelo autor

Relativamente à (1) incompatibilidade entre movimentação financeira e características específicas do indivíduo, aspectos como (1.a.) idade e renda revelaram padrões que frequentemente indicam risco elevado. Jovens com menos de 25 anos, cuja renda declarada é incompatível com altos volumes de movimentação financeira, foram apontados como casos de atenção prioritária. Um dos entrevistados destacou:

"A idade de 24 anos e o patrimônio baixo de 20 mil reais tornam a movimentação alta em 15 dias suspeita, e caberia uma investigação mais profunda." (Entrevistado_15).

Esse comportamento pode indicar tanto a lavagem de dinheiro quanto o uso da conta de um indivíduo para ocultar recursos. Outro especialista reforçou a importância dessa análise ao afirmar:

"Ele é jovem, não é um impeditivo receber esse valor, pode ser herança, mas pela renda baixa, chama atenção." (Entrevistado_7).

A (1.b.) coerência entre a profissão e o padrão de movimentação financeira foi igualmente apontada como um aspecto central na análise de risco. Inconsistências significativas nesse aspecto geram questionamentos quanto à origem dos recursos. Por exemplo, um entrevistado observou:

"A renda é um pouco baixa, mas se ele tiver uma atividade comercial por fora, pode justificar essas movimentações, então busco por qualquer pista disso." (Entrevistado_12).

Em outro caso, foi mencionada a relação participação em empresas e grandes transações:

"Confirmaria se ele tem alguma participação societária em empresas, isso pode ajudar a entender o volume de movimentação." (Entrevistado_12).

A interação entre fatores como (1.c.) renda, ocupação e volume transacional também foi apontada como determinante. Quando essas variáveis são incoerentes, o alerta de risco é elevado. Um especialista relatou:

" Quando a renda declarada não bate com o montante movimentado, eu já considero isso um sinal claro de que algo está fora do normal." (Entrevistado_11).

A análise desses casos requer uma verificação mais aprofundada, especialmente quando a profissão declarada não sustenta o volume das transações realizadas.

Quanto aos (2) fatores contextuais ou individuais, como aspectos geográficos, foram destacados como sinais de riscos. (2.a) Residência em áreas de fronteira ou regiões associadas a práticas ilícitas, como por exemplo regiões com alto fluxo de dinheiro vivo e histórico de contrabando foram apontadas como zonas de atenção. Um especialista explicou:

"A localidade interfere no jeito de você analisar o alerta. Dependendo da localização, já gera um pouco mais de risco." (Entrevistado_14).

Por fim, o impacto reputacional e o histórico legal do indivíduo foram amplamente citados como indicadores na priorização das análises. Pessoas vinculadas a (2.b.) processos legais ou citadas em mídia negativa têm maior probabilidade de estarem envolvidas em atividades suspeitas. Um dos entrevistados apontou:

"Busco informações de mídia desabonadora para ver se há algo que corrobore o risco, tanto para o cliente quanto para as contrapartes." (Entrevistado_13).

4.1.1. Operacionalização das Variáveis

Segue uma descrição de cada um dos fatores heurísticos identificados com a pesquisa qualitativa, i.e., as regras de bolso usados na priorização de alertas suspeitos.

- **Histórico de Processos Criminais do Cliente (PROC_CRIM):** Indica se o cliente possui registros de processos criminais transitados em julgado, relacionados a crimes financeiros, corrupção, tráfico ou outros delitos associados à lavagem de dinheiro. A consulta é realizada nos sites dos Tribunais de Justiça e do Supremo Tribunal de Justiça.
- **Registro de Comunicações Anteriores ao COAF (COMUN_ANTER):** Refere-se à existência de comunicações prévias do cliente ao COAF pela instituição, indicando possíveis transações suspeitas reportadas anteriormente.
- **Participação Societária (SOC_EMP):** Identifica se o cliente possui vínculo com empresas como sócio, administrador ou beneficiário final, considerando riscos associados a empresas de fachada ou movimentação incompatível com sua atividade.
- **Residência em Áreas de Risco (CEP_RISCO):** Avalia se o cliente reside em regiões com histórico de atividades ilícitas, como áreas de fronteira, zonas de grande circulação de dinheiro em espécie ou locais frequentemente mencionados em investigações financeiras. Segundo o GAFI (2023), essas áreas incluem jurisdições com deficiências estratégicas em PLD, baixa efetividade nos controles financeiros e alta exposição a crimes como contrabando e corrupção.
- **Menções Negativas na Mídia (MIDIA_NEG):** Verifica se o cliente foi citado em veículos de comunicação de grande circulação e credibilidade, investigações jornalísticas ou documentos públicos relacionados a fraudes, corrupção ou crimes financeiros.

- **Ocupação Declarada do Cliente (Profissão):** Informação fornecida pelo cliente no momento do cadastro na instituição, utilizada para verificar a compatibilidade com seus padrões de movimentação financeira.
- **Idade:** Idade declarada pelo cliente no momento do cadastro. Perfis atípicos, como jovens com movimentações elevadas sem histórico profissional ou aposentados com transações de grande valor, podem indicar risco.
- **Renda Mensal:** Valor da renda mensal informada no cadastro. Considerando sua possível variação ao longo do tempo, a instituição pode validar essa informação por meio de fontes externas para maior precisão.
- **Valor Alertado (Montante da Transação Alertado):** Refere-se ao valor da transação identificada como suspeita. Montantes elevados ou incompatíveis com o perfil do cliente podem indicar potenciais atividades ilícitas.

Das variáveis listadas, com exceção da Ocupação Declarada de Clientes (Profissão), que foi descartada devido a falta de informação para 29% dos casos, as demais foram utilizadas na etapa quantitativa.

4.2 Fase 2 – Pesquisa Quantitativa

4.2.1 Primeira Etapa: Identificação de Fatores Racionais mais eficazes

Nesta etapa, foi avaliado a eficácia das regras duras utilizadas na identificação de transações suspeitas. O objetivo consistiu em identificar quais regras apresentam maior precisão e quais possuíam altos índices de falso-positivos, permitindo ajustes e otimizações. A análise baseou-se no teste Qui-Quadrado de Pearson para verificar a associação entre as variáveis e o parecer dos analistas.

