



Universidade de Brasília - UnB

Faculdade de Economia, Administração, Contabilidade e Gestão de Pol. Públicas - FACE

Departamento de Administração - ADM

REVISÃO DE LITERATURA: ABORDAGENS DE DETECÇÃO DE ANOMALIAS EM SISTEMAS FINANCEIROS

SARAH OLIVEIRA PINTO

Brasília

2023

REVISÃO DE LITERATURA: ABORDAGENS DE DETECÇÃO DE ANOMALIAS EM SISTEMAS FINANCEIROS

Monografia apresentada ao Departamento de Administração - ADM como requisito parcial à obtenção do título de Bacharel em Administração.

Orientador: Prof. Dr. Vinicius Amorim Sobreiro.

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

Pinto, Sarah Oliveira
S243 Revisão de Literatura: Abordagens de detecção de anomalias em sistemas financeiros / Sarah Oliveira Pinto; orientador Vinicius Amorim Sobreiro. -- Brasília, 2023.
99 p.

Monografia(Graduação - Administração) -- Universidade de Brasília, 2023.

1. Detecção de anomalias. 2. Detecção de fraudes. 3. Detecção de outliers. 4. Sistemas financeiros. 5. Revisão sistemática de literatura. I. Sobreiro, Vinicius Amorim, orient. II. Revisão de Literatura: Abordagens de detecção de anomalias em sistemas financeiros.

DEDICATÓRIA

Eu dedico este trabalho à minha família, em especial, à minha mãe Laurinete.

SARAH OLIVEIRA PINTO

AGRADECIMENTOS

Agradeço ao meu orientador, Prof. Dr. Vinicius Amorim Sobreiro, por ter expandido os horizontes do meu conhecimento acerca do mundo das finanças, ajudando-me a encontrar aquilo que eu mais gosto de estudar. Agradeço a ele, também, por toda a disponibilidade, paciência e dedicação ao me ensinar e me guiar durante este trabalho. Ele abriu meus olhos para a magia do mundo acadêmico ao acreditar que eu iria conseguir mesmo quando eu enfrentei dificuldades.

Agradeço, especialmente, à minha mãe Laurinete que me acompanhou durante todos os processos da minha vida como, também, deste trabalho, sempre me dando amor e condições para realizar meus sonhos.

Agradeço ao meu pai Honório, que não conseguirá ler este trabalho, mas sempre me incentivou e acreditou em mim.

Agradeço à minha avó Ierinete por todo o diálogo, compreensão e apoio durante toda a minha trajetória acadêmica.

Agradeço ao também meu pai Adriano e à minha madrinha Marcia, que sempre estiveram ao meu lado e me deram o suporte necessário para concluir meus estudos.

Agradeço ao Thales que me acompanhou todos os dias, ouviu todas as minhas ideias e me impulsionou a ir cada vez mais longe.

Por fim, agradeço aos meus familiares Aldenete, Maria Clara, Cauê e Paulo Honório e aos meus amigos Guilherme, Laura, Clara, Caroline, Maria Fernanda, Paula, Giovanni e Alcilea que vibraram com cada uma das minhas conquistas que me trouxeram até o fim dessa etapa da minha vida.

SARAH OLIVEIRA PINTO

EPÍGRAFE

*Que nada nos limite,
que nada nos defina,
que nada nos sujeite.*

*Que a liberdade seja
nossa própria substância,
já que viver é
ser livre.*

Simone de Beauvoir
Filósofa francesa

RESUMO

Abordagens de detecção de anomalias se tornaram criticamente importantes para aprimorar sistemas de tomada de decisão, considerando o processo de redução de riscos na performance econômica de uma organização e os custos de consumidores. Estudos anteriores em detecção de anomalias examinaram, principalmente, as anormalidades que se traduzem em fraudes, como transações fraudulentas com cartões de crédito ou fraude em sistemas de seguro. No entanto, anomalias configuram irregularidades nos padrões de sistemas de dados, que podem ser provenientes de desvios, adulterações ou inconsistências, e o seu estudo abrange não somente fraudes, mas também, quaisquer anormalidades comportamentais que sinalizem riscos. Neste trabalho, é apresentada uma revisão de literatura de métodos e técnicas de detecção de anomalias em sistemas financeiros diversos utilizando a técnica dos cinco passos apresentada em Jabbour (2013, p. 145). Para tanto, um *framework* de classificação por meio de códigos foi elaborado para sistematizar as principais técnicas e conhecimentos acerca do tema, além de reconhecer oportunidades de pesquisa. Consequentemente, os resultados revelam diversas lacunas de pesquisa, dentre as quais três principais deveriam ser exploradas para o desenvolvimento dessa área: uma base de dados comum, testes com diferentes dimensões de dados e indicadores de efetividade dos modelos de detecção. Portanto, o *framework* proposto é pertinente para a compreensão de uma base de conhecimento científico já existente e, ainda, sinaliza importantes *gaps* para uma agenda de pesquisa considerando o tema de anomalias em sistemas financeiros.

Palavras-chave: Detecção de anomalias; Detecção de fraudes; Detecção de *outliers*; Sistemas financeiros; Revisão sistemática de literatura.

ABSTRACT

Anomaly detection approaches have become critically important to enhance decision-making systems, especially regarding the process of risk reduction in the economic performance of an organization and the consumer costs. Previous studies on anomaly detection have examined mainly abnormalities that translate into fraud, such as fraudulent credit card transactions or fraud in insurance systems. However, anomalies represent irregularities in system patterns data, which may arise from deviations, adulterations or inconsistencies. Further, its study encompasses not only fraud, but also any behavioural abnormalities that signal risks. In this work it is presented a literature review of methods and techniques to detect anomalies on diverse financial systems using the five-step technique presented on Jabbour (2013, p. 145). For that, it was elaborated a classification framework using codes to systematize the main techniques and knowledge on the subject, in addition to identifying research opportunities. Consequently, the results show several research gaps, among which three main ones should be explored for developing this area: a common database, tests with different dimensional sizes of data and indicators of the detection models' effectiveness. Therefore, the proposed framework is pertinent to comprehending an existing scientific knowledge base and signals important gaps for a research agenda considering the topic of anomalies in financial systems.

Keywords: Anomaly detection; Fraud detection; Outlier detection; Financial systems; Systematic literature review.

RESUMEN

Los enfoques de detección de anomalías se han vuelto de vital importancia para mejorar los sistemas de toma de decisiones, considerando el proceso de reducción de riesgos en el desempeño económico de una organización y los costos del consumidor. Estudios previos en detección de anomalías han examinado principalmente anomalías que se traducen en fraude, como transacciones fraudulentas con tarjetas de crédito o fraude en los sistemas de seguros. Sin embargo, las anomalías configuran irregularidades en los patrones de los sistemas de datos, que pueden provenir de desviaciones, manipulaciones o inconsistencias, y su estudio abarca no solo el fraude, sino también cualquier anomalía de comportamiento que señale riesgos. En este trabajo, presentamos una revisión bibliográfica de métodos y técnicas para la detección de anomalías en diferentes sistemas financieros utilizando la técnica de cinco pasos presentada en Jabbour (2013, p. 145). Para ello, se desarrolló un *framework* de clasificación a través de códigos para sistematizar las principales técnicas y conocimientos sobre el tema, además de reconocer oportunidades de investigación. En consecuencia, los resultados revelan varias brechas de investigación, entre las cuales se deben explorar tres principales para el desarrollo de esta área: una base de datos común, pruebas con diferentes dimensiones de datos e indicadores de la efectividad de los modelos de detección. Por lo tanto, el *framework* propuesto es relevante para la comprensión de una base de conocimiento científico existente y, aún, señala importantes *gaps* para una agenda de investigación que considere el tema de las anomalías en los sistemas financieros.

Palabras clave: Detección de anomalías; Detección de fraude; Detección de *outliers*; Sistemas financieros; Revisión sistemática de la literatura.

LISTA DE FIGURAS

2.1	Benefícios da transformação digital em serviços financeiros.	9
2.2	Processo de funcionamento e aplicação de Inteligência Artificial.	12
2.3	Como uma transação entra na <i>Blockchain</i>	15
2.4	Três gerações de <i>Blockchain</i>	16
2.5	Diferença entre <i>open banking</i> e bancos tradicionais.	17
2.6	Processo de segurança do sistema de pagamentos PIX.	19
2.7	Processos de <i>Machine Learning</i>	22
2.8	Distribuição logarítmica da frequência do primeiro dígito dos 4000 primeiros números da sequência de Fibonacci.	24
2.9	Relação entre o comportamento de estimadores e o aumento da dimensionalidade.	26
3.1	<i>Graphical abstract</i> da revisão de literatura.	44
4.1	Ciclo de implementação de modelos de análise de dados.	52
4.2	Contexto econômico.	57
4.3	Localização geográfica de pesquisa.	58
4.4	Método de pesquisa.	60
4.5	Área de aplicação.	62
4.6	Foco.	63
4.7	Base de dados utilizada.	65
4.8	Disponibilidade.	66
4.9	Técnica utilizada.	67
4.10	Número de estratégias.	69
4.11	Área do periódico.	71
5.1	Sugestões de pesquisas futuras.	78

LISTA DE TABELAS

2.1	Cenários de tecnologia relacionados à Inteligência Artificial.	II
2.2	Características do sistema de pagamentos PIX.	18
2.3	Abordagens de <i>Machine Learning</i>	23
2.4	Frequência de distribuição do primeiro dígito significativo dos primeiros 4000 números da sequência de Fibonacci.	23
3.1	Número de artigos selecionados da <i>ScienceDirect</i> [®]	33
3.2	Breve resumo dos artigos selecionados.	42
3.3	<i>Framework</i> de classificação e codificação.	46
4.1	Artigos categorizados com base no <i>framework</i> de classificação e codificação. . .	56
5.1	Proporções do primeiro e segundo dígito da Lei de Benford.	99

SUMÁRIO

DEDICATÓRIA	v
AGRADECIMENTOS	vii
EPÍGRAFE	ix
RESUMO	xi
ABSTRACT	xiii
RESUMEN	xv
LISTA DE FIGURAS	xvii
LISTA DE TABELAS	xix
SUMÁRIO	xxiii
1 INTRODUÇÃO	I
2 REVISÃO DA LITERATURA	7
2.1 A transformação digital nos sistemas financeiros e as tecnologias atuais	9
2.2 Inteligência Artificial	10
2.3 <i>Big Data</i> em finanças	12
2.4 Tecnologias financeiras	14
2.4.1 <i>Blockchain</i>	14
2.4.2 <i>Open Banking</i>	17
2.4.3 Pagamento instantâneo brasileiro (PIX)	18
2.5 Anomalias, fraudes e <i>outliers</i> no setor financeiro	20
2.6 Métodos de detecção de anomalias	21
2.6.1 Aplicação de <i>Machine Learning</i> em dados financeiros	21
2.6.2 Lei de Benford e a análise de anomalias financeiras	22
2.7 A maldição da dimensionalidade	25
3 MÉTODO	29
3.1 Procedimento	30
3.2 Coleta de dados	31
3.3 Classificação e codificação	43
3.3.1 Contexto econômico	46
3.3.2 Localização geográfica de pesquisa	47
3.3.3 Método de pesquisa	47
3.3.4 Área de aplicação	48
3.3.5 Foco	48

3.3.6	Base de dados utilizada	49
3.3.7	Disponibilidade	49
3.3.8	Técnica utilizada	49
3.3.9	Número de estratégias	49
3.3.10	Área do periódico	49
4	RESULTADOS E DISCUSSÕES	51
4.1	Contexto econômico	55
4.2	Localização geográfica de pesquisa	57
4.3	Método de pesquisa	60
4.4	Área de aplicação	61
4.5	Foco	62
4.6	Base de dados utilizada	64
4.7	Disponibilidade	66
4.8	Técnica utilizada	67
4.9	Número de estratégias	69
4.10	Área do periódico	70
5	CONCLUSÃO	73
	REFERÊNCIAS	79
	APÊNDICE A	97

CAPÍTULO I

INTRODUÇÃO

“Uma jornada de mil quilômetros
precisa começar com um simples passo.”

—Lao Tzu.

O grande volume de transações financeiras *online* e a elevada quantidade de registros correspondentes torna inviável que especialistas humanos consigam identificar anomalias presentes em sistemas financeiros (Gianini et al., 2020, p. 549) e, ainda, que consigam distinguir essas anomalias de um erro aleatório/sistemático que pode não estar diretamente relacionado às operações e produtos financeiros e, muitas vezes, só pode ser identificado por meio de testes de validação (Carneiro, Figueira, e Costa, 2017, p. 94; Climent, Momparler, e Carmona, 2019, p. 888). E, isso se torna cada vez mais complexo com a transformação digital que constantemente modifica serviços, modelos de negócios e, conseqüentemente, a vida das pessoas (Hanafizadeh e Kim, 2020, p. 1). Isso ficou ainda mais evidente com a chegada da pandemia de COVID-19, em 2020, que acelerou e transformou os serviços, incorporando inovações digitais a estratégias de diversos negócios (Shaikh, Sharma, e Karjaluto, 2020, p. 1).

Essa era de digitalização e transformação das organizações dentro da chamada 4ª Revolução Industrial incentiva o uso de tecnologias digitais para aprimorar a forma como processos são feitos, de modo a monitorar, diagnosticar, prever e otimizar dados e performances (Karki e Porras, 2021, p. 2). Ela possibilita o estabelecimento de um novo ecossistema em que os modelos de negócios digitais se tornam pilares da Indústria 4.0 (Shaikh et al., 2020, p. 1), devido à crescente chegada e uso de tecnologias emergentes, como *Big Data*, computação em nuvem, *Internet das Coisas*, Inteligência Artificial, *blockchain* e *criptomoeda*¹ (Hanafizadeh e Kim, 2020, p. 1).

No entanto, as necessidades e demandas humanas são a base principal que move os avanços em direção ao desenvolvimento tecnológico (Karki e Porras, 2021, p. 1). Nesse contexto, quando se trata de transações digitais, a criação de valor está em montar novas estruturas e novas formas de governança que consigam conectar as informações com as pessoas (Marikyan, Papagiannidis, Rana, e Ranjan, 2022, p. 2), uma vez que, por mais que as tecnologias digitais

¹Nota do autor: Uma forma de dinheiro digital.

sejam importantes no processo de transformação digital, é necessário que haja uma combinação das tecnologias com uma lógica de negócios que possa de fato inserir um modelo de negócios e a sua forma de gerenciamento à digitalização (Hanafizadeh e Kim, 2020, p. 1). Com base nisso, a digitalização de processos financeiros a partir de novas tecnologias possibilita o acesso otimizado e constante a ferramentas financeiras por empresas e indivíduos (Bollaert, de Silanes, e Schwienbacher, 2021, p. 1), que podem auxiliar na facilitação do processo de tomada de decisão em termos financeiros e operacionais, buscando prevenir e detectar anomalias financeiras de fato.

Porém, essa proximidade de contato com sistemas financeiros *online* também possibilita aos indivíduos a realização de tarefas transacionais com facilidade (Timmons, McGowan, e Lunn, 2019, p. 161), como pagamentos, transferências bancárias, aquisição de seguros, compra e venda de ativos financeiros ou retirada de previdências sociais. A crescente adoção e evolução das tecnologias financeiras modernas e das aplicações de tecnologia da informação, em conjunto com o desenvolvimento de inteligências artificiais e das inovações automatizadas (Cho e Chen, 2021, p. 2), em atividades diárias cria também vulnerabilidades em rápida evolução, expondo os sistemas a ataques cibernéticos e maliciosos (Wessels, van den Brink, Verburgh, Cadet, e van Ruijven, 2021, p. 1). Desse modo, por mais que tecnologias financeiras tornem os serviços financeiros e processos mais eficientes por meio da aplicação de novas formas de tecnologia (Thompson, 2017, p. 152), a flexibilidade proveniente desses sistemas *online* revela uma fragilidade que pode ser contrastada com a crescente necessidade sistemas de detecção de fraudes e anomalias (Gianini et al., 2020, p. 549).

