

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
FACULDADE DE ESTUDOS SOCIAIS
CURSO DE ADMINISTRAÇÃO

ANA LUIZA WEZEM RIBEIRO

FRAUDES EM ICOS:
UMA ABORDAGEM UTILIZANDO *MACHINE LEARNING*

MANAUS

2025

ANA LUIZA WEZEM RIBEIRO

FRAUDES EM ICOS:
UMA ABORDAGEM UTILIZANDO *MACHINE LEARNING*

Trabalho de Conclusão de Curso
apresentado ao Curso de Administração da
Universidade Federal do Amazonas (UFAM), como
requisito para obtenção do título de Bacharel em
Administração.

Orientador: Prof(a). Dr(a). Ana Cláudia de Araújo Moxotó

MANAUS
2025

ANA LUIZA WEZEM RIBEIRO

FRAUDES EM ICOS:
UMA ABORDAGEM UTILIZANDO *MACHINE LEARNING*

Trabalho de Conclusão de Curso
apresentado ao Curso de Administração da
Universidade Federal do Amazonas (UFAM) como
requisito parcial para obtenção do grau de Bacharel
em Administração.

Este trabalho foi defendido e aprovado pela banca em 12/12/2025.

BANCA EXAMINADORA

Prof.^a Dr.^a Ana Cláudia de Araújo Moxotó - UFAM
Orientadora

Prof. Dr. Márcio Palheta - UFAM
Avaliador

Prof. Dr. Daniel Armond de Melo - UFAM
Avaliador

Fraudes em ICOs : Uma abordagem usando *Machine Learning*

Ana Luiza Wezem Ribeiro

ana.ribeiro@ufam.edu.br

Ana Cláudia de Araújo Moxotó

anaclaudiaMoxotó@ufam.edu.br

Edjard de Souza Mota

edjard@icomp.ufam.edu.br

RESUMO

Este artigo examina se o aprendizado de máquina pode apoiar a detecção de fraudes em *Initial Coin Offerings* (ICOs). Utilizando uma base de 162 ICOs e 11 variáveis financeiras, informacionais e ESG, aplicamos o algoritmo de *clustering K-means* para calcular distâncias euclidianas até os centróides dos grupos e classificar como anomalias os projetos acima do 95º percentil. Os rótulos de fraude, derivados de atividade *on-chain* na rede Ethereum (API Etherscan), são então utilizados como *ground truth* para avaliar o desempenho do modelo. O algoritmo sinaliza nove ICOs anômalas, todas fraudulentas, resultando em alta precisão e ausência de falsos positivos, porém com sensibilidade muito baixa, pois apenas 9 das 130 fraudes são detectadas. A análise de importância das variáveis mostra que a detecção de anomalias é dominada pelo preço do token e pelos *hard* e *soft caps*, enquanto sinais informacionais e escores ESG desempenham papel residual. Os achados evidenciam limites de métodos não supervisionados baseados em distância e apontam para a necessidade de dados mais ricos e modelos supervisionados.

Palavras-chave: *K-means*; anomalias; ESG; *Blockchain*; Aprendizado de Máquina.

1. INTRODUÇÃO

As Ofertas Iniciais de Moedas (ICOs) emergiram como um mecanismo inovador de captação de recursos para novas empresas, caracterizadas pela emissão e venda de *tokens* digitais diretamente a investidores (Howell et al., 2020; Momtaz et al., 2019). Esses tokens, geralmente configurados como criptomoedas, são criados para operar no ecossistema do próprio empreendimento, funcionalidade possibilitada pela tecnologia de contabilidade distribuída (DLT) (Fisch, 2019). Especificamente, a blockchain funciona como um livro-razão público e descentralizado (P2P), onde todas as transações são registradas, rastreadas e validadas

coletivamente pela rede de forma imutável, graças à criptografia. Essa arquitetura técnica, fundamentada no consenso distribuído e no encadeamento inalterável de blocos, é o que garante a segurança, tornando os dados resistentes a fraudes, e a transparência, permitindo que o histórico seja auditável por qualquer participante (Pilkington, 2016).

As ICOs consolidaram-se como um mecanismo de financiamento com vantagens relevantes em relação às alternativas tradicionais. Em primeiro lugar, As ICOs proporcionam acesso rápido e global a capital, permitindo a captação de recursos de investidores ao redor do mundo. Esse modelo configura-se como uma forma de financiamento desintermediado, particularmente atraente para *startups* de *blockchain* por reduzir a dependência de intermediários financeiros (Momtaz, 2020). Além disso, os custos regulatórios e de emissão associados às ICOs tendem a ser menores do que aqueles observados em Ofertas Públicas Iniciais (IPOs) e no financiamento via *venture capital*, o que reduz barreiras de entrada para empresas jovens que buscam capital (Adhami et al., 2018).

No entanto, a ausência de regulamentação clara e a forte assimetria de informação tornam o ambiente de ICOs especialmente propenso a golpes e fraudes (Hornuf et al., 2022). Entre as práticas fraudulentas mais recorrentes destacam-se os esquemas Ponzi, nos quais os retornos são pagos com o capital de novos investidores, tornando o modelo insustentável no longo prazo (Yu et al., 2021), e os esquemas de *pump-and-dump*, baseados na manipulação artificial do preço dos *tokens* para geração de lucros rápidos em detrimento dos investidores que compram no pico de valorização (Li et al., 2021). Outras condutas incluem o desaparecimento dos emissores após a captação de recursos, a criação de equipes fictícias para conferir legitimidade ao projeto (Hornuf et al., 2022; Sapkota et al., 2020), o não pagamento de recompensas em programas de *bounty* (*bounty scam*) e o plágio de *whitepapers*, com cópias de documentos de projetos legítimos para simular credibilidade (Florysiak & Schandlbauer, 2019; Feng et al., 2019).

Estudos recentes estimam que a taxa de ICOs com características fraudulentas varia entre 10% (Tiwari et al., 2020), 26% (Hornuf et al., 2022) e 80% (Liebau & Schueffel, 2019), indicando um cenário de elevado risco para investidores. Além disso, a fraude em ICOs ocorre com frequência significativamente maior do que em ofertas públicas iniciais tradicionais, aproximadamente oito vezes mais (Lecompte, 2024). Nesse contexto, a identificação precoce de sinais de fraude torna-se crucial para a proteção dos investidores e para a construção de um mercado de criptoativos mais transparente e sustentável.

Ferramentas baseadas em inteligência artificial e aprendizado de máquina despontam, assim, como alternativas promissoras para a detecção de padrões indicativos de fraude em ICOs, ao explorar grandes volumes de informações financeiras, operacionais e textuais de forma automatizada (Karimov & Wójcik, 2021; Manful et al., 2024; Raghu et al., 2024). A relevância desta pesquisa decorre justamente da combinação entre a expansão das ICOs como mecanismo global de financiamento e a elevada incidência de fraudes em um ambiente pouco regulado, bem como do potencial da inteligência artificial para apoiar a triagem de projetos suspeitos e reduzir assimetrias informacionais que expõem, sobretudo, investidores de varejo a perdas expressivas.

Diante desse cenário, formulam-se as seguintes perguntas de pesquisa: Em que medida o algoritmo *K-means* (utilizado para detecção de anomalias) é capaz de identificar ICOs fraudulentas, e qual é a contribuição relativa das características financeiras, informacionais e socioambientais (ESG) para a formação dos padrões atípicos que sinalizam fraude?

O objetivo principal do estudo é determinar a contribuição relativa das características financeiras, informacionais e socioambientais (ESG) para a detecção de fraudes em Ofertas Iniciais de Moedas (ICOs), utilizando o algoritmo *K-means* como método de identificação de anomalias.

Além desta introdução, o trabalho está estruturado em quatro seções. A seção seguinte apresenta a revisão de literatura, abordando os fundamentos teóricos sobre ICOs, ESG e fraudes em ofertas de *tokens*. Na terceira seção, descrevem-se em detalhe a base de dados, as variáveis utilizadas e os procedimentos metodológicos adotados, com ênfase na aplicação do *K-means* e na estratégia de detecção de anomalias. A quarta seção expõe e discute os resultados empíricos obtidos. Por fim, a quinta seção traz as conclusões, destacando as principais contribuições teóricas e práticas, as limitações do estudo e sugestões para pesquisas futuras.