O índice médio de falso positivo ao longo do período analisado foi de 84,3%, indicando que a maioria dos alertas gerados pelas regras duras não se confirma como suspeita. Nos primeiros meses de 2022, o índice atingiu 100%, evidenciando a ineficiência dessas regras na detecção de atividades ilícitas. A Tabela 1 apresenta os cruzamentos entre cada uma das 12 regras duras e o parecer dos analistas, além do teste Qui-Quadrado de Pearson para avaliar a associação entre as variáveis categóricas.

Tabela 1 – Relação do Parecer dos Analistas e Regras Duras

Variável		Parecer Analista				Total		Qui-Quadrado	
		Sem Risco		Com Risco		N	%		
		N	%	N	%				
R1	Não	36.228	80,91	8.545	19,09	44.773	99,36	Estatística valor-p	32,52 0.000
	Sim	272	94,12	17	5,88	289	0,64		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
R2	Não	35.531	80,71	8.492	19,29	44.023	97,69	Estatística valor-p	103,92 0.000
	Sim	969	93,26	70	6,74	1.039	2,31		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
R3	Não	34.867	80,74	8.318	19,26	43.185	95,83	Estatística valor-p	45,83 0.000
	Sim	1.633	87,00	244	13,00	1.877	4,17		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
R4	Não	34.558	81,71	7.733	18,29	42.291	93,85	Estatística valor-p	228,62 0.000
	Sim	1.942	70,08	829	29,92	2.771	6,15		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
R5	Não	27.221	81,37	6.231	18,83	33.452	74,24	Estatística valor-p	11,79 0.001
	Sim	9.279	79,92	2.331	20,08	11.610	25,76		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
R6	Não	36.421	81,03	8.524	18,97	44.945	99,74	Estatística valor-p	13,85 0.000
	Sim	79	67,52	38	32,48	117	0,26		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
R7	Não	14.200	85,05	2.497	14,95	16.697	37,05	Estatística valor-p	282,10 0.000
	Sim	22.300	78,62	6.065	21,38	28.365	62,95		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
R8	Não	34.306	80,29	8.424	19,71	42.730	94,82	Estatística valor-p	273,50 0.000
	Sim	2.194	94,08	138	5,92	2.332	5,18		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
R9	Não	36.323	81,00	8.521	19,00	44.844	99,52	Estatística valor-p	0,01 0.942
	Sim	177	81,19	41	18,81	218	0,48		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
R10	Não	36.381	81,03	8.518	18,97	44.899	99,64	Estatística valor-p	6,79 0.009
	Sim	119	73,01	44	26,99	163	0,36		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
R11	Não	31.286	82,46	6.653	17,54	37.939	84,19	Estatística valor-p	334,45 0.000
	Sim	5.214	73,20	1.909	26,80	7.123	15,81		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
R12	Não	35.854	81,90	7.924	18,10	43.778	97,15	Estatística valor-p	808,74 0.000
	Sim	646	50,31	638	49,69	1.284	2,85		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		

Fonte: Elaborado pelo autor

As regras R1, R2 e R8 registraram índices de falso positivo superiores a 93%. Em contrapartida, as regras R6 e R12 destacaram-se com os menores índices, 67,52% e 50,31%, respectivamente, indicando maior precisão na identificação de transações suspeitas. Os dados analisados mostram variações na eficácia das regras, com algumas apresentando altos índices de falso positivos, enquanto outras demonstram maior precisão. Em vez de descartar regras com baixo desempenho.

4.2.2 Segunda Etapa: Integração de Fatores Heurísticos e Racionais em ML *Random Forest*

Nesta fase, os fatores heurísticos identificados na análise qualitativa – residência em áreas de risco, menções negativas na mídia, processos criminais, histórico de comunicações anteriores ao COAF, participação societária, idade, renda e valor alertado – foram analisados em conjunto com o parecer dos analistas para avaliar sua associação com o risco de lavagem de dinheiro.

Os dados apresentados na Tabela 2 indicam que processos criminais, comunicações anteriores e participação societária possuem associação significativa com o risco ($p < 0,001$), o que corrobora com os comentários capturados nas entrevistas. Já o uso da informação de residência em áreas de risco e menções na mídia não demonstraram relevância estatística.

Tabela 2 – Relação do Parecer dos Analistas e Fatores identificados na fase qualitativa

Variável		Parecer Analista				Total		Qui-Quadrado	
		Sem Risco		Com Risco		N	%		
		N	%	N	%				
CEP_RISCO	Não	35.441	80,97	8.327	19,03	43.768	97,13	Estatística	0,61
	Sim	1.059	81,81	235	18,16	1.294	2,87		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
MIDIA_NEG	Não	36.314	81,02	8.506	18,98	44.820	99,46	Estatística	2,71
	Sim	186	76,86	56	23,14	242	0,54		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
PROC_CRIM	Não	35.385	82,25	7.636	17,75	43.021	95,47	Estatística	965,89
	Sim	1.115	54,63	926	45,37	2.041	4,53		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
COMUN_ANTER	Não	33.242	86,24	5.306	13,76	38.548	85,54	Estatística	4,7e +03
	Sim	3.258	50,02	3.256	49,98	6.514	14,46		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		
SOC_EMP	Não	12.732	70,01	5.455	29,99	18.187	40,36	Estatística	2,4e +03
	Sim	23.768	88,44	3.107	11,56	26.875	59,64		
	Total	36.500	81,00	8.562	19,00	45.062	100,00		

Fonte: Elaborado pelo autor

O histórico de processos criminais apresentou forte associação com propensão de lavagem de dinheiro. A análise da distribuição dos casos revelou que 45% dos alertas de clientes com processos criminais foram classificados como suspeitos pelos analistas. Esse resultado reforça a relevância de antecedentes criminais como um indicador potencial de envolvimento em atividades de lavagem de dinheiro.

A comunicação anterior ao COAF também apresentou associação significativa. Aproximadamente 50% dos alertas de clientes que haviam sido comunicados anteriormente

foram classificados como suspeitos, um percentual substancialmente superior à média geral de 19% dos casos identificados como de risco.

Em relação à participação societária, observou-se que, embora haja associação estatisticamente significativa, o impacto na probabilidade de um alerta ser considerado suspeito se manifesta de forma distinta. Apenas 12% dos alertas de clientes com participação societária foram classificados como suspeito, um percentual inferior à taxa geral de 19%. Esse dado sugere que, apesar da relevância estatística, a participação societária pode estar associada a outras dinâmicas que reduzem a probabilidade de classificação como suspeito.