O desenvolvimento e a disseminação de computadores pessoais, *tablets* e telefones celulares podem exercer uma grande influência na vida dos consumidores (Albayati, Kim, e Rho, 2020, p. 1) e dar margem a essa vulnerabilidade financeira provocada pela digitalização das transações, principalmente considerando que, desde os anos 2000 o número de usuários de celulares cresceu exorbitantemente (Zaffar, Kumar, e Zhao, 2019, p. 62; Verbeke, Dejaeger, Martens, Hur, e Baesens, 2012, p. 211) e as aplicações de telefone são muito utilizadas pelo mundo (Sokolova, Perez, e Lemercier, 2017, p. 62), atenuando as barreiras de entrada (Shaikh et al., 2020, p. 1) e possibilitando aos usuários verificar balanços, depósitos, e realizar transações diversas (Zaffar et al., 2019, p. 62). O chamado *mobile money*², um serviço de pagamento de tecnologia financeira, modificou o modo de usar os celulares, não se limitando apenas à comunicação, entretenimento e navegação na *internet*, de modo que não é mais necessário ter contato direto ou dinheiro físico para conduzir transações financeiras e os serviços podem ser acessados em qualquer lugar e a qualquer momento (Daragmeh, Lentner, e Sági, 2021, p. 1). Um exemplo disso é o PIX, no contexto brasileiro, uma forma de pagamento instantâneo criado pelo Banco Central (A. R. Ferreira, 2022, p. 101) em que recursos podem ser transferidos entre contas bancárias de modo rápido e seguro em qualquer período do dia (Banco Central, 2020b, p. 1). Desse modo, diversas plataformas e serviços financeiros foram lançados ao redor do mundo, resultando em mais de 556 milhões de contas de *mobile money* registradas (GSMA, 2017, p. 17).

Turner (2021, p. 1), em um estudo feito em junho de 2021 pela plataforma digital *BankMyCell*, indica que o número de usuários de *smartphones* ao redor do mundo abrange 48,20% de toda a população. Por ser um formato de telefone que possui funcionalidades avançadas, com um sistema operacional inteligente e aplicativos diversos, o *smartphone* possibilita a conexão de redes de dados com a *internet*, o que é um atrativo. Ainda nessa linha, Handley (2019, p. 1) produziu um relatório, pela *World Advertising Research Center*, sinalizando que aproximadamente 51% de todos os usuários de *smartphones* acessam a *internet* apenas por meio de seus próprios aparelhos e, até 2025, 72,6% dos usuários de *internet* em geral irão aderir ao acesso da *internet* somente pelos *smartphones*, totalizando o equivalente a aproximadamente 3,7 bilhões de pessoas. A conveniência de poder utilizar a *internet* facilmente, desse modo, potencializa as

²Em tradução livre: Dinheiro móvel.

interações sociais, a realização *online* de transações financeiras e o aumento de produtividade, no entanto, esse uso significativo gera preocupações com a segurança de dados e informações pessoais, principalmente de teor financeiro (Jones e Chin, 2015, p. 561).

Dado esse aumento expressivo de interações, as últimas décadas foram marcadas pelo grande crescimento do uso de plataformas de *e-commerce*³ tanto por organizações privadas, quanto por agências do governo (Misra, Thakur, Ghosh, e Saha, 2020, p. 254). Em função disso, essas plataformas possibilitaram ainda mais a conexão entre a facilidade de acesso à *internet* por meio dos *smartphones* com a opção de realizar transações financeiras de modo virtual. A crescente popularidade dessas novas tecnologias emergentes criam, assim, oportunidades para que fraudadores possam causar prejuízos (Misra et al., 2020, p. 254). Os *smartphones*, conseqüentemente, podem ter sido facilitadores de fraudes e desvios de dados, e o entendimento e identificação das anomalias provocadas em virtude disso podem ser de fundamental importância para a criação de um sistema estratégico de tomada de decisões, uma vez que métodos de identificação de anomalias possibilitam o aprendizado de características de redes de dados com o propósito estratégico de distinguir anormalidades (Stripling, Baesens, Chizi, e vanden Broucke, 2018, p. 15), e essas abordagens se convertem em fundamentações teóricas relevantes para estimar probabilidades e prever fraudes, validando um sistema de suporte a decisões mais estruturado (Serrano-Silva, Villuendas-Rey, e Yáñez-Márquez, 2018, p. 78).

O desenvolvimento de um sistema eficiente de detecção de fraudes e anomalias diversas é importante, também, para a diminuição de perdas em instituições financeiras (Höppner, Baesens, Verbeke, e Verdonck, 2022, p. 291), uma vez que modelos de gestão negócios bem definidos na área financeira viabilizam que as organizações possam se beneficiar de suas forças internas, bem como, monitorar e aproveitar oportunidades de melhorias (Hanafizadeh e Marjaie, 2021, p. 1). Por mais que empresas financeiras em geral, muitas vezes encorajadas por lei, são mais suscetíveis a possuírem processos e sistemas que assegurem que os seus dados são autênticos e de qualidade (Elsaify e Hasan, 2021, p. 2), com a cada vez maior entrada do setor financeiro no *digital business*, é necessário estabelecer mecanismos mais rígidos de acesso e controle desses dados, de modo a revisar cada ponto e cada transação para garantir que não surjam comportamentos fraudulentos (Guo e Liang, 2016, p. 11). Através de abordagens mais inteligentes, é possível criar sistemas reguladores que considerem o acesso, a análise e o processamento de dados em tempo real (Barroso e Laborda, 2022, p. 13) e, por isso, é de suma importância reunir informações sobre as técnicas existentes na base de conhecimento científico, viabilizando que pesquisadores e gestores possam identificar as melhores técnicas conforme o cenário em que se encontram, podendo assim, explorar suas capacidades a fim de obter valor e mitigar riscos (Konopik, Jahn, Schuster, Hoßbach, e Pflaum, 2022, p. 3).

Os grandes desvios chamados de anomalias ou anormalidades, de maneira pormenorizada, representam inconsistências com base em paradigmas econômicos ou expectativas de comportamento, sendo eles considerados anormais (Parnes, 2020, p. 292). Logo, essas anomalias podem ser definidas como um conjunto de dados que se comportam de acordo com padrões diferentes daqueles que são identificados como normais ou típicos, sendo considerados exceções ou peculiaridades (León, 2020, p. 1; Rubio, Barucca, Gage, Arroyo, e Morales-Resendiz, 2020, p. 2; Stripling et al., 2018, p. 15; N. Wu e Zhang, 2006, p. 375). Esses *outliers* nos dados podem ser retratados como fraudes, uma vez que elas representam uma alteração criminosa no comportamento de um grupo de dados (Pourhabibi, Ong, Kam, e Boo, 2020, p. 7). Com base nesse contexto, métodos de detecção de anomalias são importantes processos de *Data Mining* para identificar tendências e padrões em diversos conjuntos de dados financeiros (Parnes, 2020, p. 308) e o desenvolvimento de pesquisas que possam aprimorar tais métodos é de suma importância para a diminuição de perdas financeiras, uma vez que as organizações dependem cada vez mais de dados para tomar decisões operacionais e estratégicas (Elsaify e Hasan, 2021, p. 2)

³Modalidade de comércio cujas transações são feitas de modo *online*.

e, para desenvolver melhor seus modelos de negócio, gestores precisam atentar-se a sinais alarmantes que possam aparecer de modo a tomarem ações necessárias (Koyuncugil e Ozgulbas, 2012, pp. 6238–6239), implementando técnicas e ferramentas que possam auxiliar ainda mais nesse processo (Barroso e Laborda, 2022, p. 11).

Diante do cenário atual, destaca-se a importância da identificação de anomalias e/ou fraudes nas mais diversas áreas do conhecimento, uma vez que tal processo possibilita fazer frente às situações de irregularidades ou anormalidades provenientes de desvios ou adulteração de dados. Embora essa questão seja revestida de suma importância nos últimos anos, ela não é recente e vem sendo discutida com bastante regularidade nos meios acadêmicos. A título de exemplo, na área da medicina, M. Chen, Kanade, Pomerleau, e Rowley (1999, p. 123) estudaram métodos de detecção de anomalias no corpo humano, que configuravam patologias, já Duftscheid e Miksch (2001, p. 39) abordaram o tema em erros de diretrizes clínicas, que poderiam resultar em erros de aplicação, e Smith, Dixon, Tays, e Bulman-Fleming (2004, p. 458) em identificações visuais de problemas neurológicos.

Ainda a título de exemplo, na área de sistemas computacionais, Waite (1999, p. 221) visou distinguir performances de dados anormais com técnicas probabilísticas, enquanto Saleh, Boujarwah, e Al-Dallal (2001, p. 979) analisaram anomalias de sincronização no desenvolvimento de *softwares*, e o estudo de Perdisci, Ariu, Fogla, Giacinto, e Lee (2009, p. 864) teve enfoque no uso de métodos de detecção de anomalias para prevenir ataques em redes computacionais. Na área de sistemas de rede, Muniyandi, Rajeswari, e Rajaram (2012, p. 174) e N. Wu e Zhang (2006, p. 376) discutiram sobre anomalias de modo a identificar interferências e obstruções. Yasami e Safaei (2017, p. 52), Kaur e Singh (2016, p. 213) e Savage, Zhang, Yu, Chou, e Wang (2014, p. 62) fizeram pesquisas sobre identificação de anomalias em redes sociais de modo a encontrar atividades anormais. Em uma perspectiva mais recente, Ardizzi, Franceschis, e Giammatteo (2018, p. 105), K. Singh e Best (2019, p. 2) e Rtayli e Enneya (2020, p. 13) trataram de técnicas de detecção de anomalias em sistemas financeiros, abordando questões como *outliers* em sistemas de pagamentos, indícios de lavagem de dinheiro e fraudes em transações financeiras. Isso mostra como essa temática continua em voga, sendo importante e caracterizando uma tendência de pesquisa.

Dentre os artigos analisados a respeito desse tema de estudo, pode-se perceber que há a existência de problemas não resolvidos ou não considerados em pesquisas anteriores, que podem ser relevantes para o presente trabalho. De maneira mais precisa, considerando artigos de revisão de literatura publicados recentemente, pode-se observar que os métodos de análise escolhidos consistiram em revisar e examinar técnicas de *Data Mining* para detecção de fraudes financeiras e classificá-las conforme os tipos de fraude (Al-Hashedi e Magalingam, 2021, p. 5), identificar e analisar pesquisas que usam técnicas no estilo *Graph-Based Anomaly Detection (GBAD)*⁴, baseadas em gráficos visuais, no contexto de detecção de fraudes financeiras (Pourhabibi et al., 2020, p. 1) e explorar vantagens e desvantagens de métodos que possam revelar comportamentos fraudulentos e desonestos em organizações (Houdek, 2020, p. 1), conforme apresentado a seguir:

It reveals the most frequent financial fraud detection technique used in data mining that can result in disclosing fraud in financial domains with high-performance accuracy and provides a new classification framework for financial fraud as well as creating a roadmap for researchers and practitioners seeking to better understand this field.

Al-Hashedi e Magalingam (2021, p. 3).

We focus on research undertaken between 2007 and 2018 to provide a synthesized understanding of the state-of-the-art GBAD methods, identify key research issues raised against application contexts, and establish future directions to expand GBAD research in fraud detection.

Pourhabibi et al. (2020, p. 2).

This article supplies an interdisciplinary and multilevel overview of methods how to identify (not

⁴Em tradução livre: detecção de anomalias com base em gráficos.

only) organizational misdeeds, the pros and cons of the use of each individual method, and the limitations and strengths of their results based on cherry-picked current studies.

Houdek (2020, p. 1).

É oportuno destacar que, embora existam tais sistematizações de literatura, a abordagem utilizada em todas é voltada, principalmente, à detecção de fraudes financeiras, não considerando especificamente a identificação de anomalias presentes nos sistemas, que podem configurar tais fraudes ou não, como, por exemplo, podem indicar transações financeiras ilegais de lavagem de dinheiro (Ardizzi et al., 2018, p. 106) por meio de padrões de comportamento anormais nos dados. Observando esse contexto, *o objetivo nessa revisão de literatura é sistematizar o conhecimento a respeito de detecção de anomalias e/ou fraudes no setor financeiro de modo a consolidar as principais técnicas utilizadas nessa área e identificar lacunas existentes.* Para a realização dessa pesquisa, foi empregada a técnica dos cinco passos utilizada por Jabbour (2013, p. 145) e Lage Junior e Godinho Filho (2010, p. 14) e, posteriormente, excelentemente utilizada em outros *papers* (Amui, Jabbour, de Sousa Jabbour, e Kannan, 2017, p. 311; Barboza, Kimura, Sobreiro, e Basso, 2016, pp. 328–329; Ferraz, Falguera, Mariano, e Hartmann, 2021, p. 7; S. Singh et al., 2019, p. 164; Stefanelli, Teixeira, Oliveira, Ferreira, e Sehnem, 2019, pp. 2050–2051; Talan e Sharma, 2019, p. 355). De maneira resumida, tais autores recomendam classificar e codificar características relacionadas a artigos relevantes, construir breves resumos das contribuições discutidas, analisar e descrever forças, fraquezas e os métodos abordados por cada um, criar um *framework* dos artigos analisados e, ainda, fornecer uma agenda de pesquisas futuras pertinente ao tema.

Desse modo, essa sistematização, poderá contribuir para a identificação de técnicas e resultados relevantes para essa área de estudo na perspectiva de um tomador de decisões, uma vez que sistemas de suporte a decisões, os quais auxiliam na predição de riscos e fraudes por meio da detecção de anomalias, possuem direto impacto na performance econômica de uma organização (Severino e Peng, 2021, p. 2). Com base nisso, um gestor precisa escolher as técnicas e os parâmetros de intervenção tecnológicos mais pertinentes a serem aplicados aos seus sistemas, de modo que a gestão financeira possa ser feita de maneira mais eficiente por meio de métricas de risco robustas e coerentes (Byers, Popova, e Simkins, 2021, p. 1) que assegurem a continuidade de seus processos, reforçando a segurança cibernética (Wessels et al., 2021, p. 1), evitando que a organização seja alvo de atos ilegais ou erros e desvios transacionais e, assim, viabilizando a sobrevivência organizacional (Konopik et al., 2022, p. 1).

Assim, em termos amplos, a técnica dos cinco passos utilizada neste trabalho pode sinalizar as abordagens de detecção de anomalias financeiras correlacionando-as a diferentes fatores conforme as classificações predefinidas que irão conduzir a análise descritiva (Cabral e Dhar, 2019b, p. 2245), de modo que futuros pesquisadores poderão encontrar uma base de dados e estatísticas para inspirar-se em seus estudos a serem desenvolvidos (Mallinguh e Zéman, 2020, p. 163) e, com isso, aprimorar ainda mais os métodos de detecção de anomalias dentro da área financeira. Já, gestores de organizações poderão fazer escolhas de maneira mais embasada de técnicas que viabilizem uma maior interpretabilidade conforme os cenários nos quais se encontram, a fim de otimizar o processo de tomada de decisão (Sá, Pereira, e Pappa, 2018, p. 26). Isso beneficiará não somente os pesquisadores e gestores, mas como também os sistemas financeiros em geral, uma vez que uma alocação mais precisa de métodos em sistemas específicos poderá conectar as capacidades dinâmicas das técnicas de detecção com o processo de transformações digital, definindo oportunidades e transformando os recursos disponíveis (Konopik et al., 2022, p. 2), o que pode resultar em modelos de negócios que prestam serviços financeiros de modo mais eficiente (Barroso e Laborda, 2022, p. 1).

Além desta introdução, o resto do trabalho é estruturado da seguinte maneira: no Capítulo 2 são discutidos os termos e teorias relevantes para uma melhor compreensão deste trabalho. No

Capítulo 3 é feita uma detalhada apresentação do método de pesquisa utilizada, além do método de classificação e codificação utilizado nos artigos relevantes. O Capítulo 4, por sua vez, foi estruturado de modo a apresentar os resultados obtidos e discussões relevantes para a presente revisão de literatura. Por fim, as conclusões desse trabalho, assim como as orientações para trabalhos futuros, são apresentadas no Capítulo 5.

CAPÍTULO 2

REVISÃO DA LITERATURA

“Todo conhecimento humano começou com intuições, passou daí aos conceitos e terminou com ideias.”
—Immanuel Kant.