2. REVISÃO DE LITERATURA

2.1 ICOs e as dimensões financeira, informacional e ESG

O financiamento via ICOs expandiu-se rapidamente na última década, desafiando pesquisadores, investidores, empreendedores e reguladores (Adhami et al., 2018). Em uma ICO, novos empreendimentos baseados em *blockchain* levantam capital por meio da emissão de *tokens* digitais em troca de criptomoedas ou moeda fiduciária (Fisch, 2019; Tiwari et al., 2020). A *blockchain* funciona como um livro-razão distribuído em rede *peer-to-peer*, permitindo o registro seguro de transações e viabilizando a criação de ativos digitais escassos

(Chen, 2018). Entre 2015 e 2021, o mercado passou por um ciclo de rápida expansão, ajustes regulatórios e posterior recomposição, com forte concentração de volumes em jurisdições como Ilhas Virgens Britânicas, Cingapura, Estados Unidos e Suíça (Adhami et al., 2018; Andrieu & Sannajust, 2025).

O desempenho de uma ICO está associado a um conjunto de sinais financeiros e informacionais que reduzem a assimetria de informações percebida pelos investidores. ICOs estabelecem *soft cap* (mínimo) e *hard cap* (máximo) para assegurar viabilidade financeira do projeto. Se o mínimo não for atingido, os recursos são devolvidos, protegendo investidores e sinalizando credibilidade do projeto (Amsden & Schweizer, 2018). Evidências indicam que fatores como transparência na divulgação, código-fonte aberto, existência de *pre-sale* de *tokens*, desenho claro de direitos econômicos (*tokenomics*), listagem em *exchanges* e equipe fundadora experiente contribuem para o sucesso da oferta (Adhami et al., 2018; Howell et al., 2020; Gartner & Moro, 2024; Moxotó et al., 2025; Huang et al., 2020). Elementos informacionais, tais como *whitepapers* detalhados e presença ativa em mídias sociais, também aumentam a probabilidade de captação bem-sucedida (Campino et al., 2022; Meoli & Vismara, 2022). Ao mesmo tempo, a governança desse mercado permanece marcada por incerteza regulatória, vulnerabilidade a fraudes e fragilidades em segurança cibernética (Ivashchenko et al., 2018; Agarwal et al., 2024; Hornuf et al., 2022).

Paralelamente, a dimensão ESG vem ganhando relevância nas decisões de investimento, sendo frequentemente associada a menor risco e melhor desempenho financeiro de médio e longo prazo (Apergis et al., 2022; Zhou et al., 2022; dos Santos & Moxotó, 2024), ainda que persistam controvérsias quanto à mensuração e padronização das métricas (Berg et al., 2020). No contexto de ICOs, estudos recentes mostram que a sinalização de compromissos ESG pode elevar as avaliações iniciais e favorecer a captação, sobretudo quando combinada com boa governança, transparência e *tokenomics* bem estruturados (Bitetto & Cerchiello, 2023; Mansouri & Momtaz, 2022; Moxotó et al., 2025a). Evidências empíricas sugerem que startups com forte orientação ESG recebem avaliações mais favoráveis e podem captar até cerca de 28% mais recursos em ICOs, apresentando, adicionalmente, menor risco percebido pelos investidores (Mansouri & Momtaz, 2022; Moxotó et al., 2025a). À luz da Teoria dos *Stakeholders* (Freeman, 2010), empresas socialmente responsáveis tendem a ser percebidas como mais confiáveis, atraindo maior volume de investimento, o que reforça a hipótese de que práticas ESG se relacionam positivamente com o desempenho de ICOs, em especial em mercados desenvolvidos (Bitetto & Cerchiello, 2023).

2.2 Fraudes em ICOs

Após o surgimento das criptomoedas, o campo dos ativos digitais experimentou uma explosão repentina de interesse entre investidores institucionais. No entanto, no caso das *ICOs*, esse crescimento veio acompanhado de numerosos golpes, frequentemente marcados pelo desaparecimento de empresas após a captação de quantias significativas de recursos (Karimov & Wójcik, 2021). Hornuf et al. (2022) apontam a existência de diferentes tipos de fraude e constataam que *ICOs* fraudulentas apresentam, em média, altos volumes de captação.

Embora representem uma forma rápida e simples de captação de recursos em comparação às ofertas públicas tradicionais, as *ICOs* também se tornaram alvos frequentes de fraudes, com estimativas indicando que cerca de 10% dos fundos captados foram perdidos por esse motivo (Tiwari et al., 2020). Wats et al. (2024) realizam uma revisão sistemática das principais dimensões de pesquisa em *ICOs* e destacam as “zonas cinzentas” do instrumento. A baixa regulação, assimetria de informação extrema e incerteza jurídica que abrem espaço para fraudes e oportunismo por parte dos emissores. Nessa mesma linha, Bai e Zhang (2025) analisam características fundamentais das *ICOs* (qualidade da equipe, *white paper*, modelo de negócios, governança do token) e evidenciam que tais atributos ajudam a separar projetos mais sólidos de iniciativas potencialmente oportunistas, ainda que não permitam identificar diretamente a fraude.

Shifflett e Jones (2018) observam que essa escassez e opacidade de informação criam um potencial considerável para práticas fraudulentas. Chen (2019) revela que a apresentação de sinais informativos em *ICOs* desempenha papel central na redução da assimetria de informação: a clareza e a confiabilidade da fonte desses sinais influenciam significativamente tanto o sucesso da venda inicial de tokens quanto sua negociação subsequente. Tal evidência reforça a importância de uma comunicação transparente e acessível para mitigar os riscos de fraude.

Uma vertente recente da literatura busca sistematizar os sinais de alerta (*red flags*) de projetos fraudulentos. Lecompte (2024) propõe uma taxonomia de *red flags* em *ICOs*, organizada em dimensões como informações sobre a equipe e *governance*, características do *white paper*, promessas de retorno incompatíveis com o risco, estrutura econômica dos tokens e opacidade tecnológica, mostrando como esses elementos se combinam em projetos posteriormente identificados como fraudulentos. Em linha semelhante, Thewissen et al. (2025)

mostram que a complexidade lexical de *white papers* excessivamente técnicos ou obscuros pode funcionar tanto como sinal de sofisticação quanto como “cortina de fumaça” para ocultar riscos.

Entre os diversos tipos de fraudes em *ICOs*, Hornuf et al. (2022) destacam o papel ambíguo da divulgação de código em plataformas como o *GitHub* (ambiente em que startups podem enviar e compartilhar seu código-fonte). Por um lado, essa prática aumenta a transparência do projeto; por outro, torna o código mais suscetível a ataques de *phishing* e de hackers, o que evidencia um *trade-off* entre transparência e exposição a fraudes externas.

Segundo Liebau e Schueffel (2019), houve um aumento expressivo no número de golpes que combinam *spoofing* e *phishing* para oferecer *ICOs* falsas. Estima-se que aproximadamente 80% das *ICOs* em 2017 tenham sido identificadas como golpes. O *phishing* representa, em particular, uma ameaça considerável. Um exemplo notório é o caso da *ICO* The Bee Token, em que fraudadores conseguiram roubar cerca de US\$ 1 milhão: vítimas receberam e-mails de *phishing* e enviaram seus fundos para carteiras falsas. A empresa confirmou o ataque, suspeitando que os invasores obtiveram informações pessoais por meio de acesso ilegal a um fornecedor terceirizado (Hornuf et al., 2022).

O esquema Ponzi tradicional opera com a premissa de que investidores antigos recebem retornos oriundos do capital aportado por novos investidores. Esse ciclo é intrinsecamente insustentável e colapsa quando o fluxo de novos aportes diminui, resultando em perdas generalizadas (Yu et al., 2021). Outro tipo recorrente de fraude são os esquemas de *pump-and-dump* (*P&Ds*), que afetam o mercado de criptomoedas, incluindo as *ICOs*. Nesses casos, fraudadores inflacionam artificialmente o preço de um token por meio de informações falsas, a fim de vender, a preços mais elevados, tokens adquiridos inicialmente a baixo custo. Após a venda massiva, o preço despencar, ocasionando perdas significativas para os demais investidores (Li et al., 2021).

Conforme Hornuf et al. (2022) e Sapkota et al. (2020), um dos grandes riscos nas *ICOs* reside no fato de que, em muitos casos, a intenção primária dos emissores não é desenvolver um negócio viável e listá-lo no mercado secundário. Em vez disso, observa-se a prática em que desenvolvedores e promotores da *ICO* desaparecem abruptamente após arrecadar os fundos, abandonando os investidores sem qualquer informação — os chamados *exit frauds*. Soma-se a isso a apresentação de falsas equipes, com pessoas fictícias listadas como membros do projeto, utilizada para conferir aparência de legitimidade ao empreendimento.