Por outro lado, as variáveis área de risco e menções negativas na mídia não apresentaram associação significativa com o risco de lavagem de dinheiro. A diferença na classificação de risco entre os grupos dessas variáveis manteve-se em torno de 20%, um valor próximo à taxa geral de 19%, indicando que a influência desses fatores no risco de lavagem de dinheiro é limitada.

A Tabela 3 apresenta a análise das variáveis numéricas identificadas na etapa qualitativa, i.e., idade, renda e valor alertado, previamente identificadas na etapa qualitativa. A idade dos clientes não difere significativamente entre os alertas classificados como suspeito ou não suspeito, portanto, isoladamente, a idade não é um fator discriminatório nessa classificação. Por outro lado, a renda mensal apresenta uma diferença estatisticamente significativa entre os grupos, onde clientes objeto de alertas classificados como "não suspeitos" apresentam uma renda média superior em comparação aos clientes objeto de alertas classificados como "suspeitos". Esse resultado sugere que incompatibilidades entre a renda declarada e o volume transacionado podem indicar maior probabilidade de influenciar a classificação do alerta como suspeito, seja por possíveis irregularidades financeiras ou pelo uso de uma renda declarada artificialmente inferior para reduzir a visibilidade no monitoramento. Por fim, a variável valor alertado mostrou uma diferença estatisticamente significativa entre os grupos, confirmando que transações de maior montante possuem maior probabilidade de serem classificadas como suspeitas. Essa tendência reforça a necessidade de monitoramento aprimorado para transações de alto valor, que frequentemente são utilizadas em esquemas de ocultação de recursos ilícitos.

Tabela 3 – Relação do Parecer dos Analistas e variáveis contínuas

Variável	Parecer Analista	N	Média	Desv_Pad	Min	Máx	t-Student	
Idade	Sem Risco	36.476	40,00	9,88	39,90	40,10	Estatística	1,08
	Com Risco	8.559	39,87	10,00	39,66	40,08	valor-p	0.2789
	Combinado	45.035	39,98	9,90	39,88	40,07		
Renda_Mensal	Sem Risco	36.500	8.423,94	8.527,24	8.326,46	8.501,42	Estatística	19,44
	Com Risco	8.562	6.463,98	7.559,54	6.303,83	6.624,12	valor-p	0.000
	Combinado	45.062	8.043,44	8.386,89	7.966,00	8.120,87		
Valor Alertado	Sem Risco	36.500	71.891,54	75.806,49	71.113,82	72.669,26	Estatística	-6,6535
	Com Risco	8.562	77.713,08	87.475,68	76.859,94	80.566,23	valor-p	0.000
	Combinado	45.062	73.187,67	9,90	39,88	40,07		

Fonte: Elaborado pelo autor

4.3 Avaliação dos Modelos Aplicados na Geração de Alertas

Nesta seção, foi analisada a eficácia dos modelos aplicados para a identificação de alertas de possíveis atividades suspeitas. Os modelos foram construídos de forma sequencial, partindo da aplicação de “regras duras até a integração com fatores heurísticos e a seleção automática de variáveis pelo próprio modelo.

Para essa análise, utilizamos o modelo de *Random Forest*, técnica baseada em múltiplas árvores de decisão. Essa abordagem foi escolhida devido à sua robustez na classificação de padrões complexos e à sua capacidade de lidar eficientemente com variáveis categóricas e numéricas (Shwartz-Ziv & Armon, 2022). O modelo foi treinado utilizando 80% da base de dados e testado nos 20% restantes, seguindo uma prática comum em Machine Learning para garantir um bom equilíbrio entre aprendizado e generalização (Hastie, Tibshirani & Friedman, 2009). Além disso, foi aplicada a validação cruzada k-fold, com a divisão dos dados em cinco partes (*folds*), permitindo que o modelo fosse treinado e testado em diferentes subconjuntos da base, reduzindo vieses e melhorando a confiabilidade dos resultados (Powers, 2011).

A Tabela 4 apresenta as métricas de desempenho dos modelos testados, com destaque para F1-score, que permite avaliar o equilíbrio entre precisão e capacidade de detecção de alertas suspeitos.

Tabela 4 – Estatísticas para comparação dos modelos

Fatores	Mod_Regras Duras	Mod_Heurísticos	Mod_Regras Duras + Heurísticos	Mod_Regras Duras + Heurísticos (seleção pelo modelo)
<i>F1-score</i>	0.3520	0.5012	0.5619	0.5405
Área curva ROC	0.6419	0.8142	0.8424	0.8263
classificação correta (total)	67,0%	79,0%	82,0%	81,0%
Falso positivo	22,8%	12,7%	10,5%	11,3%
Falso negativo	10,1%	8,4%	7,5%	7,8%

Fonte: Elaborado pelo autor

Inicialmente, foi utilizado o modelo baseado apenas em regras duras que demonstrou limitações, com um F1-score baixo, com taxa de falso positivo de 22,8%. Estes resultados ratificaram o entendimento que embora as regras duras ofereçam um critério objetivo, sua aplicação isolada não é suficiente para capturar a complexidade das operações financeiras suspeitas.

No segundo modelo foram consideradas apenas fatores heurísticos identificados na pesquisa qualitativa, que refletem a experiência dos especialistas (como histórico de processo criminal, comunicados anteriormente ao COAF, participação societária, residência em áreas de risco, menções negativas na mídia, idade, renda e valor alertado), o modelo demonstrou ser 42,4% melhor que o modelo de regras duras, o que indica que a inclusão de fatores qualitativos complementa as regras fixas, permitindo uma melhor distinção entre transações legítimas e suspeitas.

A maior eficiência foi observada na abordagem híbrida, que combinou regras duras e fatores heurísticos. Esse modelo apresentou o melhor desempenho, com um F1-score de 0,5619, o que representa um aumento de 59,6% em relação ao modelo baseado apenas em regras duras, acurácia de 82% e redução da taxa do falso positivo para 10,5%. Estes resultados confirmam que a complementaridade entre regras objetivas e conhecimento analítico melhora significativamente a eficácia da geração de alertas.