Considerando as informações apresentadas na introdução, ao se estudar anomalias dentro de um contexto financeiro, é necessário compreender melhor alguns conceitos que fazem parte desse campo de estudos, principalmente considerando as transformações pelas quais o setor financeiro vem passando. Isso porque as constantes mudanças e inovações financeiras se relacionam entre si de um modo diferente do sistema financeiro tradicional, produzindo dados e informações em processos e dinâmicas que precisam ser entendidos antes de identificar como os *outputs* advindos deles podem sinalizar anomalias e, ainda, antes de identificar quais os métodos de detecção mais pertinentes a serem utilizados em cada cenário financeiro. Desse modo, o objetivo deste capítulo é introduzir alguns assuntos que, mesmo de maneira pormenorizada, possibilitem a assimilação do conteúdo que será abordado ao longo do trabalho, como também, viabilizem que o leitor consiga fazer uma relação entre eles, a temática abordada e as informações apresentadas pelos artigos da base de revisão de literatura.

Além disso, a compreensão dos conceitos abordados a seguir revela-se importante, uma vez que os artigos selecionados no presente trabalho abordam modelos de identificação de anomalias em sistemas financeiros e tecnológicos que, muitas vezes, estão inseridos dentro dos contextos a serem compreendidos ou, então, consideram tais contextos em seus trabalhos. Com base nisso, neste capítulo será discorrido sobre os seguintes tópicos:

- A transformação digital nos sistemas financeiros e as tecnologias atuais;
- Inteligência Artificial;
- *Big Data* em finanças;
- Tecnologias financeiras;

- *Blockchain*;
- *Open Banking*;
- Pagamento instantâneo brasileiro (PIX);
- Anomalias, fraudes e *outliers* no setor financeiro;
- Métodos de detecção de anomalias;
- A maldição da dimensionalidade.

Esses temas a serem desenvolvidos no presente trabalho se considerados de forma isolada podem não parecer possuir uma relação entre si. No entanto, a sua compreensão é criticamente importante para que haja um maior entendimento acerca da totalidade do trabalho, bem como do conteúdo a ser analisado. Isso porque a tecnologia modifica e envolve diversas vertentes da sociedade, possibilitando o rápido acesso a informações por consumidores e exigindo uma maior capacidade de processamento dessas informações pelas organizações (Damaso, 2021, p. 4), de modo que os métodos de detecção de anomalias precisam seguir tais transformações, bem como se adequar às novas tecnologias.

Para Venancio (2021, p. 1), as transformações digitais, por exemplo, são capazes de impactar todo o modelo de negócios, uma vez que criam novos formatos de relacionamento com *stakeholders*, modificam as questões operacionais e reestruturam a organização de modo que ela possa estar preparada para enfrentar novos e futuros cenários. Assim, compreender as transformações digitais e as tecnologias advindas delas é crucial para um entendimento atualizado do sistema financeiro, assim como das possíveis intervenções realizadas nele, uma vez que a digitalização financeira atua no desenvolvimento sustentável e crescimento de alta qualidade da economia (X. Chen, Teng, e Chen, 2022, p. 1). Além disso, a transformação digital se mostra como um fator importante e determinante nas tomadas de decisões dentro dos modelos de negócios (Y. Chen e Xu, 2023, p. 1).

Após a crise financeira global de 2008, a profunda integração de inteligência artificial, *blockchain* e *big data* com finanças remodelou o setor financeiro, de modo que as tecnologias financeiras atuais trazem mudanças revolucionárias para a indústria financeira como um todo (X. Chen et al., 2022, p. 1). Isso é feito de maneira estratégica ao estimular os gestores a acelerarem a digitalização de seus processos por meio das novas tecnologias, para gerar novos caminhos de criação de valor para não só os stakeholders, mas como para o desenvolvimento sustentável da organização (Niu, Wen, Wang, e Li, 2023, p. 1). Com base nisso, o uso de Inteligência Artificial e a capacidade de coletar e analisar *big data* prometem tornar as organizações mais enxutas, eficientes e responsivas às necessidades dos consumidores (Newman, Mintrom, e O'Neill, 2022, p. 1). Desse modo, a transformação digital pode ser tida como um processo que visa melhorar entidades desencadeando mudanças significativas em suas propriedades por meio de combinações de tecnologias de informação, computação, comunicação e conectividade (Vial, 2019, p. 1). Ela desperta, portanto, o desenvolvimento das tecnologias que compõem os novos sistemas financeiros, que envolvem não somente inovação de serviços, mas, como também, novas formas de segurança. Logo, a compreensão da transformação digital e do impacto gerado por ela torna-se imprescindível para entender e analisar modelos de detecção de anomalias em sistemas financeiros.

A contínua transformação digital da economia introduz no cotidiano uma variedade de tecnologias capazes de modificar progressivamente o panorama financeiro empresarial (Elisa, Annalisa, Armin, Massimo, e Alexander, 2021, p. 1). Isso se intensificou ainda mais com as mudanças de hábitos relacionadas à pandemia de COVID-19 (A. R. Ferreira, 2022, p. 102). Com base nesse contexto, as tecnologias digitais demonstraram vantagens naturais quando se trata de otimizar a alocação de recursos, a produção, os serviços e as mudanças organizacional (Z. Zhang, Jin, Li, e Zhang, 2023, p. 2). Isso porque a digitalização faz com que organizações de todos os setores, formas e tamanhos se engajem na transformação de seus modelos de negócios, avançando tecnologias digitais para se manterem competitivas no mercado (Fernandez-Vidal, Perotti, Gonzalez, e Gasco, 2022, p. 29). A inovação deixa de ser delimitada por um *locus* central (Johnson, Laurell, Ots, e Sandström, 2022, p. 1) e passa a combinar avanços tecnológicos com avanços relacionados a análises de dados (Königstorfer e Thalmann, 2020, p. 1). Desse modo, a transformação digital permite que as empresas melhorem a eficiência da gestão de seus recursos, além de possibilitar a integração e transparência de informações (Niu et al., 2023, p. 2).

Essa transformação decorre da aplicação de tecnologias digitais emergentes, como *blockchains*, Inteligência Artificial, *big data*, infraestruturas digitais e plataformas digitais como um todo, influenciando tanto as estratégias das organizações, como também, as políticas administrativas (Z. Zhang et al., 2023, p. 1). Tais tecnologias viabilizam a coleta, armazenamento, transmissão e identificação de informações que podem afetar o processo de tomada de decisões em organizações (G. Tian, Li, e Cheng, 2022, p. 1), sendo importantes ferramentas para ultrapassar as barreiras de sistemas financeiros tradicionais (S. Liu, Koster, e Chen, 2022, p. 1), superar limites e atingir objetivos no contexto competitivo (Niu et al., 2023, p. 1). Isso porque a ligação entre a tecnologia digital e as indústrias tradicionais quebra as fronteiras organizacionais e muda o processo de criação de valor, gerando novos modelos de negócios e variedades de produtos e serviços (Z. Zhang et al., 2023, p. 1). Alguns benefícios da transformação digital nos serviços financeiros podem ser vistos na Figura 2.1.

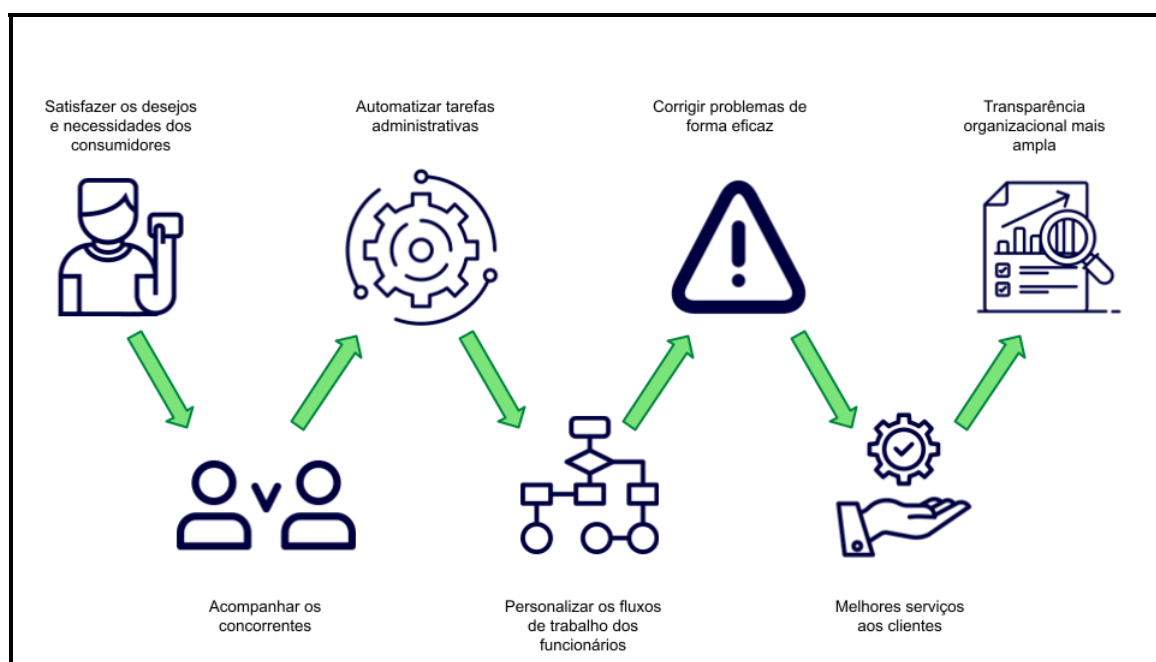


Figura 2.1: Benefícios da transformação digital em serviços financeiros.

Notas: Adaptado de Rishabh (2021, p. 1).

Com base nisso, a transformação digital e as tecnologias advindas dela são promissoras quando se trata da remodelagem da indústria financeira, uma vez que possibilitam a diminuição de custos, aprimoram a qualidade dos serviços financeiros oferecidos e diminuem a assimetria de informações, viabilizando a construção de um cenário financeiro mais diversificado e estável (Y. Wu e Huang, 2022, p. 2). Assim, as modificações não se encontram somente nas ferramentas utilizadas para aperfeiçoar os sistemas, mas como também, nos próprios sistemas, que passam a utilizar-se de novas práticas financeiras (Rodrigues, Ferreira, Teixeira, e Zopounidis, 2022, p. 2). Dessa iniciativa de transformação surgiram também as instituições pautadas, principalmente, em tecnologias financeiras, as *fintechs*¹, que englobam inovações tecnológicas no setor de finanças (Elisa et al., 2021, p. 1).

No entanto, considerando esse contexto de mudança e de tecnologia, é importante ressaltar que as novas ferramentas também são relevantes com relação ao controle de riscos em sistemas financeiros (S. Liu et al., 2022, p. 1). As aplicações de Inteligência Artificial, por exemplo, são críticas para capturar comportamentos lineares e não lineares de variáveis financeiras, de modo que elas permitem solucionar problemas que outros modelos de análise de dados não conseguem (S. Ahmed, Alshater, Ammari, e Hammami, 2022, p. 1). Em outras palavras, elas conseguem reconhecer padrões em conjuntos de dados (Goodell, Kumar, Lim, e Pattnaik, 2021, p. 3), porque são construídas com base em algoritmos que viabilizam uma simulação da inteligência humana para analisar informações (Najem, Amr, Bahnasse, e Talea, 2022, p. 633). Esse processo é o que auxilia na detecção de anomalias ou desvios nos dados financeiros, porque a cobertura das novas tecnologias é capaz de romper as limitações de tempo e espaço em dados (S. Liu et al., 2022, p. 2). A digitalização faz com que as empresas consigam processar informações massivas, em tempo real, podendo analisar esses dados em prol de melhorar a eficiência do serviço e aumentar a eficiência na tomada de decisões (G. Tian et al., 2022, p. 2). Desse modo, estudar os alcances e diversidades das tecnologias advindas da transformação digital quando se trata da detecção de anomalias em sistemas financeiros é criticamente importante para uma boa gestão de organizações.

2.2 INTELIGÊNCIA ARTIFICIAL

Em 1995, foi feita uma proposta de estudo sobre Inteligência Artificial (IA) em que foi declarado que todos os aspectos relacionados à aprendizagem ou qualquer outro recurso de inteligência podem ser tão bem descritos que uma máquina pode ser criada para fazer uma simulação deles (McCarthy, Minsky, Rochester, e Shannon, 2006, p. 2; Königstorfer e Thalmann, 2020, p. 2). Desde então, IA se tornou um ramo da ciência que atua ajudando máquinas a acharem soluções para problemas complexos de um modo mais humanizado, sendo aplicada em diversas áreas a partir de uma base de conhecimento e regras que caracterizam seus sistemas (Tiwari, Srivastava, e Gera, 2020, p. 149). Isso, porque as máquinas criadas a partir de IA têm suas performances construídas com base em funções cognitivas que normalmente são associadas à mente humana, como aprendizagem, interação e solução de problemas e, por isso, sua aplicação está se difundindo contínua e rapidamente (Johnson et al., 2022, p. 2). A diversidade de cenários em que tecnologias de IA são aplicadas pode ser vista na Tabela 2.1.

Devido a essa variedade de aplicações, espera-se que a IA revolucione e transforme os processos de produção, influenciando o comportamento dos atores econômicos (Miller, 2022, p. 1). Na esteira desse processo, a atuação das IAs ocorre por meio da construção de algoritmos que vão simular a inteligência humana de modo mais inteligente e autônomo (Najem et al., 2022, p.

¹Nota do autor: A palavra *fintech* é uma junção das palavras em inglês *financial* (financeira) e *technology* (tecnologia).

Cenários de Tecnologia	
Computação neuromórfica.	Sistemas autônomos.
Cibersegurança cognitiva.	<i>Machine Learning</i> .
Assistentes pessoais robóticos.	<i>Deep Learning</i> .
Robôs cirúrgicos autônomos.	Redes neurais.
Robótica em nuvem de última geração.	Identificação de padrões.
Jogos controlados pelo pensamento.	Processamento de linguagens.
Tradução universal em tempo real.	<i>Chatbots</i> .
Companheiros virtuais.	Análise de emoções em tempo real.

Tabela 2.1: Cenários de tecnologia relacionados à Inteligência Artificial.

Notas: Adaptado de N. A. Ahmed (2022, p. 1).

633). Os sistemas dessa tecnologia são desenvolvidos utilizando várias tecnologias de ciências de dados, além de métodos e técnicas advindos dos campos de computação, comunicação e sociologia (Miller, 2022, p. 2). Eles são capazes de ganhar habilidades adaptativas a partir do uso em conjunto de múltiplos algoritmos inteligentes que podem ser divididos entre aqueles que dependem de treinamento e podem ser atualizados para aprimorar performances ou entre aqueles que variam sua ações com base em *inputs* e *outputs*, sendo assim mais amplamente aplicáveis (J. Li, Herdem, Nathwani, e Wen, 2023, p. 3).

É importante ressaltar que os sistemas de IA já estão incorporados nos processos de tomada de decisão de vários setores (comércio eletrônico, pontuação de crédito, serviços financeiros, transporte público, saúde, educação, emprego, segurança pública, entre outros) e o seu aprimoramento para melhorar a eficiência dos sistemas tem sido cada vez mais necessário (Belli e Zingales, 2022, p. 2). Como o uso de IA está se tornando mais acessível em termos de custos tecnológicos, poder computacional, distribuição de competência e disponibilidade de dados, mais empresas estão aderindo à sua implementação em processos internos (Johnson et al., 2022, p. 2). Isso tudo ocorre, uma vez que, por causa dessa forma de tecnologia, os gestores possuem uma ferramenta para racionalizar o processo de inovação ao eliminar arbitrariedade e caos do processo de decisão, se pautando nos algoritmos que computam cálculos em grandes conjuntos de dados e automatizam decisões com base em regras e objetivos predefinidos (Pietronudo, Croidieu, e Schiavone, 2022, p. 1).