Outra ferramenta associada a golpes são os chamados *bounty programs*, que prometem recompensas em tokens por atividades de divulgação realizadas por usuários (promoção em fóruns, mídias sociais etc.). Contudo, muitos projetos não honram essas promessas, levando à categorização como *bounty scam* quando há acusações de não cumprimento por parte dos “caçadores de recompensas” (Sapkota et al., 2020).

Por fim, o plágio de *white papers* constitui mais uma tática recorrente de fraude. Uma *ICO* geralmente tem início com a publicação de um *white paper*, documento principal que descreve o projeto, seus objetivos, roteiro, uso de recursos e equipe (Florysiak & Schandlbauer, 2019). A confiabilidade dessas informações é crucial, mas a ausência de regulação e supervisão implica que sua veracidade não é garantida (Feng et al., 2019). Golpistas exploram essa lacuna ao plagiar ou falsificar *white papers*, copiando trechos de projetos legítimos e bem-sucedidos para criar uma credibilidade artificial.

Dada a dificuldade inerente em prever fraudes com informações pré-emissão, torna-se crucial a intervenção de terceiros para certificar a qualidade dos emissores. Essa certificação pode ser realizada por plataformas especializadas ou pela participação de investidores institucionais e fundos de capital de risco, capazes de conduzir uma *due diligence* aprofundada para verificar a robustez do projeto (Hornuf, 2022).

ICOs são alvos frequentes de fraudes, com estimativas indicando que cerca de 10% dos fundos foram perdidos por esse motivo (Tiwari et al., 2020). Agarwal et al. (2024), revisam técnicas de *blockchain forensics* e mostram como rastreamento de fluxos on-chain, análise de endereços e heurísticas de *clustering* têm sido empregados para investigar fraudes em cripto, incluindo *fake ICOs* e esquemas de desvio de fundos após a captação.

Tabela 1 apresenta uma consolidação dos principais tipos de fraudes em ICOs, conforme a seguir:

Tabela 1: Tipos de Fraudes em ICOs

Tipos de Fraudes	Definição	Referência
Esquemas Ponzi	Modelo fraudulento em que os retornos de investidores antigos são pagos com recursos de novos aportes. O sistema colapsa quando cessam as entradas de novos investidores.	Yu et al. (2021).
Golpes “pump and dump”	Manipulação do preço de um <i>token</i> por meio de informações enganosas, permitindo lucro dos fraudadores, seguido de queda abrupta do valor.	Li et al. (2021).
<i>Phishing</i>	Fraude em que golpistas induzem vítimas a enviar fundos para carteiras falsas por meio de e-mails falsificados.	Hornuf et al. (2022).
“Exit fraud”	Os organizadores da ICO arrecadam recursos e desaparecem sem desenvolver o projeto, deixando os investidores sem retorno ou informação.	Hornuf et al. (2022).
“Bounty Scam”	ICOs promovem campanhas prometendo recompensas por divulgação do projeto e, após o trabalho ser realizado, não efetuam o pagamento prometido.	Sapkota et al. (2020).
Plágio de <i>whitepaper</i>	Fraude em que partes ou a totalidade do <i>whitepaper</i> são copiadas de outros projetos, apresentando ideias de terceiros como se fossem próprias.	Florysiak & Schandlbauer (2019).

Fonte: Dados da pesquisa

À luz da revisão de literatura, verifica-se que características financeiras, informacionais, de governança e socioambientais das ICOs funcionam como sinais importantes para distinguir projetos legítimos de potenciais golpes em um ambiente marcado por alta assimetria de informação, baixa regulação e múltiplos esquemas de fraude. Nesse contexto, o uso de algoritmos de aprendizado de máquina baseados em clusterização, como o *K-means*, permite identificar automaticamente padrões atípicos em grandes bases de dados, destacando ICOs com maior probabilidade de estarem associadas a fraudes. Baseado na revisão de literatura sobre o tema, apontamos as seguintes hipóteses:

H1: *ICOs que apresentam padrões atípicos em características financeiras, indicadores informacionais e métricas ESG têm maior probabilidade de serem classificadas como potenciais fraudes por métodos de aprendizado de máquina baseados em detecção de anomalias, em particular pelo algoritmo de clusterização K-means.*

H1a (dimensão financeira): *ICOs com padrões atípicos em características financeiras têm maior probabilidade de serem detectadas como potenciais fraudes pelo K-means, indicando predominância dessa dimensão em relação às dimensões informacional e ESG.*

H1b (dimensão informacional): *ICOs com padrões atípicos em indicadores informacionais (como ICO score, presença em redes sociais, MVP, GitHub e KYC) têm maior probabilidade de serem detectadas como potenciais fraudes pelo K-means, indicando predominância dessa dimensão em relação às dimensões financeira e ESG.*

H1c (dimensão ESG): *ICOs com padrões atípicos em métricas ESG têm maior probabilidade de serem detectadas como potenciais fraudes pelo K-means, indicando predominância dessa dimensão em relação às dimensões financeira e informacional.*

3. METODOLOGIA E ANÁLISE DE DADOS

Esta pesquisa buscou verificar a hipótese proposta na seção anterior. O objetivo do estudo é identificar padrões atípicos entre projetos de ICO e avaliar se tais anomalias, identificadas por métodos de aprendizado de máquina, podem indicar maior probabilidade de comportamento fraudulento. O estudo abrange 162 ICOs previamente selecionadas a partir dos dados da plataforma ICOmarkS (<https://icomarks.com/>), do período de 2017 até 2021.

A pesquisa é classificada como aplicada, de abordagem quantitativa e de caráter explicativo. A metodologia aplicada utiliza técnicas de ciência de dados para enfrentar um problema prático associado a um fenômeno social emergente: a detecção de fraudes em ICOs (Momtaz, 2020; Howell et al., 2020). O componente quantitativo se evidencia no uso de algoritmos de aprendizado de máquina empregados para detectar padrões atípicos e identificar potenciais anomalias nos dados (Tiwari, Gepp & Kumar, 2020). Por fim, o estudo assume caráter explicativo, uma vez que procura compreender como discrepâncias multivariadas nas características dos projetos podem estar associadas a um maior risco de fraude, em linha com pesquisas que analisam comportamentos irregulares em mercados cripto (Liebau & Schueffel, 2019).

3.1 Variáveis

A variável Fraude (FRD) é utilizada como rótulo de referência (ground truth) para avaliar o desempenho do modelo de detecção de anomalias, e não como variável dependente no algoritmo *K-means*, que é não supervisionado. As variáveis independentes estão organizadas em três grupos principais, em linha com as dimensões teóricas consideradas nas hipóteses do estudo. O primeiro grupo corresponde ao grau de aderência ESG (ESG), que captura práticas ambientais, sociais e de governança. O segundo reúne os sinais informacionais e de governança, representados por *Scores* (SCR), *Minimum Viable Product* (MVP), *Know Your Customer* (KYC), presença no Twitter (TWT), GitHub (GTH) e Reddit (RDD). O terceiro contempla as

características financeiras da oferta, expressas por *Tokens for Sale* (TFS), *Token Price* (TPR), *Soft Cap* (SCP) e *Hard Cap* (HCP), conforme sintetizado na Tabela 2.

Essa estruturação dos dados permite testar a hipótese geral (H1) por meio do cruzamento entre a classificação de anomalias gerada pelo *K-means* e o indicador de fraude (FRD), verificando se padrões atípicos nessas três dimensões estão associados a maior incidência de fraude. Da mesma forma, possibilita avaliar as hipóteses específicas (H1a, H1b e H1c), que comparam a contribuição relativa das dimensões financeira, informacional e ESG na detecção de ICOs potencialmente fraudulentas, sem recorrer a modelos de regressão, mas sim à análise da correspondência entre anomalias e FRD..

Tabela 2: Descrição das variáveis.