Por fim, foi testado um quarto modelo que aplicou a seleção automática de variáveis, no qual o próprio algoritmo selecionou quais regras duras e fatores heurísticos eram mais relevantes, descartando variáveis de menor impacto. Porém, os resultados foram inferiores ao modelo híbrido completo e houve um leve aumento na taxa falso positivo para 11,3%, sugerindo que, embora a otimização computacional possa melhorar a eficiência, ela pode comprometer a captura de padrões críticos identificados pelos especialistas.

Como resultado dessa análise, conclui-se que o modelo híbrido demonstrou um melhor equilíbrio entre sensibilidade e especificidade, proporcionando maior confiabilidade na geração de alertas e reduzindo as limitações operacionais associadas a um grande volume de falsos positivos, o que proporcionará o descarte dos alertas inefetivos, gerando uma eficiência operacional e foco maior nos casos de risco.

Considerando o cenário da amostra onde índice de falso positivo é 84,3%, se for feita uma aplicação conservadora do modelo, utilizando uma taxa de probabilidade de acerto “falso positivo” superior a 0,995, seria possível descartar 24,7% dos alertas “não suspeitos”, sem comprometer o descarte indevido de alertas “suspeitos”, com isso o índice de falso positivo sairia dos atuais 84,3% para 61,7%.

A Tabela 5 evidencia a distribuição do peso entre regras duras e fatores heurísticos nos modelos aplicados, destacando que, no modelo híbrido, os fatores heurísticos tiveram maior importância do que as regras duras, mantendo um uso mais equilibrado das variáveis. Já no modelo com seleção automática de variáveis, observou-se um aumento na relevância dos fatores heurísticos e uma redução das regras duras, sugerindo que o modelo privilegiou a experiência prática dos especialistas na tomada de decisão. No entanto, essa reconfiguração e a seleção de algumas variáveis pode ter levado à perda de desempenho.

Tabela 5 – Tabela de importância das variáveis por modelo

Fatores	Importância (%)	
	Racionais (Regras duras)	Heurísticos (Analistas)
Mod_Regras Duras	100,00%	n/a
Mod_Heurísticos	n/a	100,00%
Mod_Regras Duras + Heurísticos	28,93%	71,07%
Mod_Regras Duras + Heurísticos (seleção pelo modelo)	20,85%	79,15%

Fonte: Elaborado pelo autor

5. DISCUSSÃO

Os resultados obtidos neste estudo reforçam a importância da integração entre regras duras (i.e., fatores racionais) e ‘regras de bolso’ (i.e., fatores heurísticos) para aprimorar a detecção de atividades suspeitas de lavagem de dinheiro. A análise demonstrou que o uso isolado de regras duras, embora padronizado e objetivo, apresenta limitações expressivas, principalmente na geração excessiva de falsos positivos, um desafio recorrente nos sistemas tradicionais de monitoramento financeiro (Jullum et al., 2020). Esse problema compromete a eficiência operacional das instituições financeiras e pode gerar impactos negativos na alocação de recursos limitados para investigações.

O modelo híbrido, que combinou fatores racionais e fatores heurísticos, apresentou melhoria significativa no equilíbrio entre precisão e detecção de transações suspeitas, reduzindo substancialmente os falsos positivos. Essa abordagem reforça achados de Ferwerda et al. (2020), que destacam que a combinação de conhecimento especializado e aprendizado de máquina melhora a acurácia dos sistemas de monitoramento ao incorporar fatores subjetivos que modelos puramente algorítmicos poderiam desconsiderar.

Além disso, a adoção de fatores heurísticos permitiu uma adaptação mais dinâmica às características das transações analisadas, sugerindo que a expertise humana pode agregar valor ao aprendizado de máquina ao refinar padrões de risco. Entretanto, desafios ainda permanecem, como a necessidade de calibração contínua do modelo híbrido para evitar viés analítico e garantir conformidade regulatória. Adicionalmente é recomendado manter uma rotina de *feedback* com os especialistas para garantir que as variáveis heurísticas se mantêm e verificar se novas variáveis possam surgir. Essa preocupação é compartilhada por Brock e Von Wangenheim (2019), que enfatizam que a adoção de novas tecnologias no setor financeiro exige ajustes organizacionais e operacionais, incluindo treinamento de equipes e transparência nos critérios de modelagem.

Assim, esta pesquisa demonstra empiricamente que a integração de heurísticas com fatores racionais, e o aprendizado de máquina, podem aprimorar a detecção de atividades suspeitas, tornando os sistemas mais eficientes sem comprometer a gestão de risco e a capacidade de identificação de transações ilícitas. Além disso, os achados fornecem *insights* práticos para instituições financeiras, sugerindo que a adoção de modelos híbridos pode reduzir significativamente os custos operacionais associados à análise de falsos positivos e aprimorar a eficiência do processo de geração de alertas, permitindo um monitoramento mais estratégico e menos oneroso.

6. CONCLUSÕES

Este estudo demonstrou que a integração de abordagens heurísticas e racionais em modelos de ML pode aprimorar a detecção de atividades suspeitas de lavagem de dinheiro. A análise confirmou que o uso isolado de regras duras apresenta limitações significativas, enquanto a incorporação de heurísticas melhora a geração de alertas e reduz a sobrecarga operacional das instituições financeiras.

Além disso, os resultados indicam que a abordagem híbrida reduz a taxa de falsos positivos e melhora a eficiência operacional, equilibrando a precisão analítica dos modelos com a expertise dos especialistas. Essa sinergia reforça a necessidade de modelos flexíveis e adaptáveis, capazes de integrar conhecimento humano e aprendizado de máquina para otimizar sistemas de monitoramento financeiro.

A aplicabilidade de modelos híbridos explicáveis e auditáveis também foi evidenciada, demonstrando seu potencial para facilitar a aceitação regulatória e aprimorar os processos institucionais de detecção de crimes financeiros. A necessidade de ajustes contínuos e aprimoramento das metodologias empregadas se mantém como um desafio central, especialmente diante da constante evolução das técnicas criminosas.

Conclui-se que a integração entre conhecimento especializado e inteligência artificial é essencial para fortalecer a detecção de crimes financeiros. A implementação de modelos híbridos pode elevar a eficácia dos sistemas de monitoramento, reduzindo custos operacionais e fortalecendo a integridade do sistema financeiro.