Além disso, a incorporação de IA nos processos internos é relevante uma vez que máquinas construídas com base em algoritmos de IA possuem vantagens computacionais como a capacidade de realizar operações de controle precisas, reduzir a dependência da interação humana e prevenir eficientemente erros que poderiam ser causados por fatores humanos (Sami, 2022, p. 1). Como os sistemas de IA normalmente lidam com o processamento de enormes quantidades de dados, a forma de interpretar e proteger os dados é importante para o desenvolvimento sustentável da IA, de modo a prevenir usos indevidos ou sancionar possíveis abusos, bem como identificar quaisquer anomalias no processo (Belli e Zingales, 2022, p. 2). Assim, os modelos, a partir dos algoritmos, podem aprender com os dados de modo seguro e originar informações pertinentes aos gestores, como previsões ou recomendações que podem influenciar a tomada de decisões (Miller, 2022, p. 2). Um algoritmo considerado realmente inteligente será aquele que conseguir autonomamente oferecer a melhor saída com base em variadas entradas (J. Li et al., 2023, p. 3).

Desse modo, as soluções baseadas em IA podem ser definidas como sistemas com a capacidade de agir de forma inteligente, interpretando corretamente dados externos e usando esses objetivos para executar tarefas específicas por uma configuração flexível, até o ponto de reproduzir comportamentos humanos com inteligência cognitiva, social e emocional (Johnson et al., 2022, p. 2), o que as tornam uma saída multifacetada com influências racionalizadoras

intencionais e não intencionais, em busca de problemas para resolver (Pietronudo et al., 2022, p. 2). Um resumo do funcionamento desse sistema que incorpora algoritmos que aprendem a partir de dados a originar informações relevantes para uma tomada de decisão pode ser visto na Figura 2.2.

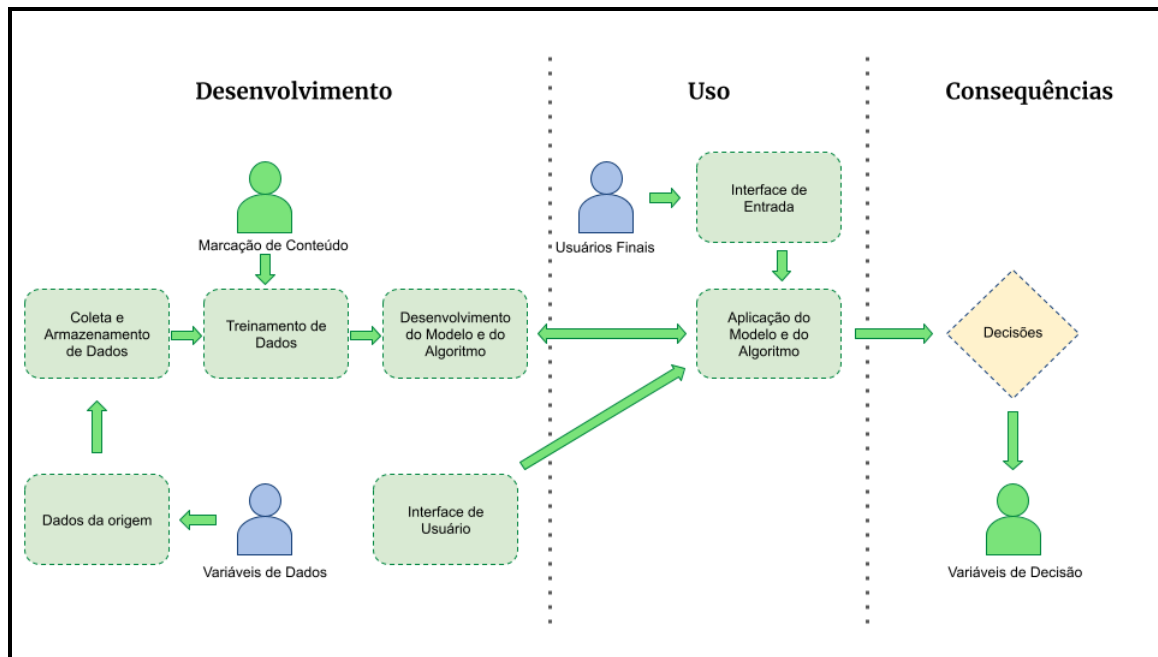


Figura 2.2: Processo de funcionamento e aplicação de Inteligência Artificial.

Notas: Adaptado de (Miller, 2022, p. 3).

Ademais, é importante destacar que o uso de IA em indústrias financeiras poderia aumentar o nível de inteligência nos processos organizacionais a serem realizados (Sami, 2022, p. 1). Para S. Ahmed et al. (2022, p. 1), o uso de aplicações de IA em tecnologias financeiras já iniciou um processo de transformação dos serviços financeiros como um todo. Os avanços na tecnologia da informação gerencial levaram ao desenvolvimento de novas tecnologias que são altamente valorizadas no setor financeiro, de modo que a IA pode tornar o trabalho dos agentes financeiros mais fácil, confiável e rápido de quando se trata da administração dessas novas tecnologias financeiras e dos dados por elas originados (Najem et al., 2022, p. 634). Ainda, a principal vantagem dessas combinações de tecnologia é sua onipresença, de modo que a IA aumenta o número de serviços disponíveis, acelera os processos para torná-los instantâneos e, finalmente, permite a personalização deles (Najem et al., 2022, p. 634). Logo, os sistemas de IA são criticamente importantes para capturar dados e comportamentos que modelos tradicionais não conseguem, o que faz com que a aplicação de modelos e algoritmos de IA sejam pertinentes para a evolução do campo de finanças (S. Ahmed et al., 2022, p. 1).

2.3 *BIG DATA* EM FINANÇAS

O fenômeno chamado *Big Data* está transformando a economia moderna (Begenau, Farboodi, e Veldkamp, 2018, p. 83), em uma era em que a transição dos fluxos de informação é mais rápida e instantânea do que nunca (Lien, 2017, p. 144). Com o *boom* da tecnologia e da ciência da computação, uma variedade de dados em formatos não estruturados ou semiestruturados surgiram e se acumularam, formando o *Big Data* e descrevendo o mundo de diferentes

perspectivas (Tang et al., 2022, p. 1). Com base nisso, *Big Data* parece ser um dos paradigmas tecnológicos disruptivos do século XXI e uma das principais alavancas da transformação digital (Dehbi, Lamrani, Belgnaoui, e Lafou, 2022, p. 440). Isso, porque estimula o desenvolvimento de inteligências de negócios que permitem que as empresas tomem decisões bem informadas que afetam positivamente a inovação nos negócios (Saeed, Riaz, e Baloch, 2022, p. 1).

Da mesma forma que *Big Data* possibilita uma pluralidade de resultados e perspectivas, também possui uma variedade de definições do seu conceito (López-Robles, Rodríguez-Salvador, Gamboa-Rosales, Ramirez-Rosales, e Cobo, 2019, p. 729). De maneira pormenorizada, o conceito de *Big Data* se refere a uma coleção de informações grandes, heterogêneas e complexas que só podem ser processadas usando várias técnicas estatísticas e analíticas avançadas (Saeed et al., 2022, pp. 2–3) e representa uma tendência relacionada à informação em sistemas financeiros que afeta a abundância de informações em grandes e pequenas empresas (Begenau et al., 2018, p. 71). Essa tendência ocorre em virtude de *Big Data* estar associada a análises que extraem valiosas informações, muitas vezes ocultas, dos dados (Saeed et al., 2022, p. 3).

A necessidade de técnicas avançadas para analisar o *Big Data* se pauta no fato de que ele tem uma abordagem multidimensional na qual se destacam cinco características principais, chamadas de os 5V's do *Big Data*: Velocidade, Volume, Veracidade, Variedade e Valor (Tang et al., 2022, p. 1; Yu, Wong, Chavez, e Jacobs, 2021, p. 2; López-Robles et al., 2019, p. 729). O volume indica a quantidade de dados, a variedade captura a diversidade de tipos de dados, que inclui dados estruturados (como planilhas) e não estruturados (como textos, imagens, áudios, etc.), a velocidade se refere à rapidez de geração e análise de dados (Saeed et al., 2022, p. 3). Já, as características de valor e veracidade de dados indicam o poder preditivo do *Big Data* de transportar informações valiosas e reais, capazes de enriquecer modelos de previsão (Tang et al., 2022, p. 2).

Nesse contexto, o *Big Data* se torna cada vez mais importante para o controle gerencial, uma vez que a análise eficiente dos dados permite uma melhor qualidade e a relevância de informações contábeis e financeiras para a gestão empresarial, viabilizando a transparência e um melhor processo de tomada de decisões pelos *stakeholders* (Dehbi et al., 2022, p. 440). Além disso, a importância do *Big Data* surge da convergência de várias tecnologias que viabilizaram uma geração de dados desenfreada, de modo que sua análise e processamento só são viáveis a partir de avanços em algoritmos de otimização de modelos estatísticos usando computação distribuída e IA (Routledge, 2018, pp. 88–89).

A partir disso, para desenvolver modelos de negócios capazes de suportar, processar, analisar e aplicar esse grande volume de dados é necessário implementar técnicas e ferramentas capazes de auxiliar nesses processos como, por exemplo, modelos de *machine learning* (Barroso e Laborda, 2022, p. 11), uma vez que a otimização dos fluxos financeiros usando os dados dessas tecnologias requer uma capacidade significativa de processamento de informações (Yu et al., 2021, p. 2) e as técnicas de *machine learning* são excelentes para extrair padrões dos dados e transformá-los em conhecimento (Baesens, Höppner, e Verdonck, 2021, p. 2). Assim, a análise baseada na computação de *Big Data* pode beneficiar as organizações bancárias e financeiras de hoje em muitos aspectos e fornecer informações muito valiosas para as organizações obterem negociações mais inteligentes, o que pode ajudá-las a obter uma grande vantagem competitiva (X. Tian, Han, Wang, Lu, e Zhan, 2015, p. 33).

Com *insights* valiosos provenientes de grandes conjuntos de dados, as empresas estarão investindo cada vez mais em sua busca para determinar como os dados podem ser utilizados para criar valor (Saeed et al., 2022, p. 1). Porém, vale ressaltar que para isso é necessário uma série de técnicas e ferramentas avançadas (López-Robles et al., 2019, p. 730), e o desenvolvimento de modelos analíticos baseados nelas ainda se mostra como um problema, devido a dificuldade de projetar um sistema de computação de alta eficiência para processar os dados históricos e

recebidos distribuídos em tempo real (X. Tian et al., 2015, pp. 34–35).

2.4 TECNOLOGIAS FINANCEIRAS

Diferentes tecnologias têm se tornado uma engrenagem de crescimento que estimula a economia global (Metawa, Dogan, e Taskin, 2022, p. 386). Nessa mesma linha, as tecnologias financeiras, comumente chamadas de *fintechs* quando relacionadas à empresas que oferecem serviços de inovação tecnológica e financeira, representam um fenômeno que promete mudar o mercado financeiro favorecendo a inclusão financeira e democratizando o acesso a serviços financeiros e desafiando as regulamentações vigentes (Sánchez, 2022, p. 1). Isso é feito por meio do redesenho de processos dos bancos, comércio, investimentos financeiros e até mesmo da remodelagem do dinheiro, a partir da aplicação de tecnologias avançadas que baseiam-se, por exemplo, em pagamento *online*, *criptomoedas* e IA (Palmié, Wincent, Parida, e Caglar, 2020, p. 3). Essas tecnologias atualizadas podem revolucionar drasticamente os serviços financeiros, tornando as transações financeiras mais seguras, convenientes e acessíveis (Cho e Chen, 2021, p. 2; Metawa et al., 2022, p. 386). Isso devido à sua lógica de aplicar soluções e serviços baseados em tecnologia da informação para aumentar a eficiência nos mercados financeiros e transações bancárias para consumidores, bancos, empresas e todos os membros do ecossistema financeiro (Palmié et al., 2020, p. 3). Desse modo o estudo dessas tecnologias e a sua aplicação são de grande relevância para o campo das finanças.

2.4.1 BLOCKCHAIN

A tecnologia de *blockchain*² é uma base de dados descentralizada em uma rede de pessoa para pessoa, em que um comprador e um vendedor interagem entre si por meio de um contrato inteligente, que representa um bloco dentro da cadeia e pode ser visualizado por outros usuários, chamados nódulos, do sistema (J. Li et al., 2023, p. 12). Ela também é uma tecnologia de contabilidade replicável e subjacente a muitas inovações incrementais e disruptivas e se baseia em blocos algorítmicos contendo informações de transações, que são conectadas e validadas em ordem cronológica e formam uma cadeia que contém registros permanentes, imutáveis, transparentes e invioláveis (Patel, Migliavacca, e Oriani, 2022, p. 1). Nesse contexto de descentralização, abertura, transparência e resistência a adulterações propiciado pela *blockchain*, os contratos inteligentes possibilitam que os participantes possam negociar e colaborar sem qualquer autoridade ou intermediário confiável de terceiros, o que expande significativamente a aplicação dessa tecnologia (Ouyang, Zhang, e Wang, 2022, p. 2). Na Figura 2.3 é possível compreender visualmente o funcionamento de uma transação na *blockchain*.

O conceito de *blockchain* foi inicialmente introduzido em conjunto com o conceito de *criptomoedas*, mas essa tecnologia de transações internas possui uma infraestrutura que vai além disso e, que pode retirar a noção de confiança de organizações centralizadas para plataformas descentralizadas que são matematicamente verificáveis e criptograficamente seguras (Prakash, Anoop, e Asharaf, 2022, p. 2). Por mais que a *blockchain* tenha ficado conhecida devido ao crescimento das *Bitcoins*³, vale ressaltar que essa tecnologia não se limita apenas à área financeira (Chang et al., 2020, p. 2). Desse modo, nos mais variados setores, as partes que participam de uma *blockchain* (empresas, instituições, indivíduos, entre outros) podem interagir entre si e criar todos os tipos de registros (informações de produtos, certificados, dados de localização,

²Do inglês: Cadeia de blocos.

³Nota do autor: Um tipo de *criptomoeda* muito reconhecido. Além dessa, outras *criptomoedas* conhecidas são *Ether* e *Ripple*.

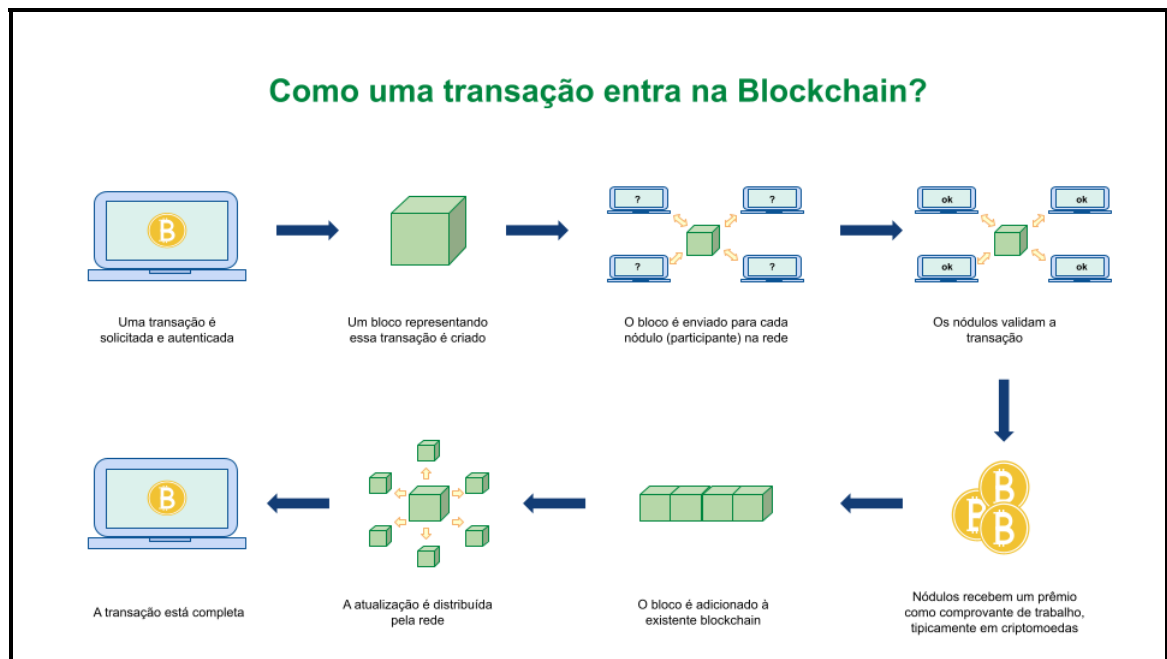


Figura 2.3: Como uma transação entra na *Blockchain*.