Variável	Sigla	Descrição	Escala de Medida	Fonte
Fraude	FRD	Indicador de indícios de comportamento fraudulento da ICO, obtido em sites especializados em criptomoedas	Dummy (0–1)	Etherscan
ESG	ESG	Mede o grau de aderência ESG das ICOs (<i>Environment, Social, Governance</i>)	0–0,1	Magalhães et al. (2024)
<i>Scores</i>	SCR	Avaliação de especialistas em ICOs	0–10	ICOMarkS
<i>Minimum Viable Product</i>	MVP	Indica a presença de protótipo funcional	Dummy (0–1)	ICOMarkS
<i>Know Your Customer</i>	KYC	Indica a existência de validação obrigatória da identidade do investidor	Dummy (0–1)	ICOMarkS
<i>Twitter</i>	TWT	Indica se o projeto possui conta ativa no <i>Twitter</i>	Dummy (0–1)	ICOMarkS
<i>GitHub</i>	GTH	Indica se o projeto disponibiliza o código no <i>GitHub</i>	Dummy (0–1)	ICOMarkS
<i>Reddit</i>	RDD	Indica se o projeto possui canal ativo no <i>Reddit</i>	Dummy (0–1)	ICOMarkS
<i>Tokens for Sale</i>	TFS	Número de tokens disponíveis para venda no ICO	Numérica	ICOMarkSdS
<i>Token Price</i>	TPR	Preço do token no momento da oferta inicial	USD	ICOMarkS
<i>Soft Cap</i>	SCP	Valor mínimo a ser arrecadado para a execução do projeto	USD	ICOMarkS
<i>Hard Cap</i>	HCP	Valor máximo permitido para arrecadação do projeto	USD	ICOMarkS

Fonte: Elaboração da Autora

3.1.1 Variável Fraude (FRD)

A identificação de fraudes a partir da movimentação na *blockchain* foi realizada com base em dados baixados por meio da Etherscan API, a partir do endereço do *smart contract* (*token contract*) informado para cada ICO na base de dados, e não por meio de buscas diretas no site. Inicialmente, criou-se uma conta no Etherscan para obtenção de uma *API key* e, em seguida, os dados de transações foram extraídos de forma automatizada por um script em Python. Esse script lia o endereço do *smart contract* de cada ICO e enviava requisições HTTP à Etherscan API, informando o endereço do contrato como parâmetro. A API retornava a lista de transações associadas a cada contrato; com base nesse retorno, o *script* contabilizava o número de movimentações para cada projeto.

A partir desses dados, definiu-se a variável referência Fraude (FRD) como uma *dummy* que mede a legitimidade operacional da ICO. Os projetos foram classificados como “Normais” (FRD = 0) quando os dados baixados via Etherscan API indicavam pelo menos uma transação associada ao endereço do *smart contract*, sinalizando que o contrato foi efetivamente utilizado na *blockchain*.

Em contraste, foram classificados como “Fraude” (FRD = 1) os casos em que nenhuma transação foi encontrada para o endereço informado, isto é, quando a API retornou uma lista vazia de movimentações. Essa ausência de atividade *on-chain* foi interpretada como forte indício de inoperacionalidade ou possível fraude, seja pela inexistência prática do contrato, pela completa inatividade do projeto ou pelo uso de um endereço de contrato falso ou divergente.

Ao recorrer exclusivamente a dados públicos, imutáveis e transparentes da *blockchain* Ethereum, obtidos de forma automatizada via Etherscan API, o procedimento estabeleceu um critério objetivo, replicável e auditável para identificar ICOs sem movimentação e classificá-las como casos suspeitos de fraude.

3.1.2 ESG e outras variáveis explicativas

No presente estudo, a variável ESG foi obtida diretamente da base construída por Magalhães et al. (2024), e não calculada neste estudo. Magalhães et al. (2024), estimam um *score* ESG para 6.513 startups de criptomoedas com base no método de Mansouri & Momtaz (2022), que mede características ESG a partir de textos dos próprios projetos (*white papers*, comunicados, sites, GitHub, Crunchbase etc.). Em termos práticos, Magalhães et al. (2024) utilizam o algoritmo em Python `ESG_Calculator()` (disponível em GitHub e em

SustainableEntrepreneurship.org), que conta a frequência de palavras associadas a temas ambientais, sociais e de governança em um modelo *bag of words*, usando dicionários específicos de ESG.

Assim, neste artigo, o índice ESG de cada ICO corresponde exatamente ao *score* previamente calculado por Magalhães et al. (2024), assegurando consistência e comparabilidade com a literatura recente sobre *ESG* em startups de criptomoedas.

As demais variáveis utilizadas no modelo foram extraídas diretamente da base da plataforma ICOmarks (www.icomarks.com), que fornece informações padronizadas sobre características das ofertas e sinais informacionais das ICOs.

3.2 Procedimento para Análise de Dados

3.2.1 Detecção de Fraudes com *K-means* e t-SNE

O processo de análise dos dados foi conduzido em três etapas principais: pré-processamento, normalização e aplicação dos métodos de agrupamento e detecção de anomalias. Todo o processo foi implementado em *Python*, utilizando o ambiente *Google Colab* e bibliotecas como *pandas*, *scikit-learn* e *matplotlib*.

Inicialmente, realizou-se a limpeza da base, tratando valores ausentes, inconsistências e formatos inadequados. Em seguida, as variáveis foram padronizadas por meio do *StandardScaler*, garantindo que todas contribuíssem de maneira proporcional no modelo, uma vez que diferenças de escala podem distorcer os resultados.

Com a base preparada, aplicou-se o algoritmo de clusterização *K-means* para identificar grupos de ICOs com características semelhantes. De acordo com Jain (2010), o *K-means* é um dos métodos de agrupamento (ou *clustering*) mais importantes e amplamente utilizados. Descoberto de forma independente por Steinhaus (1956), Lloyd (1982), Ball e Hall (1965) e McQueen (1967), sua simplicidade e eficácia o tornaram uma ferramenta fundamental em diversas áreas.

O algoritmo *K-means* inicia com a seleção aleatória de k centros iniciais, geralmente escolhidos entre as próprias observações do conjunto de dados, etapa crucial pois a escolha inicial pode influenciar o resultado do agrupamento (Arthur & Vassilvitskii, 2006). A partir daí, o algoritmo passa a operar de forma iterativa em duas fases. Na fase de atribuição, cada ponto é associado ao centro mais próximo, com base em uma medida de distância (tipicamente a

distância euclidiana). Em seguida, na fase de computar novamente dos centros, cada centro é atualizado para a média dos pontos que lhe foram atribuídos. Essas duas etapas são repetidas até que haja convergência, isto é, até que os centros deixem de sofrer alterações significativas e os agrupamentos se tornem estáveis.

Assim, a partir do agrupamento, foi calculada para cada projeto a distância Euclidiana até os centroides do *cluster* ao qual pertence. Essa métrica funciona como uma pontuação de anomalia: quanto maior a distância, mais o projeto se afasta do padrão médio do cluster. O uso da distância ao centróide como indicador de outliers é amplamente empregado em estudos de detecção de anomalias em ambientes de risco, especialmente em dados financeiros e de *blockchain*, pois permite capturar desvios estruturais que podem indicar comportamentos atípicos ou fraudulentos (Karimov & Wójcik, 2021; Tiwari, Gepp & Kumar, 2020).

Essa medida foi utilizada como indicador de anomalia, permitindo detectar ICOs cujo comportamento se desvia significativamente do padrão geral. Os casos que apresentaram distância acima do 95º percentil foram classificados como "potenciais fraudes" para fins de análise. Após a formação dos clusters pelo algoritmo *K-means*, foi calculada para cada projeto a distância Euclidiana em relação ao centróide do grupo ao qual foi atribuído. A utilização de limites baseados em percentis é recomendada em pesquisas de machine learning por reduzir a sensibilidade a ruídos e permitir uma detecção mais consistente de pontos extremos (Liu, Deng & Chen, 2021).