6.1. Contribuições para a teoria

Esta pesquisa analisou a integração de heurísticas qualitativas e modelos quantitativos racionais na prevenção à lavagem de dinheiro, abordando lacunas na literatura sobre abordagens híbridas. Os resultados demonstraram que a combinação dessas estratégias aprimora a geração de alertas e reduz falsos positivos e negativos, tornando o modelo mais eficaz para instituições financeiras e órgãos reguladores, como sugerido por Tversky e Kahneman (1974). Essa combinação destaca a necessidade de metodologias que integrem a intuição dos especialistas a modelos estatísticos, criando uma estrutura preditiva mais ajustável às variações do mercado.

Além disso, contribui para o enriquecimento da pesquisa acadêmica sobre o uso de Machine Learning (ML) em crimes financeiros, destacando a sinergia entre fatores heurísticos e racionais na melhoria dos modelos de detecção. Estudos como os de Chen et

al. (2018) e Jullum et al. (2020) ressaltam que essa combinação pode reduzir falsos positivos, tornando os sistemas mais eficazes. Outro ponto central foi a demonstração da viabilidade de quantificar fatores heurísticos subjetivos em variáveis aplicáveis a modelos preditivos, permitindo que futuras pesquisas avancem na construção de abordagens explicáveis e auditáveis para regulação financeira.

6.2. Contribuições para a prática

As contribuições práticas desta pesquisa incluem a proposta de uma metodologia para instituições financeiras que combina fatores heurísticos qualitativos com fatores racionais, permitindo o desenvolvimento de sistemas mais alinhados às necessidades do setor. A implementação de modelos híbridos pode auxiliar instituições na otimização do processo de geração de alertas, reduzindo não apenas falsos positivos, mas também aprimorando a alocação de recursos para casos de maior risco.

Além disso, a integração dessas abordagens possibilita a redução de falsos positivos nos sistemas de alerta, ao unir a experiência prática dos analistas com análises de fatores racionais ou regras duras, conforme demonstrado nos resultados preliminares. Outro impacto relevante é a criação de procedimentos mais ágeis e eficazes, otimizando a alocação de tempo dos profissionais para a avaliação de transações de alto risco e fortalecendo o suporte operacional na detecção de atividades suspeitas. Isso sugere que a adoção de modelos híbridos pode ser um diferencial competitivo para instituições que buscam aprimorar sua conformidade regulatória sem comprometer a eficiência operacional.

6.3. Limitações e sugestões de pesquisa futura

Embora este trabalho apresente contribuições relevantes para o campo da PLD, algumas limitações merecem destaque. Primeiro, a análise quantitativa utilizou dados de uma única instituição financeira, o que pode restringir a generalização dos resultados. Para validar a aplicabilidade do modelo em diferentes contextos, pesquisas futuras devem expandir a base de análise para múltiplas instituições, considerando diferentes políticas internas de monitoramento.

Segundo, a modelagem de fatores heurísticos baseou-se nas regras de bolso identificadas qualitativamente, o que gera uma dependência significativa das percepções dos especialistas entrevistados. Essa limitação destaca a necessidade de complementar

essa abordagem com estudos longitudinais, visando validar a eficácia e consistência desses fatores em diferentes contextos e períodos temporais.

Por fim, apesar do potencial demonstrado pelo uso de ML, questões como a transparência dos modelos utilizados e sua aceitação por reguladores permanecem desafios relevantes. Pesquisas futuras poderiam explorar técnicas que aumentem a transparência e a confiança nos resultados gerados, promovendo maior aceitação por parte de reguladores e *stakeholders* no processo de decisão.

REFERÊNCIAS

ALNASSER, M. S. A. S. Money laundering in selected emerging economies: Is there a role for banks? **Journal of Money Laundering Control**, 2020. Disponível em: https://www.researchgate.net/publication/341912851_Money_laundering_in_selected_emerging_economies_is_there_a_role_for_banks

BAKRY, A. N.; ALSHARKAWY, A. S.; FARAG, M. S.; RASLAN, K. R. Automatic suppression of false positive alerts in anti-money laundering systems using machine learning. **Journal of Money Laundering Control**, 2024.

Disponível em: <https://link.springer.com/article/10.1007/s11227-023-05708-z>

BANCO CENTRAL DO BRASIL. Carta Circular nº 4.001, de 29 de janeiro de 2020. Define procedimentos de prevenção e combate à lavagem de dinheiro. **Normativos Banco Central** Disponível em:

<https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Circular&numero=4001>

BANCO CENTRAL DO BRASIL. Circular nº 3.978, de 23 de janeiro de 2020. Dispõe sobre procedimentos de prevenção à lavagem de dinheiro. **Diário Oficial da União**, 2020. Disponível em:

<https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Circular&numero=3978>

BARDIN, Laurence. **Análise de conteúdo**. Tradução de Luís Antero Reto e Augusto Pinheiro. 3. reimpr. da 1. ed. São Paulo: Edições 70, 2016. Disponível em: <https://acervo.enap.gov.br/cgi-bin/koha/opac-detail.pl?biblionumber=52319%20thumbnail-shelfbrowser>

BRASIL. Lei nº 9.613, de 3 de março de 1998. Dispõe sobre os crimes de "lavagem" ou ocultação de bens, direitos e valores, a prevenção da utilização do sistema financeiro para os ilícitos previstos nesta Lei; cria o Conselho de Controle de Atividades Financeiras (COAF) e dá outras providências. **Diário Oficial da União**, 04 mar. 1998. Disponível em:

<https://legislacao.presidencia.gov.br/atos/?tipo=LEI&numero=9613&ano=1998&ato=6f6cXSE1EeNpWTfd8>

BROCK, J.; VON WANGENHEIM, F. Demystifying AI: What Digital Transformation Leaders Can Teach You about Realistic Artificial Intelligence. *Change: The Magazine of Higher Learning*, v. 61, n. 4, p. 18-27, 2019.

Disponível em: <https://doi.org/10.1177/1536504219865226>.

BRYMAN, A. Integrating Quantitative and Qualitative Research: How Is It Done? *Qualitative Research*, v. 6, p. 97-113, 2006.