Notas: Adaptado de Learning (2020, p. 1).

registros de transações, dados adquiridos de sensores) (Agi e Jha, 2022, p. 2).

Nesse contexto, a tecnologia de *blockchain* possui um grande potencial para auxiliar modelos de negócios e indivíduos a concordarem, em escala global, sobre o verdadeiro estado de ativos dentro de um mercado sem depender de intermediários dispendiosos, por meio de uma combinação inteligente entre incentivos econômicos e criptografia (Marthews e Tucker, 2022, p. 1). Isso, porque antes de serem armazenados na *blockchain*, os registros são verificados e validados utilizando um mecanismo de consenso entre os participantes para, então, serem combinados na formação de um bloco de dados que será conectado ao restante da cadeia de blocos (Agi e Jha, 2022, p. 2). No entanto, é importante destacar que o desenvolvimento dessa tecnologia não está maduro ainda, uma vez que com ela vieram diversos desafios como escalabilidade, segurança, privacidade, latência, entre outros, de modo que os mercados, inclusive o financeiro considerado neste trabalho, busquem uma maior compreensão da indústria de *blockchain* para identificar soluções robustas para cada desafio (Chang et al., 2020, p. 2).

Na esteira desse processo, é preciso conhecer bem os elementos que caracterizam a *blockchain*: (1) lógica computacional; (2) transmissão ponto a ponto; (3) irreversibilidade de registros; (4) banco de dados distribuído e (5) transparência com pseudônimo (Chang et al., 2020, p. 2). A lógica computacional relaciona-se ao fato de que as transações podem ser automaticamente ativadas por algoritmos; a transmissão ponto a ponto permite que haja uma interação descentralizada de participante para participante sem que haja a necessidade de um nó central; a irreversibilidade de registros implica na imutabilidade das transações que são organizadas de modo cronológico; o banco de dados distribuído é aquele que não pode ser controlado por um único indivíduo, mas pode ser modificado ao adicionar novas transações e validado por todos os participantes; por fim, a transparência com pseudônimo permite que as transações sejam transparentes e visíveis em toda a cadeia por qualquer participante, mas ao mesmo tempo, cada nó pode decidir se vai compartilhar detalhes sobre si ou usar um pseudônimo (Patel et al., 2022, p.1).

A *blockchain* passou por três tecnologias de desenvolvimento que podem ser vistas na Figura 2.4. No entanto, é preciso atentar-se que, por mais que essa tecnologia tenha evoluído, ainda há vários ataques cibernéticos possíveis que podem ocorrer nessas cadeias, de modo que

é importante criar uma assessoria de riscos ou um modelo que possa prever e lidar com esses ataques na *blockchain* (Prakash et al., 2022, p. 12). Considerando isso, a tecnologia oferece algumas formas de segurança, como o fato de que se um nó é hackeado ou atacado, o agente malicioso só consegue mudar ou afetar aquele nó, porque cada bloco é isolado um do outro, mesmo em cadeia e, além disso, toda vez que há uma nova transação, todos os participantes da cadeia precisam validar a integridade do bloco, de modo que o controle é maior (J. Li et al., 2023, p. 12). Para aumentar ainda mais a segurança surgiu a inteligência *blockchain* que é a convergência entre *blockchain* e IA, de modo que as duas tecnologias consegue se complementar dentro do contexto de transformação digital (Ouyang et al., 2022, p. 1).

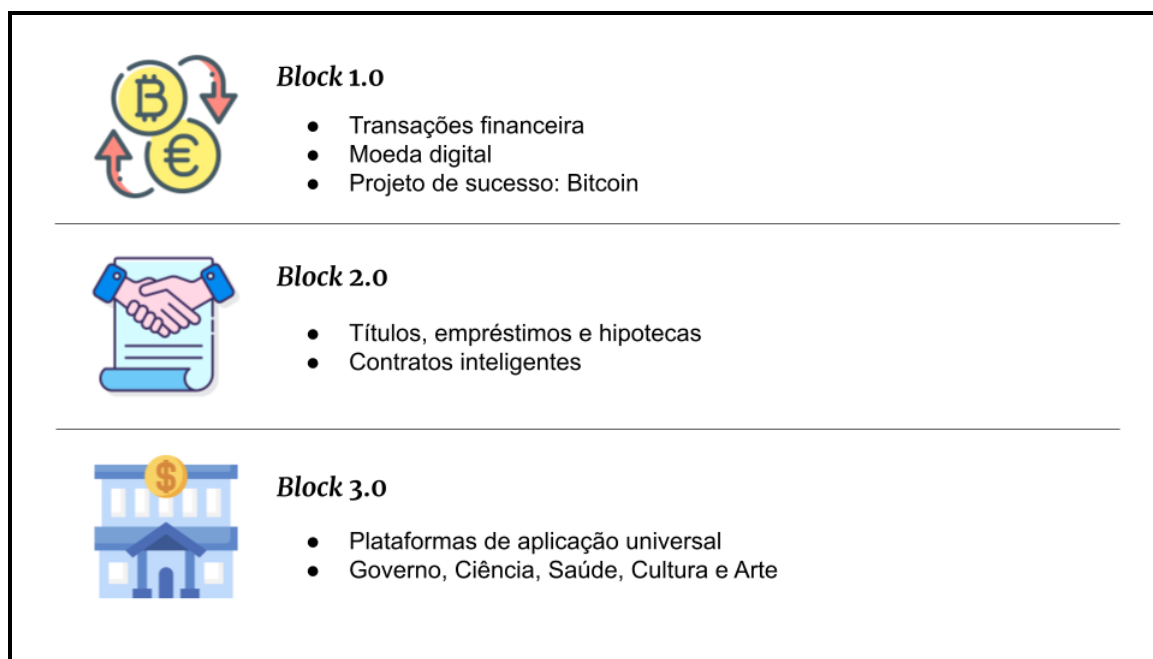


Figura 2.4: Três gerações de *Blockchain*.

Notas: Adaptado de (Chang et al., 2020, p. 3).

Outra questão importante é o fato de que, como a *blockchain* é descentralizada, e não há uma autoridade central para lidar com as disputas e transações, quaisquer transações consideradas fraudulentas, ou qualquer sinal de anomalia deve ser discutido diretamente com o outro usuário em negociação e isso pode configurar um problema caso o outro usuário negue a solicitação de revisão dos dados (J. Li et al., 2023, p. 1). Essa falta de coordenação central pode limitar o uso da *blockchain* (Chang et al., 2020, p. 3) e indica a necessidade do aumento de estudos acerca de como identificar, prever e lidar com anomalias e fraudes que possam surgir dentro dessa tecnologia.

O *Open Banking* é uma tendência de inovação e integração de serviços (Liao, Guan, Cheng, e Yuan, 2022, p. 450) em que os clientes do sistema bancário, sejam eles bancos, *fintechs* ou quaisquer instituições de pagamentos, podem viabilizar o compartilhamento e disseminação de suas informações com outras instituições que, no contexto brasileiro, são verificadas e autorizadas pelo Banco Central (Coimbra, Ikeno, Ikeda, e Rodrigues, 2022, p. 168). Nesse processo, as instituições precisam fornecer uma interface de programação de aplicativos na qual os usuários dentro de um cenário de interoperabilidade podem realizar operações bancárias por meio dessa interface (Mansfield-Devine, 2016, p. 8). Além disso, as contas bancárias desses usuários podem ser utilizadas e movimentadas por meio de outras plataformas que não somente o aplicativo ou site da instituição bancária (Coimbra et al., 2022, p. 168). Nos sistemas bancários tradicionais, os dados de transações da conta do usuário geralmente são considerados um ativo do banco, mas o *open banking* devolve o controle aos clientes para compartilhar informações da conta com provedores de serviço terceiros por meio das interfaces de usuários (Liao et al., 2022, p. 451). Uma comparação das relações com o consumidor nos bancos tradicionais e no *open banking* é apresentada na Figura 2.5.

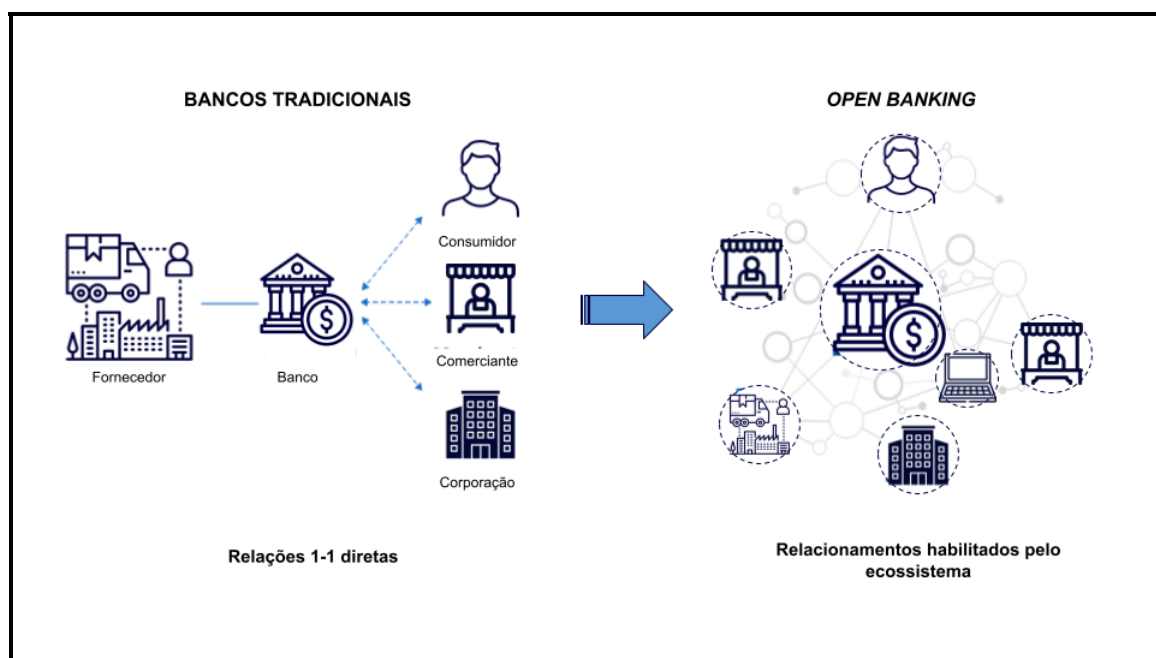


Figura 2.5: Diferença entre *open banking* e bancos tradicionais.

Nota: Adaptado de Raisin (2022, p. 1).

É importante destacar que esse sistema de *open banking* sinaliza tanto oportunidades como desafios para bancos ao redor do mundo, principalmente quando se considera gerenciamento de dados (H. Wang, Ma, Dai, Imran, e Wang, 2020, p. 812). Isso porque ao divulgar as interfaces para os provedores de serviços terceiros, as instituições financeiras também precisam atentar-se a possíveis ataques maliciosos e adulteração de contas por *hackers* (Liao et al., 2022, p. 450). Para especialistas em segurança, qualquer serviço acessível pela *internet* será atacado por pessoas que desejam explorá-lo para fins nefastos (Mansfield-Devine, 2016, p. 10). Além disso, outro desafio encontrado no *open banking* está relacionado à privacidade dos usuários, de modo que as instituições bancárias possuem mecanismos para impedir que os proprietários de dados compartilhem e troquem dados fora de suas instituições, evitando assim situações de fraude e abuso (H. Wang et al., 2020, p. 812).

Características PIX	Definição
Rápido.	Transações concluídas em poucos segundos.
Disponível.	Vinte e quatro horas por dia, todos os dias.
Fácil.	Experiência facilitada ao usuário.
Barato.	Gratuito para pessoa física, custos baixos para demais casos.
Seguro.	Robustez de mecanismos de segurança.
Aberto.	Estrutura de ampla participação de instituições.
Versátil.	Instrumento multiproposta para vários pagamentos.
Integrado.	Informações para conciliação e automação.

Tabela 2.2: Características do sistema de pagamentos PIX.

Nota: Adaptado de Banco Central (2020b, p. 1).

Uma saída encontrada pelos bancos para aumentar a segurança é a adoção da autenticação de dois fatores em que um código é enviado ao celular dos usuários de modo a se certificar de que é ele mesmo tentando acessar, mas não acabou com todos os problemas relacionados à segurança (Mansfield-Devine, 2016, p. 9). De todo modo, o usuário deve estar no centro da transformação digital, e as instituições devem levar em consideração seu perfil, necessidades, objetivos e interesses (Damaso, 2021, p. 2), inclusive as questões de proteção e privacidade de seus dados, sendo esta uma das mais importantes questões gerenciais quando se trata de *open banking* (H. Wang et al., 2020, p. 821). Logo, os estudos acerca de modelos de análise, compreensão e proteção dos dados devem ser explorados com relação a essa nova tecnologia, de modo a prevenir fraudes ou perdas financeiras para qualquer um dos participantes do sistema.

2.4.3 PAGAMENTO INSTANTÂNEO BRASILEIRO (PIX)

O pagamento instantâneo brasileiro (PIX) é um sistema de pagamentos econômicos instantâneos desenvolvido no ano de 2020 e controlado pelo Banco Central brasileiro (A. R. Ferreira, 2022, p. 101). A instantaneidade de pagamento está relacionada ao fato de que a transferência de recursos, considerando a transmissão do pagamento e a disponibilidade de recursos, é feita em tempo real, em um serviço contínuo e considerado ininterrupto (Coimbra et al., 2022, p. 162). A transferência dos recursos é feita em qualquer período do dia rapidamente, podendo ser feita a partir de conta corrente, poupança ou até mesmo por meio de uma conta pré-paga de pagamento e o Banco Central garante que é um processo seguro (Banco Central, 2020b, p. 1). Para realizar o PIX, é necessário que os usuários cadastrem chaves de transação que podem ser CPF, CNPJ, *e-mail*, número de celular ou, então, uma chave sem estar relacionada com os dados pessoais que utiliza de números e letras aleatórios (Coimbra et al., 2022, p. 162). As chaves são necessárias para que o sistema de pagamentos instantâneos consiga reconhecer os dados das contas (Coimbra et al., 2022, p. 162).

Esse sistema PIX foi criado como parte das reformas de regulação do sistema financeiro que são apresentadas na Agenda BC#, do Banco Central, e pode ser visto como uma iniciativa para esse fim (A. R. Ferreira, 2022, p. 2). Algumas características dele é que não possui limite de horário para que as transações e transferências sejam realizadas entre quaisquer bancos e *fintechs*⁴ e que não há limite mínimo de valores para realizar as transferências ou pagamentos (Banco Central, 2020b, p. 1). No entanto, as instituições financeiras são livres para estabelecerem seus próprios limites aos seus usuários com relação ao PIX, principalmente visando diminuir o risco de fraudes, lavagem de dinheiro ou financiamento de terrorismo (Banco Central, 2020b, p. 1). Outras características do sistema PIX podem ser vistas na Tabela 2.2.

⁴Nota do autor: Empresas que oferecem serviços financeiros com base em inovações tecnológicas.

Devido a praticidade, rapidez, versatilidade e facilidade de ter a tecnologia de pagamentos instantâneos em qualquer lugar, a utilização do PIX está cada vez maior, de modo que a cada dia supera o número de recorde transacional, sendo um dos métodos de pagamento mais utilizados no Brasil e amplamente disponível nas instituições financeiras (Ávila, Silva, e Bender, 2022, p. 12013). Além disso, por causa da pandemia do COVID-19, houve um aumento da bancarização digital, de modo que os clientes das instituições financeiras que não faziam transferências TED ou DOC de modo virtual, passaram a realizar o PIX (Febraban, 2021, p. 55). No entanto, em conjunto com a digitalização é importante atentar-se aos mecanismos de segurança do sistema de pagamento, que deve ser abrangente e viabilizar uma estrutura em que o usuário se sinta protegido e seguro e não apenas fornecer alertas de ações criminosas (Lima, 2020, p. 1). Na Figura 2.6 é possível visualizar o processo de segurança do Banco Central para o PIX.