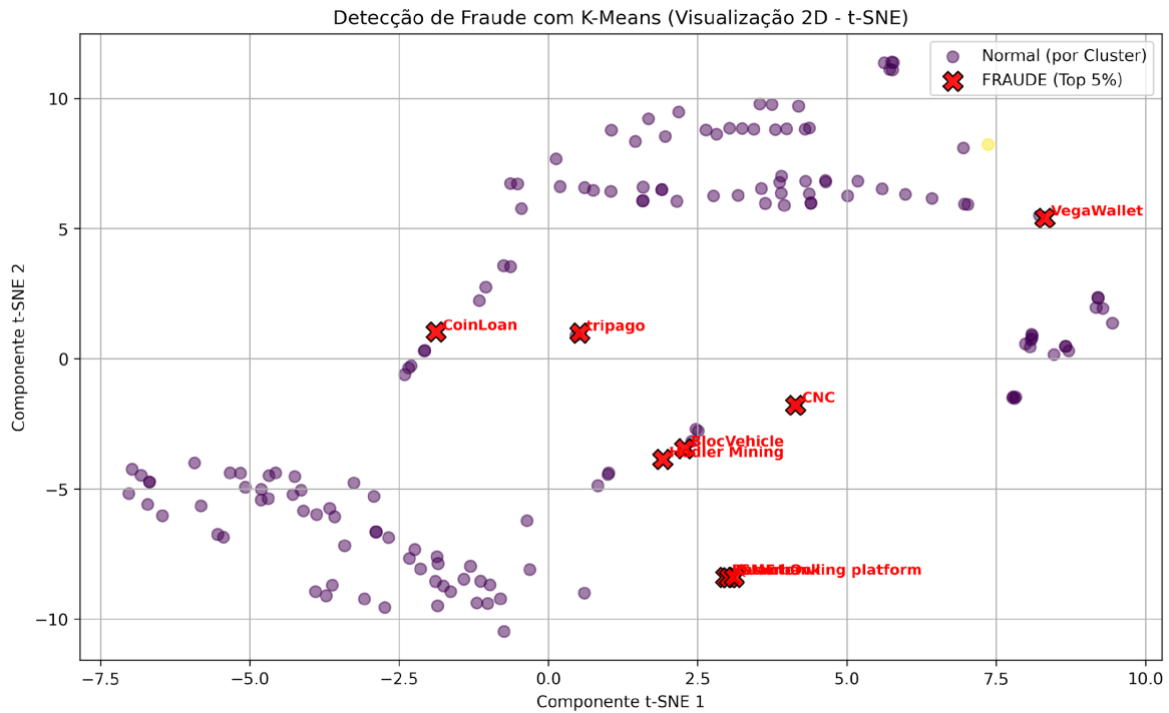
Além do cálculo das anomalias, empregou-se a técnica *t-SNE* (*t-Distributed Stochastic Neighbor Embedding*) para reduzir a dimensionalidade dos dados e representá-los em um espaço bidimensional. O *t-SNE* é amplamente utilizado em estudos que envolvem múltiplas variáveis porque preserva relações locais entre observações e evidencia padrões de agrupamento de forma visualmente clara (Van der Maaten & Hinton, 2008). Dessa forma, a técnica permitiu plotar graficamente os clusters gerados e destacar as ICOs classificadas como anômalas, facilitando a interpretação e comparação dos resultados.

4. RESULTADOS

Nesta etapa foram analisados os dados utilizando o modelo de clusterização, resultando na identificação de $k=2$ *clusters* distintos. A Figura 1 apresenta a visualização bidimensional do agrupamento *K-means* obtida a partir da técnica de redução de dimensionalidade t-SNE, o que permite observar como as ICOs se organizam em função de suas características. Os pontos destacados em vermelho representam as anomalias, identificadas a partir do cálculo da distância

Euclidiana em relação ao centróide de cada cluster. Esses projetos aparecem afastados dos agrupamentos principais, indicando comportamentos atípicos dentro da base analisada e sugerindo maior probabilidade de irregularidades ou risco de fraude.

Figura 1: Detecção de Fraude com *K-means* ($k=2$) (Visualização 2D – t-SNE)



Nota: Elaboração da Autora

Na base de 162 registros analisados, o modelo *K-means* identificou 9 ICOs como anomalias, correspondendo a 5,56% do total da amostra. Estas 9 observações representam os projetos cujas características apresentaram o maior desvio em relação ao comportamento médio dos dois *clusters*.

A Tabela 3 apresenta as ICOs identificadas como mais anômalas pelo modelo de *K-means*, ordenadas pela distância ao centróide do grupo ao qual foram atribuídas. Quanto maior essa distância, mais o comportamento da ICO se afasta do padrão médio dos demais projetos. Observa-se que a *CoinLoan* exibe a maior distância ao centróide (8,6573), indicando que é o projeto mais isolado e fora do padrão dentre todas as observações analisadas — e, coerentemente, é rotulado como fraude na base de dados.

De forma consistente, todas as ICOs listadas na tabela (Quantor, SME Banking Platform, Potion Owl, BlocVehicle, Hodler Mining, CNC, VegaWallet e Tripago) também

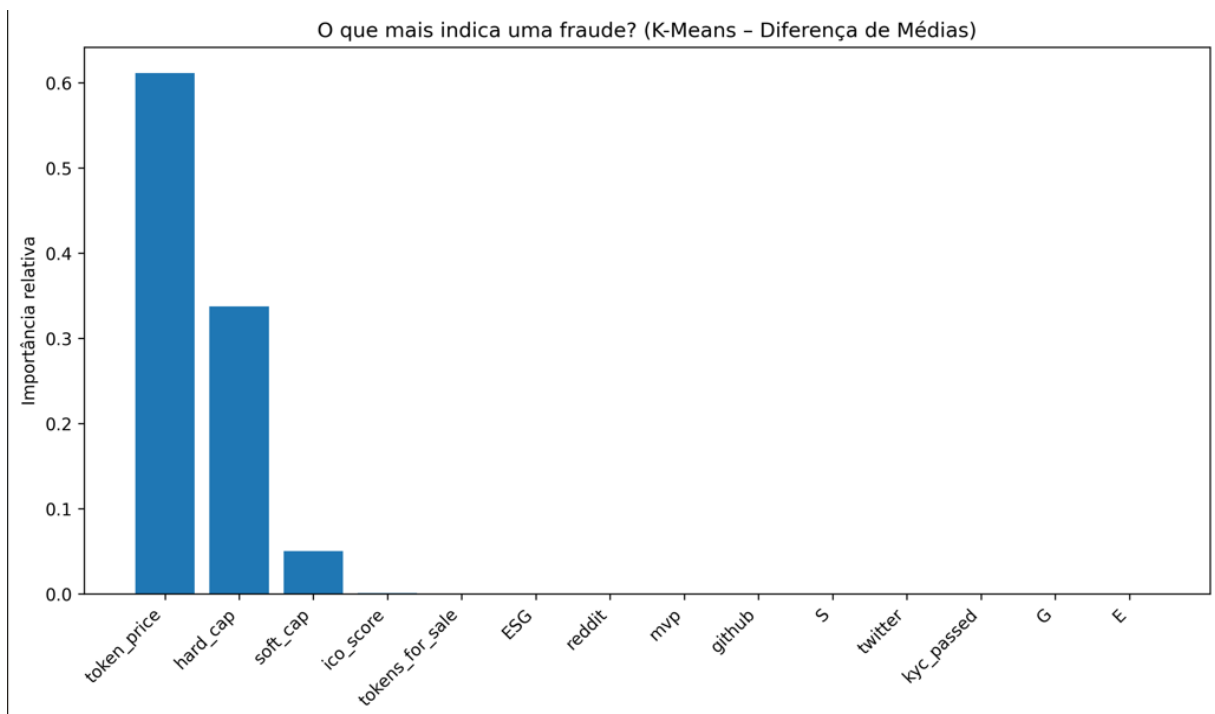
apresentam distâncias relativamente elevadas em relação ao centróide e são igualmente rotuladas como fraudulentas (Fraude = “sim”).

Tabela 3: Anomalias Detectadas por ICO

Nome da ICO	Distância ao Centróide	Fraude (Rótulo Real)
CoinLoan	8,6573	sim
Quantor	6,8637	sim
SME banking platform	6,7775	sim
Potion Owl	5,8936	sim
BlocVehicle	5,1697	sim
Hodler Mining	4,9727	sim
CNC	4,8695	sim
VegaWallet	4,7500	sim
Tripago	4,5991	sim

Nota: Elaboração da Autora

Figura 2: Fatores que influenciam na detecção de Fraudes



Nota: Elaboração da Autora

A Figura 2, intitulada “Fatores que influenciam na detecção de fraudes”, apresenta a importância relativa das variáveis na detecção de anomalias pelo algoritmo *K-means* e revela um descompasso claro entre as hipóteses formuladas e a evidência empírica. Em vez de padrões combinando dimensões financeira, informacional e ESG, a separação entre ICOs anômalas e normais é explicada quase inteiramente por três variáveis estritamente financeiras (*token_price*, *hard_cap* e, em menor grau, *soft_cap*). Esse resultado sugere que o modelo está essencialmente capturando *outliers* de dimensão financeira, isto é, projetos com preços de *token* e metas de captação muito distantes da média da amostra, em linha com evidências de que estruturas de oferta mais agressivas podem amplificar volatilidade e risco nas emissões de criptoativos (Wang et al., 2022), confirmando, assim, a hipótese H1a.