Disponível em: <http://dx.doi.org/10.1177/1468794106058877>

CANHOTO, A. I. Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective. *Journal of Business Research*, p. 1–12, 2020. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33100427/>

CHEN, Z.; et al. Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: A review. *Knowledge and Information Systems*, v. 57, n. 2, p. 245–285, 2018. Disponível em: <https://doi.org/10.1007/s10115-017-1144-z>

CRESWELL, J. W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 3. ed. Thousand Oaks: **Sage Publications**, 2009. Disponível em: <https://psycnet.apa.org/record/2008-13604-000>

CRESWELL, J. W.; PLANO CLARK, V. L. *Designing and Conducting Mixed Methods Research*. 3. ed. Los Angeles- **Sage Publications**, 2018

Disponível em: <https://search.worldcat.org/title/Designing-and-conducting-mixed-methods-research/oclc/1000385901>

DAMASIO, A. R. **Descartes' Error: Emotion, Reason, and the Human Brain**. New York: Avon Books, 1994. Disponível em: <https://philpapers.org/rec/DAMDEE>

DEMETIS, D. S. Fighting money laundering with technology: A case study of Bank X in the UK. *Decision Support Systems*, Elsevier, jan. 2018. Disponível em: <https://doi.org/10.1016/j.dss.2017.11.005>

DOSHI-VELEZ, F.; KIM, B. Towards a Rigorous Science of Interpretable Machine Learning. **Computer Science, Philosophy**, 2017. Disponível em: <https://www.semanticscholar.org/paper/Towards-A-Rigorous-Science-of-Interpretable-Machine-Doshi-Velez-Kim/5c39e37022661f81f79e481240ed9b175dec6513>

FATF/OECD - GAFILAT. Anti-money laundering and counter-terrorist financing measures – Brazil Fourth Round Mutual Evaluation Report. **FATF**, Paris, 2023. Disponível em: <https://www.fatf-gafi.org/en/publications/Mutualevaluations/Brazil-mer-2023.html>

FERWERDA, J. *et al.* Estimating money laundering flows with a gravity model-based simulation. **Scientific Reports**, v. 10, n. 1, p. 18552, 2020. Disponível em: <https://doi.org/10.1038/s41598-020-75653-x>

GIGERENZER, G. Why Heuristics Work. **Perspectives on Psychological Science**, v. 3, n. 1, p. 20-29, 2008. Disponível em: <https://journals.sagepub.com/doi/10.1111/j.1745-6916.2008.00058.x>

GIGERENZER, G.; GAISSMAIER, W. Heuristic Decision Making. **Annual Review of Psychology**, v. 62, p. 451-482, 2011. Disponível em: <https://www.annualreviews.org/content/journals/10.1146/annurev-psych-120709-145346>

GREENE, J. C.; CARACELLI, V. J.; GRAHAM, W. F. Toward a Conceptual Framework for Mixed-Method Evaluation Designs. **Educational Evaluation and Policy Analysis**, v. 11, n. 3, p. 255-274, 1989. Disponível em: <https://doi.org/10.3102/01623737011003255>

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2. ed. **Springer**, 2009. Disponível em: <https://doi.org/10.1007/978-0-387-84858-7>

JOHNSON, R. B.; ONWUEGBUZIE, A. J. Mixed Methods Research: A Research Paradigm Whose Time Has Come. **Educational Researcher**, v. 33, n. 7, p. 14-26, 2004. Disponível em: <https://doi.org/10.3102/0013189X033007014>

JOHNSON, R. B.; ONWUEGBUZIE, A. J.; TURNER, L. A. Toward a definition of mixed methods research. **Journal of Mixed Methods Research**, v. 1, n. 2, p. 112-133, 2007. Disponível em: <https://doi.org/10.1177/1558689806298224>.

JULLUM, M. *et al.* Detecting money laundering transactions with machine learning. **Journal of Money Laundering Control**, v. 23, n. 1, p. 173–186, 2020. Disponível em: <https://doi.org/10.1108/JMLC-07-2019-0055>

KUTE, D. V. *et al.* Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering: A Critical Review. **IEEE Access**, v. 9, p.82304-82325, 2021. Disponível em: <https://ieeexplore.ieee.org/document/9446887>

KRABUANRAT, K.; PHELPS, R. Heuristics and Rationality in Strategic Decision Making: An Exploratory Study. **Journal of Business Research**, v. 41, n. 1, p. 83-93, 1998. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0148296397000143>

KVALE, S.; BRINKMANN, S. *InterViews: Learning the Craft of Qualitative Research Interviewing*. 2nd ed. Thousand Oaks, CA: **Sage Publications**, 2009.

LOPEZ-ROJAS, E.; AXELSSON, S. Money laundering detection using synthetic data. In: **European Intelligence and Security Informatics Conference (EISIC)**, 2012. p. 78-85. Disponível em: https://www.researchgate.net/publication/224952602_Money_Laundering_Detection_using_Synthetic_Data

MEDAR, R.; RAJPUROHIT, V. S.; BACHULKAR, R. B. Impact of Training and Testing Data Splits on Accuracy of Time Series Forecasting in Machine Learning. In: 2017 **IEEE International Conference on Smart Technologies for Smart Nation**, Bangalore, 2017. p. 385-391.

PARISI, G. I. *et al.* Continual lifelong learning with neural networks: A review. **Neural Networks**, v. 113, p. 54-71, 2019. Disponível em: <https://doi.org/10.1016/j.neunet.2019.01.012>

POWERS, D. M. **Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation.** *Journal of Machine Learning Technologies*, v. 2, n. 1, p. 37-63, 2011. Disponível em: <https://doi.org/10.48550/arXiv.2010.16061>

RODRIGUEZ, J. D.; PÉREZ, A.; LOZANO, J. A. Sensitivity analysis of k-fold cross-validation in prediction error estimation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2010. Disponível em: https://www.researchgate.net/publication/261226686_Fundamental_study_on_curved_fold_ed_dipole_antenna

SCHOTT, P. A. Reference guide to anti-money laundering and combating the financing of terrorism. 2. ed. Washington, D.C.: **The World Bank**, 2006. Disponível em: <https://openknowledge.worldbank.org/handle/10986/14701>

SHWARTZ-ZIV, R.; ARMON, A. **Tabular Data: Deep Learning is Not All You Need.** *Journal of Machine Learning Research*, v. 23, p. 1-42, 2022. Disponível em: <https://doi.org/10.48550/arXiv.2106.03253>

SINGH, K.; BEST, P. Anti-Money Laundering: Using data visualization to identify suspicious activity. **International Journal of Accounting Information Systems**, v. 34, p. 100418, 2019. Disponível em: <https://doi.org/10.1016/j.accinf.2019.06.001>.