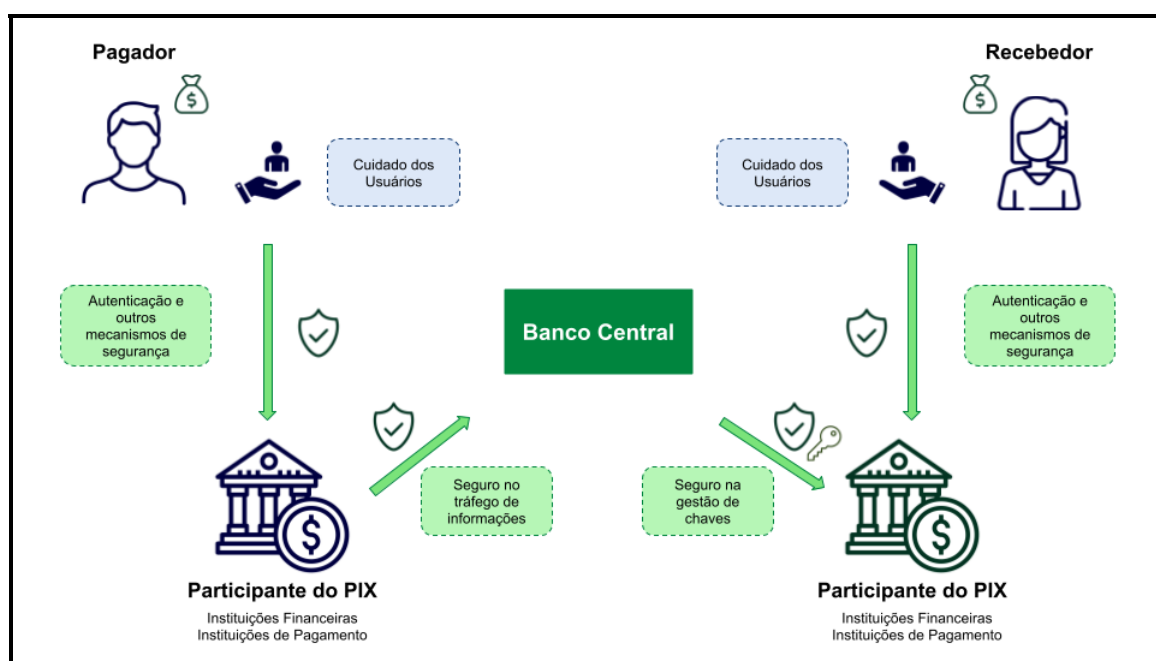


Figura 2.6: Processo de segurança do sistema de pagamentos PIX.

Fonte: Banco Central (2020a, p. 1).

No fim de 2021, o Banco Central alterou a regulamentação do PIX visando aumentar os mecanismos de segurança do sistema, incluindo um bloqueio cautelar a partir do provedor de serviços de pagamentos caso haja alguma anomalia que sinalize indícios de fraude, uma obrigatoriedade de notificação de infração na suspeita de fraude, aumento da responsabilidade das instituições financeiras, criação de mecanismos de proteção de dados e uso de dados para prevenir e identificar ações criminosas (Coimbra et al., 2022, p. 163). Segundo o relatório da Federação Brasileira de Bancos de 2021, as pessoas físicas realizam as transações PIX mais do que as pessoas jurídicas e, isso pode ser relativo às questões de segurança relacionadas à ferramenta (Febraban, 2021, p. 54). Desse modo, essa tecnologia financeira, apesar de muito recente, já é foco de fraudes e merece atenção quando se trata de analisar dados para detectar anomalias.

Muitos autores já propuseram diversas definições para anomalia, não havendo uma definição universal adotada, uma vez que o conceito por trás dessa palavra depende da estrutura de dados analisada e das aplicações consideradas (Hilal, Gadsden, e Yawney, 2022, p. 3). Em geral, anomalias podem ser observações, padrões, casos ou pontos inconsistentes em dados com distribuições pré estabelecidas (Xia et al., 2022, p. 498). Em outras palavras, dados anômalos são aqueles que não estão associados a comportamentos esperados em um sistema (Chatterjee e Ahmed, 2022, p. 2), o que pode indicar diversos problemas, principalmente considerando o setor financeiro, uma vez que tais dados podem ser resultado de falhas no sistema, ataques externos (Chatterjee e Ahmed, 2022, p. 2) e/ou fraudes financeiras (Audibert, Michiardi, Guyard, Marti, e Zuluaga, 2022, p. 1). De todo modo, anomalias podem sinalizar inconveniências nos sistemas financeiros, e o monitoramento das variáveis que se relacionam a elas é imprescindível para o reconhecimento de riscos e para o controle e gerenciamento dos dados de modo a prevenir dificuldades financeiras (Climent et al., 2019, pp. 885–886).

O termo anomalia também pode ser encontrado na literatura com outros nomes, como *outlier* ou fraude, sendo eles usados alternadamente, principalmente se tratando de métodos de detecção desses dados que saem do padrão (Hilal et al., 2022, p. 1). De modo similar à anomalia, *outliers* se referem a desvios de informações com base em uma região de distribuição de dados (Tra, Amayri, e Bouguila, 2022, p. 2), em que seus valores se diferenciam significativamente da maioria dos dados coletados, levando a parâmetros de estimação enviesados e resultados incorretos (Koren, Koren, e Peretz, 2023, p. 1). Esses desvios podem levar a especificações incorretas de um sistema, previsões ruins e inferências inválidas, de modo que a identificação e correção dos *outliers* é um objetivo importante da modelagem financeira (Ané, Ureche-Rangau, Gambet, e Bouverot, 2008, p. 326). Ao mesmo tempo, o estudo dos *outliers* também é importante para entender os mecanismos que os originam e a influência que eles causam nas análises estatísticas de dados (K. Li et al., 2022, p. 1).

Na mesma linha, fraude também se refere a diferenças de comportamento em dados, mas esse termo é mais comumente utilizado para se falar de anomalias e *outliers* quando estão relacionadas a crimes bem pensados e organizados, inseridos nos dados de modo oculto e que podem se manifestar de diversas formas (Baensens et al., 2021, p. 1). Fraudes se relacionam com apropriação ilegal de propriedade, incluindo dinheiro e, quando são feitas de modo *online* envolvem a reconstrução de transações financeiras e de fluxos de capital entre múltiplas partes (Nikkel, 2020, p. 2). Com base nisso, as fraudes são manifestadas como sendo um comportamento coletivo dentro das redes de dados, uma vez que os defraudadores tentam coordenar o comportamento delas com o comportamento do grupo de dados (Pourhabibi et al., 2020, p. 7), tornando assim a detecção desses crimes complexa, uma vez que as alterações e desvios nos dados são feitas de modo que aparentemente parecem legítimas (Carneiro et al., 2017, p. 91).

Assim, as anomalias se manifestam de diversas formas não só na literatura, mas como também nos sistemas financeiros, podendo sinalizar crimes ou *insights* importantes quando se trata da análise de dados. Os crimes advindos da interferência em transações e dados financeiros podem variar em natureza e ter o efeito de possivelmente desestabilizar economias, aumentar o custo de vida e impactar a sensação de segurança do consumidor (Hilal et al., 2022, p. 6). Logo, o estudo aprofundado em abordagens de detecção de anomalias em sistemas financeiros podem melhorar a precisão, confiabilidade e velocidade dos processos relacionados a essa temática, levando a alertas financeiros mais pontuais, consistentes e que podem ser executados em tempo real (Rubio et al., 2020, p. 16).

O estudo de técnicas de detecção de anomalias aborda o problema de identificar instâncias de dados ou eventos que não estão de acordo com o comportamento esperado (Chandola, Banerjee, e Kumar, 2009, p. 1; Hilal et al., 2022, p. 1) a partir da análise de desvios em distribuições normais de dados (Audibert et al., 2022, p. 1). As diversas abordagens estabelecem se os dados coletados ou produzidos pelos sistemas refletem o comportamento esperado, sinalizando a ocorrência de eventos que merecem atenção (Sgueglia, Sorbo, Visaggio, e Canfora, 2022, p. 172).

É importante destacar que cada método de detecção de anomalias é construído a partir de diferentes perspectivas, de acordo com o cenário financeiro em que os dados se encontram. Porém, em geral, eles seguem a mesma ideia de aprender modelos de dados de instâncias normais e genuínas para então identificar as transações e dados que não estão em conformidade (Jurgovsky et al., 2018, p. 235). De todo modo, a detecção de anomalias é uma importante atividade de mineração de dados, usada para identificar padrões, tendências e novidades interessantes e emergentes em diferentes conjuntos de dados financeiros (Parnes, 2020, p. 308).

2.6.1 APLICAÇÃO DE *MACHINE LEARNING* EM DADOS FINANCEIROS

Machine learning é uma forma de tecnologia que se manifesta como uma IA que possibilita que máquinas aprendam a reconhecer padrões em conjuntos de dados (Goodell et al., 2021, p. 3). Essa habilidade viabiliza a criação de modelos e ferramentas para solucionar desafios em diversas áreas (Gan, Wang, e Yang, 2020, p. 1), principalmente, com os avanços tecnológicos, que permitiram o desenvolvimento de diversas dessas aplicações em setores financeiros (Sarlin e Björk, 2017, p. 1). Isso fez com que toda a indústria financeira fosse transformada, tornando esse método de IA dominante a partir dos anos 2020 (S. Ahmed et al., 2022, p. 1). De maneira pormenorizada, *machine learning* se refere a uma classe de modelos de ciências de dados que conseguem aprender a partir de dados pré estabelecidos e, a partir disso, aprimorar a sua performance conforme seu próprio aprendizado, replicando o aprendizado humano por meio de programas computacionais (Ghoddusi, Creamer, e Rafizadeh, 2019, p. 710).

As técnicas de mineração de dados, como as de *machine learning* são divididas em duas abordagens principais: supervisionadas e não supervisionadas, com base em informações registradas no passado (Carneiro et al., 2017, p. 93). Os métodos supervisionados utilizam bases de dados conhecidas para construir um modelo de análise, enquanto os métodos não supervisionados não possuem dados preliminares para comparação (K. Singh e Best, 2019, p. 4). Normalmente, os métodos não supervisionados são aplicados em dados não rotulados como normais ou anormais, de modo que o objetivo do algoritmo de aprendizado é identificar padrões escondidos no conjunto de dados, para poder fazer previsões ou visualizar desvios (Rubio et al., 2020, p. 2). Como dados já rotulados estão frequentemente em falta ou inexistem nos problemas do mundo real, as técnicas de aprendizado não supervisionado costumam ser o foco da maioria das pesquisas e modelos de análise (Pourhabibi et al., 2020, p. 5).

As vantagens das técnicas de *machine learning* sobre outros métodos estatísticos ou econômicos se pauta no fato de que os algoritmos de *machine learning* conseguem gerenciar uma grande quantidade de dados estruturados e não estruturados e tomar decisões ou fazer previsões de forma rápida e precisa (Ghoddusi et al., 2019, p. 710). O processo de aplicação dos algoritmos desde os dados não tratados até os dados estruturados podem ser vistos na Figura 2.7.

Além disso, os modelos de *machine learning* oferecem uma maior potência computacional e flexibilidade funcional para decifrar padrões complexos em dados de alta dimensionalidade (Goodell et al., 2021, p. 2), oferecendo novas oportunidades em termos de inovação tanto na área econômica quanto na área financeira (Ghoddusi et al., 2019, p. 709). Consequentemente, a

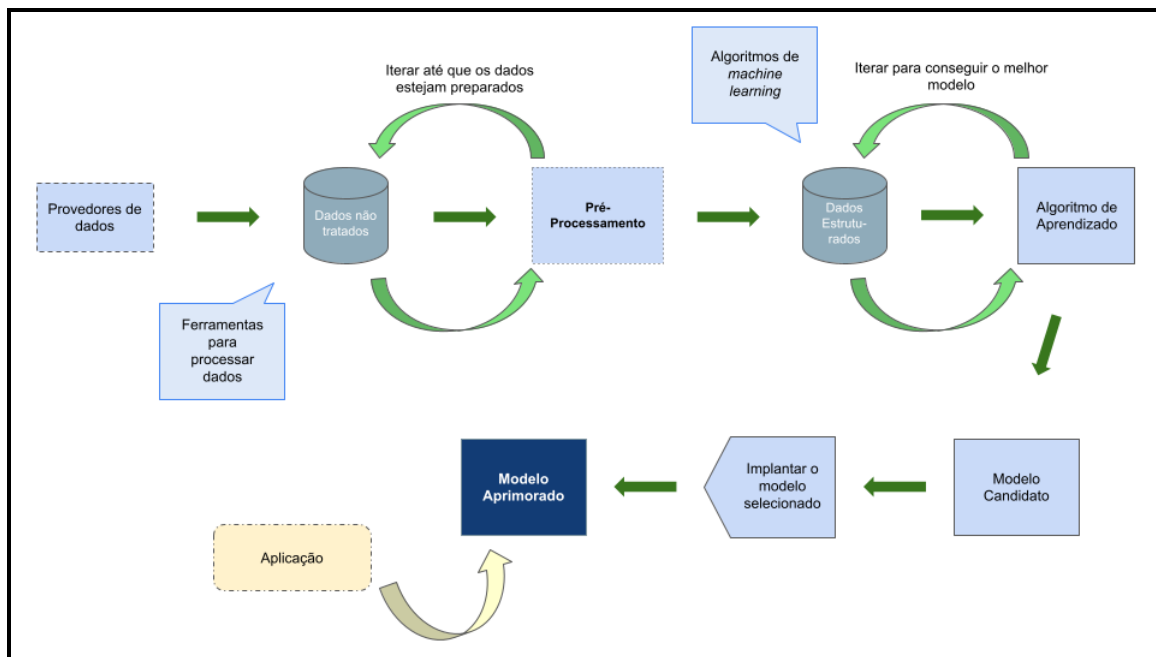


Figura 2.7: Processos de *Machine Learning*.

Notas: Adaptado de Mittal (2017, p. 1).

aplicação das técnicas de *machine learning* é conveniente e relevante para prever padrões intermitentes em dados, assim como para detectar anomalias dentro dos conjuntos de dados (Amini, Elmore, Özde Öztekin, e Strauss, 2021, p. 1). Exemplos de técnicas de *machine learning* que usam as abordagens supervisionadas e não supervisionadas podem ser vistos na Tabela 2.3.

Os algoritmos de *machine learning*, desse modo, são capazes de detectar vulnerabilidades no sistema e fornecer soluções de segurança adequadas para manter a plataforma segura e, por isso, são adequados na detecção de prevenção de fraudes nas empresas financeiras (Palathadka et al., 2021, p. 3). Segundo León, Barucca, Acero, Gage, e Ortega (2020, p. 1), as técnicas de *machine learning* são superiores quando se trata do desenvolvimento de ferramentas e modelos automatizados para controlar e supervisionar sistemas financeiros. Desse modo, essa técnica mostra-se muito relevante para o presente trabalho.

2.6.2 LEI DE BENFORD E A ANÁLISE DE ANOMALIAS FINANCEIRAS

A Lei⁵ de Benford, também chamada de Lei de Newcomb-Benford, em homenagem ao astrônomo e matemático Simon Newcomb e ao engenheiro e físico Frank Benford, é mais uma simples observação da realidade do que um resultado matemático demonstrável (Azevedo, Gonçalves, Gava, e Spinola, 2021, p. 1). De maneira resumida, ela surgiu da observação de manchas por excesso de toques em páginas de um livro de tabelas de logaritmos, de modo que as páginas mais desgastadas e, logo, mais lidas, apresentavam os logaritmos de números com primeiros dígitos mais baixos, como 1 e 2, enquanto as páginas menos desgastadas indicavam os logaritmos de números com primeiros dígitos mais altos, como 8 e 9 (Nigrini, 2012, p. 2). Porém, esse fenômeno não se restringe a somente as páginas de um livro. Quando se compila, por exemplo, 20

⁵Nota do autor: O entendimento do vernáculo “Lei” foi utilizado com base nos estudos e percepção dos autores Kedar e Hon (2017, p. 21), e representa uma conexão entre acidentes, ligada por uma fórmula matemática dentro de uma dicotomia de leis naturais imanentes ao mundo e de leis impostas por um Deus onipotente. Logo, tem-se, neste trabalho, a concepção de “Lei” como algo regido pela natureza, de modo que é possível saber que aconteceu ou que irá acontecer por meio de cálculos matemáticos, mas não sendo possível prever quando, uma vez que tal fato depende da natureza.