O índice ESG, por sua vez, apresenta contribuição praticamente nula, assim como as variáveis informacionais e de governança (*ICO_score*, *MVP*, *GitHub*, presença em redes sociais e *KYC_passed*). De um lado, isso indica que as fraudes da amostra não se distinguem sistematicamente nem pelo discurso ESG nem por sinais formais de transparência, o que coloca em xeque a premissa de que esses elementos, isoladamente, funcionariam como filtros robustos de risco. De outro, levanta dúvidas sobre a sensibilidade do próprio indicador ESG e sobre a capacidade da base de dados capturar nuances de risco socioambiental em projetos cripto. Em termos das hipóteses, os achados dão suporte parcial à hipótese geral H1, na medida em que padrões atípicos estão associados à identificação de fraudes, mas confirmam sobretudo H1a, ao mostrar a predominância da dimensão financeira, enquanto H1b e H1c não encontram respaldo empírico relevante para as dimensões informacional e ESG.

Do ponto de vista metodológico, essas evidências também expõem limitações importantes do *K-means* como ferramenta central de detecção de anomalias. Trata-se de um algoritmo baseado em distância euclidiana e centróides, fortemente sensível à presença de valores extremos em variáveis de alta variância. Em uma amostra relativamente pequena, alguns poucos projetos com *token_price* e *hard_cap* muito elevados tendem a dominar a formação dos agrupamentos, de modo que a estrutura de *clusters* se organiza quase exclusivamente nessas dimensões monetárias. Como consequência, o modelo não explora de forma adequada o espaço multivariado de risco proposto nas hipóteses H1b e H1c, reduzindo-se, na prática, a um detector de discrepâncias financeiras extremas, com baixa capacidade de incorporar simultaneamente informações ESG, sinais de governança e características operacionais das ICOs.

Para avaliar a eficácia do método de detecção, os casos apontados como anomalias pelo *K-means* foram cruzados com os rótulos de fraude pré-existentes na base, verificando-se que os projetos mais distantes do centróide correspondem, de fato, a ICOs previamente classificadas

como fraudulentas. Esse resultado oferece evidência adicional de que o modelo é capaz de capturar padrões atípicos associados a indícios de fraude.

Tabela 4: Matriz de Confusão - Detecção de Anomalias (*K-means*)

Classificação	Não fraude	Fraude	Total
Não anomalia	32 (Verdadeiro Negativo)	121 (Falso Negativo)	153
Anomalia	0 (Falso Positivo)	9 (Verdadeiro Positivo)	9
Total	32	130	162

Nota: Elaboração da Autora

A matriz de confusão da Tabela 4 mostra que o *K-means* não gerou falsos positivos: nenhuma ICO não fraudulenta foi classificada incorretamente como anomalia (0 casos), o que implica especificidade e precisão de 100% para a classe “anomalia”. Em contrapartida, observa-se que, das 162 ICOs da base, 130 são rotuladas como fraude, o que corresponde a aproximadamente 80% dos casos ($130/162 \approx 0,80$). Essa proporção evidencia um forte desbalanceamento de classes, com predominância de observações positivas (fraude) em relação às negativas (não fraude). Esse desbalanceamento reforça a interpretação da baixa sensibilidade do modelo: embora não gere falsos positivos, o *K-means* consegue identificar apenas 9 das 130 fraudes existentes, mostrando-se eficaz apenas para casos de anomalias mais extremas.

5. DISCUSSÃO

Os resultados empíricos mostraram que o algoritmo *K-means* foi capaz de identificar um subconjunto de fraudes mais extremas com alta precisão, sem geração de falsos positivos, mas com sensibilidade bastante limitada: apenas 9 das 130 ICOs fraudulentas foram classificadas como anomalias. Esse desempenho é compatível com a natureza não supervisionada do método e com a elevada heterogeneidade do mercado de ativos digitais descrita por Karimov e Wójcik (2021), que salientam a importância de amostras amplas e ricas em variáveis para capturar padrões de fraude mais complexos. Na configuração adotada, a hipótese geral H1 é apenas

parcialmente confirmada: padrões atípicos de fato aparecem associados à fraude, mas apenas em um grupo restrito de casos extremos.

A análise da importância relativa das variáveis indicou que a detecção de anomalias foi dominada por características estritamente financeiras, em especial *token_price*, *hard_cap* e *soft_cap*. Esse resultado está em linha com evidências de que o desenho econômico da oferta, sobretudo metas de captação e estrutura de preços, constitui um elemento central tanto para o sucesso quanto para o risco em *ICOs* (Amsden e Schweizer, 2018; Howell et al., 2020; Momtaz et al., 2019). Dessa forma, a hipótese H1a, que atribuía à dimensão financeira um papel predominante na detecção de fraudes, encontra respaldo empírico: o *K-means* captou principalmente *outliers* financeiros, isto é, projetos com preços de *token* e metas de captação muito distantes do padrão da amostra.

Em contraste, as dimensões informacionais e *ESG* apresentaram contribuição residual para a formação dos padrões atípicos, o que levou à não confirmação de H1b e H1c. Variáveis como *ICO_score*, existência de *MVP*, código no *GitHub*, presença em redes sociais e *KYC* praticamente não influenciaram a separação entre *ICOs* anômalas e normais. Do ponto de vista teórico, esse resultado dialoga com trabalhos que apontam tanto o potencial quanto os limites dos sinais informacionais na redução da assimetria de informação em *ICOs*. Embora *whitepapers* detalhados, reputação da equipe e engajamento em mídias sociais possam funcionar como sinais de qualidade (Chen, 2019; Bai e Zhang, 2025), a literatura também registra a capacidade de golpistas de mimetizar esses sinais e construir uma aparência de legitimidade (Lecompte, 2024; Thewissen et al., 2025). O fato de tais variáveis não diferenciarem, na amostra analisada, projetos fraudulentos de projetos legítimos sugere que, neste contexto, esses sinais foram amplamente apropriados por emissores oportunistas.

O desempenho nulo do índice *ESG* na detecção de anomalias é igualmente revelador. A literatura em finanças sustentáveis costuma associar orientação *ESG* a menor risco e melhor desempenho em horizontes mais longos (Apergis et al., 2022; dos Santos e Moxotó, 2024) e estudos recentes indicam que a sinalização de compromissos *ESG* pode elevar avaliações iniciais e volume captado em *ICOs* (Mansouri e Momtaz, 2022; Bitetto e Cerchiello, 2023; Moxotó et al., 2025a). Entretanto os resultados deste estudo sugerem que, na base analisada, o discurso socioambiental não se converteu em marcador robusto de risco de fraude. Esse descompasso reforça críticas à mensuração de *ESG* em ativos digitais (Berg et al., 2022; Magalhães et al., 2024) e indica que, em mercados pouco regulados, a retórica *ESG* pode ser usada de forma estratégica

para fins de *greenwashing*, sem necessariamente refletir compromisso efetivo com práticas sustentáveis.

Do ponto de vista metodológico, os achados evidenciam limitações importantes do *K-means* como ferramenta central de detecção de anomalias em *ICOs*. Trata-se de um algoritmo baseado em distância euclidiana e centróides, que tende a privilegiar discrepâncias extremas em variáveis de alta variância e grande escala. Em uma amostra relativamente pequena, alguns poucos projetos com *token_price* e *hard_cap* muito elevados acabam por dominar a formação dos agrupamentos, o que reduz a capacidade do modelo de explorar um espaço de risco multivariado que combine características financeiras, informacionais e *ESG*, conforme proposto nas hipóteses. Em termos mais gerais, os resultados confirmam que métodos não supervisionados baseados em distância tendem a capturar melhor fraudes extremas do que fraudes mais sutis. Isso reforça a conveniência de combinar abordagens de *clustering* com modelos supervisionados, como *Random Forest*, *XGBoost* ou redes neurais, e com técnicas de detecção de anomalias mais robustas, como *Isolation Forest* ou *One-Class SVM*, em pesquisas futuras.