SLOVIC, P. *et al.* Risk as Analysis and Risk as Feelings: Some Thoughts About Affect, Reason, Risk, and Rationality. **Risk Analysis**, v. 24, n. 2, p. 311-322, 2004. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/j.0272-4332.2004.00433.x>

SUDJANTO, A. *et al.* Statistical Methods for Fighting Financial Crimes. **Technometrics**, v. 52, n. 1, p. 5-17, 2010. Disponível em: https://www.researchgate.net/publication/234031532_Statistical_Methods_for_Fighting_Financial_Crimes

TVERSKY, A.; KAHNEMAN, D. Judgment under Uncertainty: Heuristics and Biases. **Science**, v. 185, n. 4157, p. 1124-1131, 1974. Disponível em: <https://doi.org/10.1126/science.185.4157.1124>.

TVERSKY, A; KAHNEMAN, D. On the psychology of prediction. **Psychological Review**, v. 80, n. 4, p. 237-251, jul. 1973. DOI: 10.1037/h0034747. Disponível em: <https://www.proquest.com/docview/614352743?fromopenview=true&pq-origsite=gscholar&sourcetype=Scholarly%20Journals>

WALKER, J. How Big is Global Money Laundering? **Journal of Money Laundering Control**, v. 3, n. 1, 1 mar. 1999. ISSN 1368-5201. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/eb027208/full/html>

GLOSSÁRIO

AML - Anti-Money Laundering: Termo em inglês para PLD (Prevenção à Lavagem de Dinheiro) que compreende o conjunto de procedimentos, regulamentos e políticas implementados por instituições financeiras e outros setores para prevenir, detectar e denunciar indícios lavagem de dinheiro.

COS: Comunicações de Operações Suspeitas

KYC (Know Your Customer): Processo utilizado pelas instituições financeiras para verificar a identidade dos clientes e garantir a conformidade com as regulamentações.

COAF (Conselho de Controle de Atividades Financeiras): Órgão brasileiro responsável pela coordenação de ações para prevenir e combater a lavagem de dinheiro.

Relatório de Investigação Financeira (RIF): Documento produzido pelo COAF contendo informações detalhadas de transações suspeitas, enviado ao Ministério Público Federal, Ministério Público Estaduais, Polícia Federal, Polícias Civis Estaduais, estes utilizam para investigações criminais.

Machine Learning (ML): Subcampo da inteligência artificial que utiliza algoritmos para que sistemas computacionais aprendam a partir de dados e identifiquem padrões de forma automática.

Heurística: Processo cognitivo que permite aos especialistas tomar decisões rápidas com base na experiência e em padrões observados, sem análise exaustiva de todas as informações disponíveis.

Regras Duras: Conjunto de critérios objetivos definidos por regulamentações e políticas institucionais para detectar transações suspeitas de forma sistemática.

Regras de Bolso: Estratégias informais baseadas na experiência prática de especialistas para priorizar e classificar casos de risco.

Falsos Positivos: Casos em que transações legítimas são erroneamente classificadas como suspeitas.

Falsos Negativos: Casos em que transações ilícitas não são detectadas pelo sistema de monitoramento.

MIDIA_NEG (Exposição Negativa na Mídia): Variável utilizada para indicar se um indivíduo ou empresa possui histórico de exposição negativa em veículos de comunicação.

PROC_CRIM (Processos Criminais): Indicador de envolvimento prévio em processos judiciais relacionados a lavagem de dinheiro e crimes correlatos.

SOC_EMP (Sociedade Empresarial): Variável que indica se a transação está vinculada a uma empresa ou pessoa jurídica.

VALOR_ALERTADO: Variável referente ao montante financeiro envolvido na transação que gerou alerta.

ANEXO 1 - Termo de Consentimento para Participação em Pesquisa

Título da Pesquisa: Machine Learning Baseado na Combinação de Fatores Heurísticos e Racionais no Auxílio da Prevenção à Lavagem de Dinheiro.

Pesquisador Responsável: Ricardo Ferreira Lima | Mestrado Profissional em Administração | Insper- Instituto de Ensino e Pesquisa

Eu, [nome completo], [nacionalidade], [idade], [profissão], [número do RG ou CPF], estou sendo convidado/a para participar do estudo denominado "[título da pesquisa]", que tem como objetivo [descrição do objetivo da pesquisa]. A minha participação no referido estudo ocorrerá por videoconferência via software Microsoft Teams.

Introdução

Você está sendo convidado(a) a participar de uma pesquisa que visa avaliar se o uso de modelos de Machine Learning (ML) e Inteligência Artificial (IA) pode auxiliar na redução do índice de falsos positivos no processo de tratamento de alertas. Antes de decidir participar, é importante que você compreenda os objetivos, procedimentos e possíveis riscos e benefícios desta pesquisa. Este documento fornece essas informações.

Objetivos da Pesquisa: O objetivo principal desta pesquisa é identificar e detectar quais fatores os especialistas de lavagem de dinheiro utilizam para identificar sinais de lavagem de dinheiro.

Voluntariedade: A sua participação é totalmente voluntária. Você pode optar por não participar ou retirar seu consentimento a qualquer momento, sem qualquer penalidade ou prejuízo.

Riscos e Benefícios

- **Riscos:** Não são esperados riscos significativos associados à participação nesta pesquisa. As respostas serão mantidas em sigilo e utilizadas exclusivamente para fins acadêmicos.
- **Benefícios:** A pesquisa poderá contribuir para o desenvolvimento de métodos mais eficazes de detecção de crimes financeiros, beneficiando o setor financeiro como um todo.

Sobre Privacidade e Sigilo de Dados

Declaro estar ciente de que minha privacidade será respeitada. Meu nome e quaisquer dados que possam me identificar serão mantidos em sigilo e divulgados apenas com meu consentimento expresso. Caso sejam usadas informações identificáveis, este termo especificará claramente quais dados serão coletados, seu uso e como serão protegidos. Análises e divulgações dos dados coletados ocorrerão de forma não identificada, respeitando minha privacidade.

Contato: Em caso de dúvidas ou para mais informações sobre a pesquisa, você pode entrar em contato com o pesquisador responsável: **Ricardo Ferreira Lima** | Email: @al.insper.edu.br | **Telefone:** (11) 9740x-xxxxx

Declaração de Consentimento: Ao assinar este documento, você confirma que leu e compreendeu as informações acima, teve a oportunidade de fazer perguntas e está ciente de que pode retirar seu consentimento a qualquer momento. Você concorda voluntariamente em participar desta pesquisa.