Supervisionado	Não Supervisionado
<ul style="list-style-type: none"> • Regressão: <ul style="list-style-type: none"> - Linear; - Polinomial. • Árvores de Decisão e Florestas Aleatórias. • Classificação: <ul style="list-style-type: none"> - KNN (Algoritmo K vizinhos próximos); - SVM (Máquina de vetores suporte); - Naive-Bayes; - Regressão logística. 	<ul style="list-style-type: none"> • Agrupamento e Redução de Dimensionalidade: <ul style="list-style-type: none"> - SDV (decomposição em valores singulares); - PCA (análise do componente principal); - Agrupamento <i>K-means</i>. • Análise de Associação: <ul style="list-style-type: none"> - <i>Apriori</i>; - <i>FP-Growth</i> (Padrões frequentes). • Modelo Oculto de Markov.

Tabela 2.3: Abordagens de *Machine Learning*.

Notas: Adaptado de Skiles (2022, p. 1).

mil primeiros dígitos retirados de diferentes fontes, é observada uma mesma mesma frequência desses dígitos, que segue uma distribuição logarítmica (Benford, 1938, p. 551).

Nesse contexto, a Lei de Benford é uma lei fenomenológica acerca da distribuição probabilística dos primeiros dígitos significantes em um conjunto de dados (Shi, Ausloos, e Zhu, 2018, p. 878). De maneira pormenorizada, a ocorrência de dígitos em dados gerados naturalmente não é uniforme (Badal-Valero, Alvarez-Jareño, e Pavía, 2018, p. 24) e segue uma distribuição específica (Druică, Oancea, e Vâlsan, 2018, p. 76): em um conjunto de dados numéricos o número 1 aparece como o dígito mais significativo cerca de 30% das vezes, enquanto o dígito 9 aparece menos de 5% das vezes (Azevedo et al., 2021, p. 1). A título de exemplo, essa distribuição pode ser vista de modo mais detalhado na Tabela 2.4, em que foram utilizados os primeiros 4000 números da sequência de Fibonacci, um conjunto de dados naturais, para demonstrar a frequência de distribuição dos dígitos significativo.

Dígito	Frequência
1	30,099%
2	17,625%
3	12,525%
4	9,650%
5	7,950%
6	6,650%
7	5,775%
8	5,149%
9	4,575%

Tabela 2.4: Frequência de distribuição do primeiro dígito significativo dos primeiros 4000 números da sequência de Fibonacci.

Essa quantidade de números foi escolhida, uma vez que para que haja uma boa conformidade com a Lei de Benford, o arranjo numérico deve possuir quatro ou mais dígitos e é necessário utilizar pelo menos 1000 números (Arboleda, Guzman-Luna, e Torres, 2018, p. 13). No entanto, é de suma importância destacar que a Lei de Benford é uma regra que descreve a distribuição esperada de números naturais aleatórios e que o número zero não pode ser considerado um primeiro dígito (Lin, Lin, Yeh, e Wang, 2018, p. 159). Esses dados também podem ser analisados de forma visual, devido à sua distribuição logarítmica (Benford, 1938, p. 551), como

na Figura 2.8.

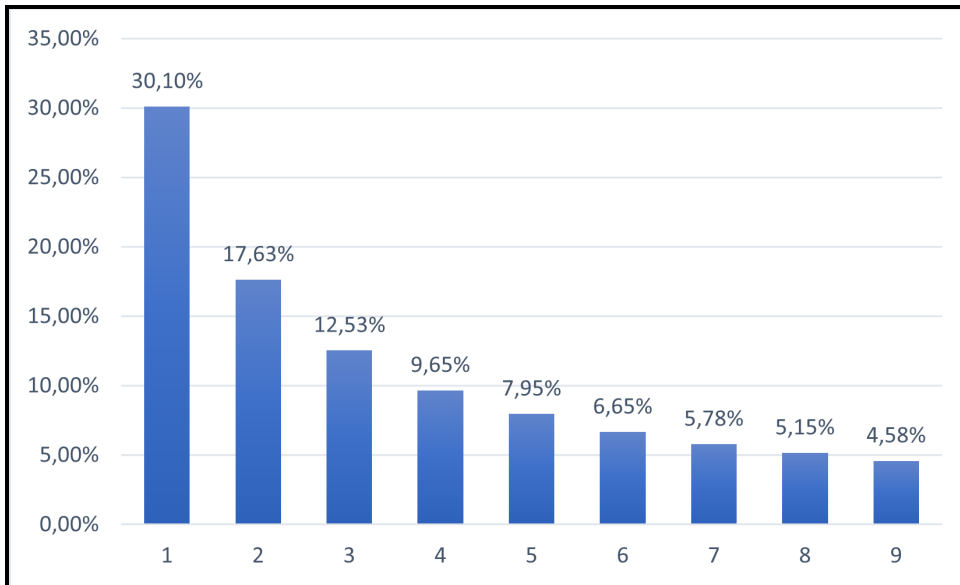


Figura 2.8: Distribuição logarítmica da frequência do primeiro dígito dos 4000 primeiros números da sequência de Fibonacci.

A partir disso, a distribuição de Benford pode ser empregada para detectar uma ampla variedade de conjuntos de números e, desde que o primeiro dígito dos números seja gerado estocasticamente, esses conjuntos de números provavelmente satisfarão a condição de aplicação Lei de Benford (Lin et al., 2018, p. 162). Por causa disso, ela pode ser utilizada como uma técnica para descobrir possíveis erros ou dados atípicos em conjuntos de dados (Arboleda et al., 2018, p. 13), como uma ajuda nos procedimentos analíticos de teste da exatidão de relatórios financeiros (Shi et al., 2018, p. 887). Assim, a Lei de Benford é utilizada como uma ferramenta de contabilidade e auditoria forense para dados financeiros (Azevedo et al., 2021, p. 1), de modo que é amplamente aplicada na investigação de manipulação de dados por pesquisadores em finanças e economia, sendo utilizada não somente para identificar dados falsos criados, mas, como também, para verificar a não confiabilidade de dados macroeconômicos (Shi et al., 2018, p. 878) e detectar possíveis anomalias (Arboleda et al., 2018, p. 13). Para isso, conforme ensina Nigrini (2012, p. 5), pode-se identificar as frequências dos dígitos com base nas equações 2.1, 2.2 e 2.3, com D_1 representando o primeiro dígito, D_2 o segundo dígito e D_1D_2 os dois primeiros dígitos de um número (Nigrini, 2012, p. 5)⁶:

$$\text{Prob}(D_1 = d_1) = \log \left(1 + \frac{1}{d_1} \right); \quad d_1 \in (1, 2, \dots, 9) \quad (2.1)$$

$$\text{Prob}(D_2 = d_2) = \sum_{d_1=1}^9 \log \left(1 + \frac{1}{d_1 d_2} \right); \quad d_2 \in (1, 2, \dots, 9) \quad (2.2)$$

$$\text{Prob}(D_1D_2 = d_1 d_2) = \log \left(1 + \frac{1}{d_1 d_2} \right); \quad d_1 d_2 \in (10, 11, \dots, 99) \quad (2.3)$$

Quando um conjunto de dados gerados naturalmente falha no teste de conformidade com a Lei de Benford, há uma evidência de que os números ali presentes podem ter sido manipula-

⁶Nota do autor: Um exemplo numérico desse cálculo é apresentado no Apêndice A.

dos (Badal-Valero et al., 2018, p. 25). Porém, é necessário ter em mente que isso por si só não pode ser considerado uma infração, uma vez que a conformidade com a Lei de Benford não é um processo dicotômico no qual os dados se conformam ou não, de modo que ela representa um processo de graus contínuos de conformidade, desde a conformidade muito forte até a conformidade marginal, até o afastamento muito forte da conformidade (Druică et al., 2018, p. 79). Isso vai também de encontro com uma certa resistência do uso dessa técnica para a detecção de anomalias em sistemas financeiros. Não só não possível ter certeza acerca da manipulação dos dados, mas como também potenciais vigaristas podem estar cientes da necessidade de os dados estarem conforme a Lei de Benford, diminuindo a confiabilidade dessa técnica (Shi et al., 2018, p. 879).

Porém, segundo os estudos de Azevedo et al. (2021, p. 11), quando usada corretamente, a Lei de Benford é uma ferramenta poderosa para identificar contas suspeitas de fraude para posterior análise e investigação. Para isso, é necessário que sejam considerados métodos alternativos que complementem a análise clássica da Lei de Benford (Shi et al., 2018, p. 879). Além disso, cada série de dados naturais tem seu próprio nível usual de conformidade com Benford, de modo que qualquer análise deve levar em conta essa realidade (Druică et al., 2018, p. 77).

Com base nisso, o uso de técnicas estatísticas, por exemplo, pode auxiliar na análise de distribuições individuais em comparação com uma distribuição esperada, como a da Lei de Benford (Azevedo et al., 2021, p. 4). De acordo com Nigrini (2012, pp. 149–150), as amostras devem ser submetidas a uma série de testes estatísticos, como qui-quadrado, teste Z, desvio médio absoluto, teste Kolmogorov-Smirnoff, teste global e Mantissa arc teste para, assim, ser possível validar com mais segurança a distribuição do conjunto de dados de acordo com a Lei de Benford. Isso, porque outros procedimentos analíticos podem equilibrar a análise a ser feita a partir da Lei de Benford (Druică et al., 2018, p. 77). De todo modo, a Lei de Benford é usada regularmente por auditores internos de grandes empresas públicas respeitáveis, por grandes e pequenas empresas de auditoria externa, por contadores forenses e examinadores de fraude (Nigrini, 2012, p. xvii).

2.7 A MALDIÇÃO DA DIMENSIONALIDADE

A compreensão das características de dados provenientes do setor financeiro é criticamente importante para que seja feita uma análise efetiva acerca das informações (Huang e Kou, 2014, p. 31). No entanto, as bases de dados contemporâneas possuem milhares de aplicações e dados com características diferentes (Setzer e Bichler, 2013, p. 62), que são continuamente gerados e processados em tempo real, resultando em um grande volume de informações sem natureza ou tamanho estruturados como, por exemplo, o *Big Data*, o que inviabiliza a análise e processamento por meio de métodos tradicionais (Trujillo, Davis, Du, Damiani, e Storey, 2021, p. 2). A habilidade de interpretação de dados em tempo real decorre da necessidade de gestão dos 5V's do *Big Data*, de modo que seja possível, a partir da interpretação correta e eficiente, fornecer suporte, também em tempo real, para sistemas de decisão (Fertier, Barthe-Delanoë, Montarnal, Truptil, e Bénaben, 2020, p. 1) na área financeira.

Essa gestão é necessária, uma vez que a coleta de dados para análise e aplicação em diversas áreas pode resultar em dados com dimensões elevadas que, normalmente, são muito maiores do que a amostra requerida para as pesquisas a serem desenvolvidas (J.-T. Zhang, Zhou, Guo, e Zhu, 2021, p. 142). Esse contexto representa um desafio chamado de Maldição da Dimensionalidade, em que os padrões de comportamento dos dados altamente dimensionados são mais complexos de serem identificados (Salimi et al., 2018, p. 28), de modo que é preciso identificar como alinhar os processos organizacionais com os dados coletados a fim de auxiliar corretamente em decisões mais embasadas (Horita, Albuquerque, Marchezini, e Mendiondo, 2017, p.

12), considerando o controle de informações em tempo real (Chaturvedi, Hutchinson, e Nazareth, 1993, p. 213).

Com base nisso, a Maldição da Dimensionalidade é a dificuldade advinda da manipulação, administração e interpretação de dados altamente dimensionados, que faz com que os modelos de gestão e análise existentes não sejam suficientes (Aremu, Hyland-Wood, e McAree, 2020, p. 1). O problema resulta do fato de que o número de sorteios⁷ necessários para um processo de estimativa, construído a partir de um modelo de análise e decisão, cresce exponencialmente com o número de parâmetros, o que rapidamente torna a simulação computacional inviável (Cherchi e Guevara, 2012, p. 322). Devido à escassez de dados em espaços multidimensionais, o comportamento de estimadores suaves não paramétricos, ou seja, aqueles que não consideram os parâmetros inerentes aos dados produzidos em larga escala, se deteriora rapidamente à medida que a dimensão aumenta (Nagler e Czado, 2016, p. 69; Lavergne e Patilea, 2008, p. 103). Esse fato, que pode ser observado na Figura 2.9, limita severamente o uso prático de modelos econômicos, uma vez que há a dificuldade de incorporar o comportamento ideal e otimizar a rotina deles (Kompas e Ha, 2019, p. 103).

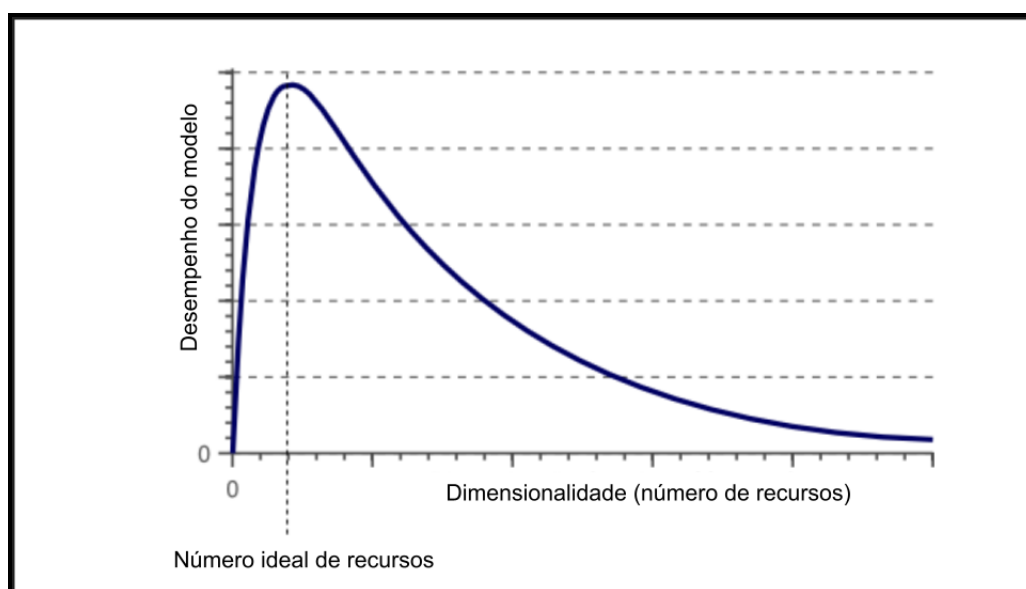


Figura 2.9: Relação entre o comportamento de estimadores e o aumento da dimensionalidade.

Notas: Adaptado de Lenine (2017, p. 1).

A Maldição da Dimensionalidade tem sido um obstáculo vexatório no processamento de consultas em dados multidimensionais (Ma, Zhang, Lin, e Chen, 2012, p. 16). As simulações computacionais podem ser consideradas ferramentas importantes para pesquisas e desenvolvimento de sistemas que identifiquem padrões e comportamentos (Zhou, Lu, Hu, e Hu, 2020, p. 1), como os necessários para a detecção de anomalias, mas o grande volume dos dados configura diversas barreiras para as aplicações práticas, porque as máquinas virtuais não conseguem considerar todas as características dimensionais (Setzer e Bichler, 2013, p. 63). Isso ocorre, uma vez que problemas baseados em dados que sofrem a Maldição da Dimensionalidade não consideram a resolução espacial dos recursos, cujo custo de discretização⁸ cresce conforme o número de dimensões (Pflüger, Peherstorfer, e Bungartz, 2010, p. 508).

Nesse contexto, como a maioria dos estudos estão confinados a sistemas que envolvem dados pouco dimensionados (Grüne, 2021, p. 317), técnicas de redução de modelos são fun-

⁷Nota do autor: Sorteios são extrações de amostras de dados.

⁸Nota do autor: Dificuldade relacionada à criação de blocos de dados separados com valores ordenados e discretos para facilitar a mineração dos dados e, posteriormente, a sua análise.

damentais para lidar com dados altamente dimensionados e gerar amostras equivalentes aos problemas do mundo real (Fernández-Martínez e Fernández-Muñiz, 2020, p. 10; Feuerriegel e Gordon, 2018, p. 89; Lavergne e Patilea, 2008, p. 103). No entanto, tal abordagem, por mais que possa sanar parte do desafio imposto pela Maldição da Dimensionalidade, ela apenas simplifica a interpretação de dados já existentes, e não considera o volume de dados continuamente gerados em tempo real. Os resultados obtidos por Ghasemaghaei, Ebrahimi, e Hassanein (2018, p. 101), em sua pesquisa, indicam que, quando são consideradas todas as dimensões para análise, a qualidade de decisão é aprimorada.