As limitações dos dados também ajudam a explicar a baixa sensibilidade observada. Em primeiro lugar, o tamanho reduzido da amostra, composta por 162 *ICOs*, restringe a representatividade dos padrões de fraude e, como apontam Karimov e Wójcik (2021), compromete a capacidade de generalização de modelos aplicados a um mercado tão heterogêneo quanto o de *ICOs*. Em segundo lugar, o conjunto de variáveis é relativamente restrito, com 11 indicadores concentrados em características estruturais da oferta e em poucos sinais informacionais. A literatura recente mostra que elementos qualitativos, como riqueza e consistência do *whitepaper*, qualidade da equipe, intensidade de engajamento em redes sociais e indicadores textuais extraídos de documentos e plataformas *on-line*, aumentam de forma significativa o poder preditivo de modelos de fraude em *ICOs* (Feng et al., 2019; Hornuf et al., 2022). A ausência desse tipo de variável tende, portanto, a limitar a profundidade analítica da modelagem proposta.

Há ainda um problema estrutural ligado ao próprio mercado de *ICOs*. Em um ambiente com baixa ou nenhuma regulamentação, as informações divulgadas pelos projetos podem ser incompletas, inconsistentes ou manipuladas (Tiwari et al., 2020; Shifflett e Jones, 2018). A literatura sobre fraudes em *ICOs* destaca que a ausência de auditoria formal, supervisão contínua e mecanismos de verificação abre espaço para divergências significativas entre o que é prometido em *whitepapers* e materiais promocionais e o que, de fato, é executado após a captação (Hornuf et al., 2022; Sapkota et al., 2020). Essa assimetria informacional compromete a qualidade das

bases de dados disponíveis para pesquisa e limita o potencial de ferramentas estatísticas e computacionais na identificação de padrões reais de comportamento fraudulento.

Em conjunto, os resultados sugerem que o desempenho observado decorre da combinação de três fatores principais: amostra reduzida, escopo limitado de variáveis e baixa confiabilidade das informações em um mercado descentralizado e pouco regulado. Esses elementos contribuíram diretamente para a baixa sensibilidade do modelo e para a predominância da dimensão financeira na detecção de anomalias. Ao mesmo tempo, apontam para a necessidade de estratégias mais robustas, que articulem bases de dados maiores, variáveis mais ricas, incluindo dados textuais e *on-chain*, e técnicas avançadas de *machine learning*, combinando algoritmos de *clustering* com modelos supervisionados e ferramentas de *blockchain forensics*, de modo a aprimorar a detecção de fraudes em *ICOs*.

6. CONCLUSÕES

Esta pesquisa investigou a aplicabilidade de métodos de aprendizado de máquina não supervisionados, especificamente o algoritmo *K-means*, na detecção de padrões de fraude em Ofertas Iniciais de Moedas (ICOs). Adotou-se uma abordagem quantitativa baseada em técnicas de detecção de anomalias, utilizando uma base composta por 162 ICOs e 11 variáveis relacionadas às dimensões financeira, informacional e ESG. A amostra é fortemente desbalanceada, com aproximadamente 80% das observações rotuladas como fraude, o que torna o problema particularmente desafiador. Após o pré-processamento e a padronização dos dados, aplicou-se o *K-means* com distância Euclidiana, classificando como anomalias os projetos situados acima do 95º percentil.

Os resultados demonstraram que o *K-means*, em sua forma padrão, é estruturalmente ineficiente para a detecção de fraudes generalizadas neste domínio, embora seja altamente preciso para outliers extremos. O modelo identificou 9 projetos anômalos, todos rotulados como fraudes na base original, evidenciando 100% de precisão e ausência de falsos positivos. Contudo, a sensibilidade foi baixa, pois apenas 9 das 130 fraudes totais foram detectadas, de modo que o método se revelou eficaz apenas para fraudes altamente discrepantes. Em um contexto de forte desbalanceamento de classes e elevada assimetria informacional, esse desempenho é compatível com a literatura que aponta limitações de modelos não supervisionados na captura de padrões mais sutis.

A análise aprofundada permitiu identificar dois fenômenos críticos que invalidam o uso ingênuo de clusterização baseada em distância para este problema, constituindo a principal contribuição empírica deste trabalho:

1. Distorção por Outliers Financeiros: Confirmando a Hipótese H1a, a detecção foi dominada por características estritamente financeiras, em especial *token_price*, *hard_cap* e *soft_cap*. A magnitude extrema dessas variáveis em alguns projetos (escalas de trilhões) comprimiu a contribuição das demais dimensões no cálculo da distância Euclidiana, fazendo com que o algoritmo captasse principalmente outliers de engenharia financeira.
2. O Efeito Camuflagem (Falha de H1b e H1c): Contrariando as hipóteses iniciais, variáveis informacionais (como presença em redes sociais e código no *GitHub*) e o índice *ESG* tiveram contribuição residual e não se mostraram discriminatórias. Esse padrão prova que projetos fraudulentos conseguem mimetizar estatisticamente os sinais tradicionais de legitimidade e confiança, misturando-se à distribuição normal dos dados legítimos e anulando o poder isolado desses indicadores.

Portanto, conclui-se que a ineficiência do modelo não decorre apenas de limitações amostrais, mas da inadequação fundamental da distância Euclidiana linear para tratar dados financeiros exponenciais e da alta sobreposição de classes no espaço de características. A fraude em ICOs, na base analisada, se revelou mais um problema de "isolamento de pontos raros" do que de "formação de grupos" coesos.

Algumas limitações ajudam a contextualizar esse quadro: o tamanho reduzido da amostra, a restrição do conjunto de variáveis disponíveis, a baixa confiabilidade inerente às informações autorreportadas em um setor pouco regulado e, sobretudo, o forte desbalanceamento de classes. Esses fatores dificultam a captação de padrões complexos e reduzem a capacidade de generalização do modelo.

Para superar as limitações identificadas e avançar na detecção de fraudes em ICOs, pesquisas futuras devem redirecionar o foco metodológico nas seguintes frentes:

1. Pré-processamento Avançado: Aplicação obrigatória de transformações logarítmicas (log-scaling) nas variáveis financeiras para mitigar a distorção por outliers identificada, permitindo que outras dimensões contribuam para a análise.
2. Engenharia de Variáveis e Ampliação de Dados: Substituição de métricas brutas por razões financeiras mais informativas e remoção de variáveis redundantes de "confiança". Recomenda-se também a incorporação de variáveis textuais (via *Natural*

Language Processing de whitepapers), métricas avançadas de atividade on-chain e indicadores comportamentais, utilizando bases de dados ampliadas.

Mudança e Combinação Algorítmica: Substituição do *K-means* por algoritmos de detecção de anomalias teoricamente mais robustos para o problema, como *Isolation Forest* ou *One-Class SVM*, projetados para isolar pontos raros. Dado o desbalanceamento, a combinação destes com métodos supervisionados de alto poder discriminatório (como *Random Forest*, *XGBoost* ou redes neurais) deve ser prioritária para capturar padrões fraudulentos mais sutis.

Implicações Gerenciais e Regulatórias

Do ponto de vista prático, os resultados sugerem que métodos automatizados, mesmo com limitações, podem funcionar como ferramenta de triagem para priorizar projetos suspeitos para *uma due diligence* manual mais aprofundada. O alto número de fraudes na amostra reforça a necessidade de investidores adotarem critérios rigorosos e complementarem análises tradicionais com ferramentas computacionais. Em termos regulatórios, os achados destacam a urgência de maior transparência, padronização e supervisão mínima no mercado de ICOs. Medidas como validação de equipes, auditoria de contratos inteligentes e certificação de *whitepapers* são essenciais para reduzir assimetrias informacionais e fortalecer a integridade do ecossistema.

Em síntese, este estudo demonstra que, em um contexto de alta incidência de fraude e forte desbalanceamento, métodos de clusterização linear capturam sobretudo desvios financeiros extremos. A detecção mais abrangente e eficaz exige uma revisão do pré-processamento dos dados, a incorporação de novas fontes de evidência e, principalmente, a adoção de algoritmos projetados para isolar anomalias e combinar abordagens supervisionadas e não supervisionadas.