Nome do Participante: _____

Assinatura do Participante: _____ **Data:** ____/____/2024

ANEXO 2 - Cenários para as entrevistas e roteiro

Introdução

Obrigado por participar desta entrevista. Nosso objetivo é entender melhor os fatores e métodos que você utiliza para identificar situações suspeitas de lavagem de dinheiro. Vamos discutir um exemplo específico de situação, e gostaríamos de ouvir suas opiniões e percepções sobre o processo de análise. Suas respostas serão mantidas em sigilo e usadas apenas para fins acadêmicos.

Exemplo de Situação 1

Imagine que você recebe um alerta na regra “**movimentação acima da capacidade financeira**” no qual um determinado cliente pessoa física, cuja movimentação financeira detectada foi de R\$679.700,00 em múltiplos créditos de diferentes contrapartes no período de 15 dias, o que aparentemente está muito acima de sua capacidade financeira, pois trata-se de um empregado registrado com salário mensal de R\$ 5.000,00.

Além das informações objetivas do alerta, você também recebe os dados do KYC, onde tem as seguintes informações:

Sexo: Masculino | **Idade:** 24 anos | **Estado Cível:** Solteiro | **PEP:** não
Renda Mensal: R\$ 5.000,00 | **Fonte de Renda:** Empregado CLT
Patrimônio Declarado: R\$20.000,00 (investimento em CDB)
Local de Residência: Umuarama - PR

Exemplo de Situação 2

Imagine que você recebe um alerta na regra “**movimentação em velocidade**” no qual um determinado cliente pessoa física, cuja movimentação financeira a crédito detectada foi de R\$340.500 em múltiplos créditos de diferentes contrapartes no período de 30 dias, seguidos de imediatas transferências a débitos no mesmo montante para diferentes contrapartes. Possui renda mensal informada de R\$ 12.000,00 e profissão autônomo.

Além das informações objetivas do alerta, você também recebe os dados do KYC, onde tem as seguintes informações:

Sexo: Masculino | **Idade:** 44 anos | **Estado Cível:** casado | **PEP:** não
Renda Mensal: R\$ 12.000,00 | **Fonte de Renda:** bureau
Patrimônio Declarado: R\$ 0
Local de Residência: São Paulo - SP

Exemplo de Situação 3

Imagine que você recebe um alerta na regra “**fracionamento em espécie**” no qual um determinado cliente pessoa Jurídica, recebeu 10 depósitos em agências diferentes no valor de R\$ 49.990 cada. Trata-se de um comercio varejista na região periférica da cidade do Rio de Janeiro. O mercado possui um faturamento de R\$ 100.000,00.

Além das informações objetivas do alerta, você também recebe os dados do KYC, onde tem as seguintes informações:

Data da Constituição: 01/02/2023

Atividade principal: COMÉRCIO VAREJISTA DE MERCADORIAS EM GERAL, COM PREDOMINÂNCIA DE PRODUTOS ALIMENTÍCIOS - MINIMERCADOS, MERCEARIAS E ARMAZÉNS.

1 sócio com idade de 35 anos

Faturamento: R\$ 100.000,00 anual

Local: Rio de Janeiro – RJ

Exemplo de Situação 4

Imagine que você recebe um alerta na regra “**movimentação inusitada**” no qual um determinado cliente pessoa física, recebeu 1 depósito de uma construtora no valor de R\$ 1.240.000,00. A movimentação nos últimos 6 meses a crédito na conta foi de R\$ 60.000,00. O cliente aplicou o montante recebido em CDB.

Além das informações objetivas do alerta, você também recebe os dados do KYC, onde tem as seguintes informações:

Sexo: Feminino | **Idade:** 32 anos | **Estado Civil:** casado | **PEP:** não

Renda Mensal: R\$ 10.000,00 | **Fonte de Renda:** bureau

Patrimônio Declarado: R\$ 0

Local de Residência: Recife -PE

Exemplo de Situação 5

Imagine que você recebe um alerta na regra “**movimentação em espécie**” no qual um determinado cliente pessoa Jurídica, recebeu 90% dos créditos em sua conta em espécie sem indícios de fracionamento de valores. O montante movimentado a crédito, em 6 meses, foi de R\$ 1.760.000,00. Os depósitos são sempre realizados no bairro próximo ao endereço da loja de revenda de carros. A revenda possui faturamento de R\$ 2.100.000,00.

Além das informações objetivas do alerta, você também recebe os dados do KYC, onde tem as seguintes informações:

Data da Constituição: 01/02/2021

Atividade principal: COMÉRCIO A VAREJO DE AUTOMÓVEIS, CAMIONETAS E UTILITÁRIOS USADOS.

2 sócios com idade de 56 anos e 24 anos

Faturamento: R\$ 2.100.000,00 anual

Local: Feira de Santana - BA

Perguntas Principais

- 1) Antes de iniciar a análise, você costuma avaliar o caso para identificar elementos que você considera mais relevante para determinar se há risco?
- 2) Já no processo de análise, pode descrever o passo a passo que você segue para investigar esta transação?
- 3) Ainda sobre o passo a passo, existe alguma ordem na qual você prioriza os diferentes critérios e sinais de alerta?
- 4) Quais são os principais fatores que você considera ao decidir se um alerta é suspeito ou não?
- 5) Além dos fatos já descritos, existem outras informações que você eventualmente procuraria para avaliar melhor esta transação?
- 6) Como sua experiência prévia influencia suas decisões atuais?
- 7) Pode compartilhar uma experiência passada onde você identificou um alerta suspeito? O que chamou sua atenção e qual foi o resultado?

Perguntas adicionais

- 8) Que tipo de treinamento ou conhecimento específico você considera essencial para identificar transações suspeitas?
- 9) Que sugestões você daria para melhorar o treinamento dos especialistas de lavagem de dinheiro?
- 10) Que melhorias você sugeriria para o processo atual de identificação de transações suspeitas?
- 11) Como o feedback dos especialistas pode ser melhor incorporado no desenvolvimento de novos sistemas de detecção?

Fechamento

Agradeço muito sua participação e suas valiosas percepções. Suas respostas ajudarão a melhorar os métodos de detecção de crimes financeiros e a tornar o processo mais eficiente. Se tiver alguma dúvida ou quiser adicionar algo, por favor, sinta-se à vontade para compartilhar