REFERÊNCIAS

- Agarwal, U., Rishiwal, V., Tanwar, S., & Yadav, M. (2024). Blockchain and crypto forensics: Investigating crypto frauds. *International Journal of Network Management*, 34(2), e2255. <https://doi.org/10.1002/nem.2255>
- Adhami, S., Giudici, G., & Martinazzi, S. (2018). Why do businesses go crypto? An empirical analysis of initial coin offerings. *Journal of Economics and Business*, 100, 64-75.
- Amsden, R., & Schweizer, D. (2018). Are blockchain crowdsales the new 'gold rush'? Success determinants of initial coin offerings. *Success determinants of initial coin offerings (April 16, 2018)*.
- Andrieu, G., & Sannajust, A. (2025). ICOs after the decline: A literature review and recommendations for a sustainable development. *Venture Capital*, 27(1), 1-19.
- Apergis, N., Poufinas, T., & Antonopoulos, A. (2022). ESG scores and cost of debt. *Energy Economics*, 112, 106186.
- Arthur, D., & Vassilvitskii, S. (2006). *K-means++: The advantages of careful seeding*. Stanford.
- Bai, Y., & Zhang, B. (2025). Fundamental analysis of initial coin offerings. *International Journal of Finance & Economics*, 30(1), 879–892. <https://doi.org/10.1002/ijfe.2948>
- Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315-1344.
- Bitetto, A., & Cerchiello, P. (2023). Initial coin offerings and ESG: Allies or enemies?. *Finance Research Letters*, 57, 104227.
- Ball, G. H., & Hall, D. J. (1965). ISODATA, a novel method of data analysis and pattern classification.
- Campino, J., Brochado, A., & Rosa, Á. (2022). Initial coin offerings (ICOs): Why do they succeed?. *Financial Innovation*, 8(1), 17.
- Chen, K. (2019). Information asymmetry in initial coin offerings (ICOs): Investigating the effects of multiple channel signals. *Electronic Commerce Research and Applications*, 36, 100858.
- Chen, Y. (2018). Blockchain tokens and the potential democratization of entrepreneurship and innovation. *Business Horizons*, 61(4), 567-575.
- Dos Santos, P. M. L., & Moxotó, A. C. D. A. (2024). ESG and risk. *Vie & sciences de l'entreprise*, 220(2), 53-74.
- Feng, C., Li, N., Wong, M. H., & Zhang, M. (2019). Initial coin offerings, blockchain technology, and white paper disclosures.

- Fisch, C. (2019). Initial coin offerings (ICOs) to finance new ventures. *Journal of Business Venturing*, 34(1), 1-22.
- Florysiak, D., & Schandlbauer, A. (2019, January). The information content of ico white papers.
- Freeman, R. E. (2010). *Strategic management: A stakeholder approach*. Cambridge University Press.
- Gartner, J., & Moro, A. (2024). C-level managers and born-digitals' scaling: The case of Initial Coin Offerings (ICOs). *Technological Forecasting and Social Change*, 198, 122943.
- Hornuf, L., Kück, T., & Schwienbacher, A. (2022). Initial coin offerings, information disclosure, and fraud. *Small Business Economics*, 58(4), 1741-1759.
- Howell, S. T., Niessner, M., & Yermack, D. (2020). Initial coin offerings: Financing growth with cryptocurrency token sales. *The Review of Financial Studies*, 33(9), 3925-3974.
- Huang, W., Meoli, M., & Vismara, S. (2020). The geography of initial coin offerings. *Small Business Economics*, 55(1), 77-102.
- Ivashchenko, A., Polishchuk, Y., & Britchenko, I. (2018). Implementation of ICO European best practices by SMEs. *Economic Annals-XXI*, 169(1-2), 67-71.
- Jain, A. K. (2010). Data clustering: 50 years beyond *K-means*. *Pattern recognition letters*, 31(8), 651-666.
- Karimov, B., & Wójcik, P. (2021). Identification of scams in initial coin offerings with machine learning. *Frontiers in Artificial Intelligence*, 4, 718450.
- Lecompte, A. (2024). The devil is in the details: a taxonomy of red flags of fraudulent initial coin offering projects. *SN Business & Economics*, 4(11), 128.
- Li, T., Shin, D., & Wang, B. (2021). Cryptocurrency pump-and-dump schemes.
- Liebau, D., & Schueffel, P. (2019). Cryptocurrencies & Initial Coin Offerings: Are they scams?-an empirical study. *The Journal of the British Blockchain Association*, 2(1).
- Liu, Y., Deng, Y., & Chen, S. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP Journal on Wireless Communications and Networking*, 2021(48).
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), 129-137.
- Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov), 2579-2605.
- MacQueen, J. (1967). Multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281-297).

- Magalhães, J. S. B., Moxotó, A. C. de A., & Mota, E. de S. (2024). The Impact of ESG on Startup ICO Fundraising: A Machine Learning Approach. *International Journal of Cryptocurrency Research*, 4(2), 112–125. <https://doi.org/10.51483/IJCCR.4.2.2024.112-125>
- Mansouri, S., & Momtaz, P. P. (2022). Financing sustainable entrepreneurship: ESG measurement, valuation, and performance. *Journal of Business Venturing*, 37(6), 106258.
- Manful, J. T., Hasford, A., & Yeluripati, G. R. (2024, October). Exploring The Use of Advanced Machine Learning to Detect Fraudulent Crypto Platforms. In *2024 IEEE 9th International Conference on Adaptive Science and Technology (ICAST)* (Vol. 9, pp. 1-8). IEEE.
- Meoli, M., & Vismara, S. (2022). Machine-learning forecasting of successful ICOs. *Journal of Economics and Business*, 121, 106071.
- Momtaz, P. P. (2020). Initial coin offerings. *Plos One*, 15(5), e0233018.
- Momtaz, P. P., Rennertseder, K., & Schröder, H. (2019). Token offerings: A revolution in corporate finance?.
- Moxotó, A. C. A., Soukiazis, E., & Melo, P. (2025). Determinants of success in initial coin offerings (ICOs): A systematic literature review. *Digital Business*, 100123.
- Moxotó, A. C. D. A., Soukiazis, E., & Melo, P. (2025a). The determinants of the Initial Coin Offering (ICO). A cross-country study. *Journal of Economy and Technology*.
- Pilkington, M. (2016). Blockchain technology: principles and applications. In *Research handbook on digital transformations* (pp. 225-253). Edward Elgar Publishing.
- Raghu, N., Kannanugo, N., Trupti, V. N., Ojashwini, R. N., Kiran, B., & Deepthi, M. (2024). Real-time fraud detection in crypto-currencies: Leveraging AI and blockchain. In *Applications of Blockchain and Artificial Intelligence in Finance and Governance* (pp. 28-66). CRC Press.
- Sapkota, N., Grobys, K., & Dufitinema, J. (2020). How much are we willing to lose in cyberspace? on the tail risk of scam in the market for initial coin offerings.
- Shifflett, S., & Jones, C. (2018). Buyer beware: hundreds of Bitcoin wannabes show hallmarks of fraud. *Wall Street Journal*, 17.
- Steinhaus, H., 1956. Sur la division des corp materiels en parties. *Bull. Acad. Polon. Sci. IV* (C1.III), 801–804
- Thewissen, J., Thewissen, J., Arslan-Ayaydin, Ö., & Herrezuelo, D. B. (2025). Hitting the right note: the impact of lexical complexity on initial coin offering success. *Financial Review*.
- Tiwari, M., Gepp, A., & Kumar, K. (2020). The future of raising finance-a new opportunity to commit fraud: a review of initial coin offering (ICOs) scams. *Crime, Law and Social Change*, 73, 417-441.

Wang, S., Cai, X., Guan, L., & Zhang, L. (2022). What do institutional investors bring to initial coin offerings (ICOs)?. *Transportation Research Part E: Logistics and Transportation Review*, 167, 102876.

Wats, S., Joshi, M., & Singh, S. (2024). Initial coin offerings: Current trends and future research directions. *Quality & Quantity*, 58(2), 1361–1387. <https://doi.org/10.1007/s11135-023-01701-z>

Yu, S., Jin, J., Xie, Y., Shen, J., & Xuan, Q. (2021). Ponzi scheme detection in ethereum transaction network. In *Blockchain and Trustworthy Systems: Third International Conference, BlockSys 2021, Guangzhou, China, August 5–6, 2021, Revised Selected Papers 3* (pp. 175-186). Springer Singapore.

Zhou, G., Liu, L., & Luo, S. (2022). Sustainable development, ESG performance and company market value: Mediating effect of financial performance. *Business Strategy and the Environment*, 31(7), 3371-3387.