CAPÍTULO 3

MÉTODOS

“Se queres prever o futuro, estuda o
passado.”
—Confúcio.

O presente estudo apresenta uma revisão sistemática de literatura, que fornece uma visão abrangente acerca de uma área de pesquisa (A. d. B. Silva et al., 2020, p. 2), por meio da identificação de artigos relevantes a partir de critérios de seleção (K. Z. Zhang e Benyoucef, 2016, p. 96) e sintetização de dados, questionamentos e técnicas de pesquisas (Ain, Vaia, DeLone, e Waheed, 2019, p. 2). Ela possibilita que seja feita uma análise aprofundada sobre o conhecimento científico a partir da seleção de uma base de dados dominante, na qual é realizada pesquisa, usando uma combinação de palavras-chave relacionadas ao tópico de estudo, e leitura de artigos pertinentes (Hosseini e Ivanov, 2020, p. 2). Isso é feito de modo a identificar artigos revisados por pares que possam auxiliar o processo de tomada de decisões científicas (Rajaeian, Cater-Steel, e Lane, 2017, p. 45), a partir do reconhecimento de perguntas de pesquisa e sugestões de uma agenda de pesquisa para futuras investigações dentro do tema de relevância (Y. Wu, Ngai, Wu, e Wu, 2020, p. 1).

Esse método já foi excelentemente empregado por diversos autores (A. Ahmed, Deokar, e Lee, 2021, p. 3; Henrique, Sobreiro, e Kimura, 2019, p. 227; Amani e Fadlalla, 2017, p. 34; Seuring, 2013, p. 1514) e mostra-se vantajoso para a comunidade científica. Dentro do tema de detecção de anomalias, essa técnica de pesquisa foi empregada recentemente por autores como Al-Hashedi e Magalingam (2021, p. 6), Pourhabibi et al. (2020, p. 1) e Houdek (2020, p. 1), considerando, principalmente, anomalias que sinalizam a existência de fraudes. Com base nisso, o presente estudo visa complementar tais pesquisas anteriores e, ainda, diferencia-se pelo enfoque na identificação técnicas de detecção de anomalias em sistemas financeiros que se manifestam também de outras formas, além de fraude.

Desse modo, o modelo de revisão de literatura a ser desenvolvido foi proposto por Jabbour (2013, p. 145) e Lage Junior e Godinho Filho (2010, p. 14) e segue os cinco passos destacados, de maneira breve, a seguir:

- Primeiro Passo: Realizar uma pesquisa sistemática de artigos disponíveis em uma base de dados confiável acerca das variações de técnicas empregadas para detectar anomalias;
- Segundo Passo: Desenvolver uma estrutura de classificação e codificação;
- Terceiro Passo: Aplicar a estrutura de classificação para sintetizar e apresentar as tendências de estudo sobre detecção de anomalias;
- Quarto Passo: Delinear um *framework* retratando a composição da revisão de literatura, seus principais resultados e o que eles representam de acordo com a codificação preestabelecida;
- Quinto Passo: Analisar e discutir os resultados e, a partir disso, propor sugestões para pesquisas futuras.

Essa técnica de revisão de literatura já foi utilizada e disseminada por diversos artigos publicados em revistas internacionais (Henriques, Sobreiro, e Kimura, 2018, p. 145; J. K. Y. Lee et al., 2021, p. 68415; Mariano, Sobreiro, e Rebelatto, 2015, pp. 34–35; Masudin e Fernanda 2019, p. 2; Nazário, Silva, Sobreiro, e Kimura, 2017, pp. 116–117; Salim, Rahman, e Wahab, 2019, p. 1449). Como os estudos científicos buscam descrever, explicar, prever ou avaliar fenômenos (Abreu, Kimura, e Sobreiro, 2019, p. 197), esse método de revisão possibilita rastrear a evolução e as principais contribuições dentro do campo de estudo (Cruz, Kimura, e Sobreiro, 2019, p. 80) ao mapear e identificar características e tópicos principais dos artigos (Pinheiro et al., 2019, p. 844) e ao agrupá-los de modo a explorar mais a fundo os tópicos de análise (M. C. R. d. C. Ferreira, Sobreiro, Kimura, e Barboza, 2016, p. 114), criando uma completa visão do conhecimento existente (Henriques, Sobreiro, Kimura, e Mariano, 2020, p. 6). Isso é feito por meio da busca e seleção de artigos utilizando uma combinação de palavras-chave que possibilite o aumento do escopo e alcance da pesquisa (Jabbour et al., 2017, p. 291) e da elaboração de um *framework* de classificação que propõe diretrizes para a categorização dos artigos selecionados (Cabral e Dhar, 2019a, p. 26).

A partir disso, essa técnica possibilita que o conteúdo e as características dos artigos possam ser destrinchados e alocados conforme as codificações preestabelecidas, criando uma nova perspectiva de análise literária por meio dos dados estatísticos que são originados (W. Silva, Kimura, e Sobreiro, 2017, p. 92). Com base em tais dados, esse modelo de revisão de literatura irá revelar as tendências de pesquisa não apenas de maneira isolada, mas como também, considerando combinações de vários fatores que envolvem os estudos acerca de detecção de anomalias no setor financeiro, mostrando-se adequado para responder perguntas de pesquisa (Pham, Hoang, e Phan, 2019, p. 848) e revelar lacunas para investigações adicionais (Bajaj, Kashiramka, e Singh, 2020, p. 176). Tais *gaps* de pesquisa permitirão a reflexão acerca de desafios e futuras direções de pesquisa dentro da área de estudo (Abu et al., 2021, p. 755; Khudzari, Kurian, Tartakovsky, e Raghavan, 2018, p. 52).

Os benefícios que podem advir disso encontram-se tanto na esfera acadêmica como na própria esfera financeira, já que esse método permite identificar, avaliar e sintetizar os trabalhos existentes e produzidos por pesquisadores, acadêmicos e profissionais (Okoli, 2015, p. 880). O resumo gráfico resultante poderá funcionar como uma ferramenta visual para que os futuros pesquisadores dentro desse tema possam usar como embasamento ou ponto de partida de modo

a desenvolver pesquisas mais direcionadas ou pertinentes à área, por representar um modo criativo, conciso e impactante de comunicar informações científicas (Ojha, 2021, p. 218). Subsequentemente, esses estudos podem impactar positivamente nos sistemas financeiros que utilizarem tais modelos mais precisos, porque sistemas de detecção de fraude poderosos conseguem não só detectar mas, como também, prevenir anomalias, reduzindo as possíveis perdas financeiras (Höppner et al., 2022, p. 291) e possibilitando uma maior proteção dos dados (Elsafy e Hasan, 2021, p. 2), o que viabiliza a criação de ambientes saudáveis para o desenvolvimento e transformação do ecossistema de negócios digitais (Hanafizadeh e Kim, 2020, p. 1).

3.2 COLETA DE DADOS

Em termos amplos, o primeiro passo dessa pesquisa foi conduzido em duas partes: entre junho e julho de 2021, que originou um estudo preliminar de mesma autoria (Pinto e Sobreiro, 2022, p. 2) e, posteriormente, em outubro de 2022 para atualizar os dados obtidos e contemplar o período de 2022. Foi feita uma busca meticulosa utilizando palavras-chave, referentes ao tema de estudo, contidas nos títulos e resumos dos artigos a serem escolhidos e analisados. Para realizar tal pesquisa, foi escolhida a base de dados *ScienceDirect*[®], operada pela *Elsevier*[®]. No presente momento, essa base possui um portfólio de 2.650 revistas com mais de 24.000 editores (Plc, 2021, p. 14), subdivididos entre as áreas Ciências da Saúde, Ciências da Vida, Ciências Físicas e Engenharia e Ciências Sociais e Humanas, e possibilita o acesso internacional à publicações científicas de diversas áreas. Os conteúdos científicos são distribuídos e acessados por meio da *ScienceDirect*[®], que integra pesquisadores a pesquisas médicas e científicas revisadas por pares (Plc, 2021, p. 14). Isso é feito por meio do acesso a artigos científicos a partir de palavras-chave, nome do autor, título da revista e outros tipos avançados de pesquisa.

Logo, de maneira mais detalhada, para realizar a presente revisão de literatura, as palavras-chave incluíram “*anomaly detection finance*”, “*outlier detection finance*”, “*anomaly detection economics*”, “*anomaly detection financial fraud*”, “*anomalous activity finance*”, além de outras variações. Ainda, foram usadas palavras na seção de pesquisa avançada utilizando o marcador booleano “*OR*” a fim de aumentar as chances de sucesso da pesquisa, como, “*financial*”, “*finance*”, “*fraud*”, “*machine learning*”, “*anomaly detection*”, “*detection*”. Para afunilar ainda mais a pesquisa na base de artigos, foram utilizados filtros relacionados aos anos de publicação, sendo escolhidos artigos publicados dentre os anos de 2017 e 2021 e, posteriormente, 2022, apenas, visando uma revisão de literatura mais atual e, além disso, como se trata de detecção de anomalias em sistemas financeiros, foram escolhidas as áreas temáticas “Economia, Econometria e Finanças” e “Negócios, Administração e Contabilidade”.

Considerando todas as pesquisas, foi encontrada uma grande quantidade de artigos com as palavras-chave e os filtros estabelecidos. No entanto, após o reconhecimento dos títulos de cada artigo, foi possível perceber que, dentro das variações de palavras-chave, as pesquisas retornavam muitos artigos repetidos, o que diminuiu a quantidade de artigos considerados novos. Com as palavras-chave “*anomaly detection finance*”, por exemplo, na primeira coleta de dados foram encontrados 71 artigos, mas, desse total, somente 22 foram considerados para a revisão de literatura. Além disso, com a consulta aos resumos dos artigos encontrados pode-se identificar pesquisas que não se encaixavam no tema a ser estudado, não envolvendo, por exemplo, métodos ou técnicas de detecção de anomalias, como a pesquisa de Bouzguenda (2018, p. 273) que abordou impacto da inteligência emocional no processo de tomada de decisões financeiras, ou, então, a pesquisa de Jung e Jeong (2020, p. 1) que, em um estudo de caso, explorou técnicas para prever níveis de engajamento em mídias sociais.

Assim, depois da seleção de artigos com base nos critérios elencados, foram identificados

66 artigos relevantes em 2021 e 79 em 2022 para o presente estudo. Porém, a leitura minuciosa de cada um desses artigos remanescentes sinalizou alguns artigos que não eram compatíveis com o critério de seleção, como, o estudo de Frömmel, Han, Li, e Vigne (2021, p. 1) acerca de anomalias de baixa liquidez no beta de ações da China, com enfoque na exploração de tais anomalias e não em técnicas de detecção de anomalias dentro do contexto do mercado financeiro Chinês. Outro exemplo de artigo desconsiderado foi a pesquisa de Cumming, Groh, e Johan (2018, p. 130) que abordou abuso de mercado, analisando diferenças de intensidade de execução da lei em países Europeus, não discutindo, portanto, questões relacionadas a detecção de anomalias ou fraudes que possam causar tal abuso. Outro artigo retirado da base de literatura foi o estudo promovido por Pinto e Sobreiro (2022, p. 3), uma vez que ele possui muita semelhança com o presente trabalho e poderia influenciar nos dados estatísticos. Em suma, a quantidade de artigos considerada para a presente revisão de literatura que se encaixaram nos critérios de seleção é de 74 artigos. Esse histórico de pesquisa e seleção de artigos é apresentado na Tabela 3.1.

É oportuno destacar que, os artigos selecionados focam em métodos e técnicas para a detecção de anomalias, fraudes e *outliers* em sistemas financeiros diversos, e um breve resumo de cada um desses estudos pode ser visualizado na Tabela 3.2.

Palavras-chave	Artigos 2017-2021		
	Encontrados na ScienceDirect®	Adicionados à Base de Estudo	Repetido ou Retirado
<i>Anomaly detection finance.</i>	71	22	49
<i>Outlier detection finance.</i>	169	7	162
<i>Literature review anomaly detection finance.</i>	118	5	113
<i>Anomaly detection economics.</i>	69	3	66
<i>Anomaly detection financial fraud.</i>	304	23	281
<i>Credit card fraud detection.</i>	70	1	69
<i>Anomalous activity finance.</i>	85	5	80
Total de Artigos.	-	66	-
<i>Excluídos após a leitura.</i>	-	-	18
Total de Artigos.	-	48	-
Palavras-chave	Artigos 2021-2022		
	Encontrados na ScienceDirect®	Adicionados à Base de Estudo	Repetido ou Retirado
<i>Anomaly detection finance.</i>	28	26	2
<i>Outlier detection finance.</i>	29	21	8
<i>Literature review anomaly detection finance.</i>	20	2	18
<i>Anomaly detection economics.</i>	23	9	14
<i>Anomaly detection financial fraud.</i>	15	3	12
<i>Credit card fraud detection.</i>	23	10	13
<i>Anomalous activity finance.</i>	15	8	7
Total de Artigos.	-	79	-
<i>Excluídos após a leitura.</i>	-	-	53
<i>Total de Artigos 2017-2021.</i>	-	48	-
Total de Artigos - Final.	-	74	-

Tabela 3-1: Número de artigos selecionados da ScienceDirect®.

Artigos	Breve Resumo
Al-Hashedi e Magalingam (2021).	Propuseram uma revisão de literatura para complementar a amostra de artigos revisados por outros autores e, ainda, para apresentar um resumo abrangente de pesquisas recentes acerca do uso de técnicas de mineração de dados de modo a promover a detecção de fraudes na área financeira.
Andor e Bohák (2017).	Propuseram uma abordagem inovadora que promove a identificação de eventos significantes em séries temporais de preços financeiros a partir de testes estatísticos para explorar a detecção de mudanças abruptas na amostra utilizada.
Arboreda et al. (2018).	Utilizaram operadores baseados em operações de <i>slice and dice</i> ^a , Lei de Benford e técnicas forenses de detecção de fraude contábil para identificar sinais de existência de possíveis fraudes em armazéns de dados.
Ardizzi et al. (2018).	Implementaram um modelo de análise econométrica acerca de depósitos financeiros feitos na Itália para definir uma série de indicadores de riscos que possam distinguir componentes ilegais e anômalos daqueles considerados legais e dentro do nível de normalidade.
Arévalo et al. (2022).	Implementaram um modelo não supervisionado com técnicas de <i>machine learning</i> que identifica grupos de dados e detecta anomalias em um sistema de pagamentos de El Salvador.
Aslam, Hunjra, Friti, Loughichi, e Shams (2022).	Propuseram um <i>framework</i> de análise de dados com base em <i>machine learning</i> , estatística e estatística Bayesiana para a detecção de fraudes em sistemas de seguro de automóveis.
Ausloos, Ficcadenti, Dhési, e Shakeel (2021).	Exploraram a conformidade das distribuições dos valores de fechamento diários da S&P 500 e dos correspondentes retornos logarítmicos diários com os dígitos significativos da Lei de Benford.
Azevedo et al. (2021).	Avaliaram a aplicação da Lei de Benford como técnica para detectar fraudes de pagamentos promovidos pelo programa de benefícios brasileiro Bolsa Família de modo a selecionar regiões que propiciam uma distribuição suspeita de capital.

Continua na próxima página.

^aQuando informações e dados são quebrados em pequenas partes para que possam ser analisadas com mais facilidade a partir de diferentes perspectivas.

Artigos	Breve Resumo
Badal-Valero et al. (2018).	Apresentaram uma ferramenta para identificar, com base em dados contábeis, empresas fraudulentas, detectar criminosos que promovem ações de lavagem de dinheiro e reduzir a probabilidade de erros na distinção entre empresas com operações ilegais e aquelas operando corretamente.
Baesens et al. (2021).	Apresentaram o uso de técnicas de engenharia de dados, decompostas em engenharia de recursos e engenharia de instâncias, para aperfeiçoar a performance de um modelo analítico de detecção de fraudes com base em dados reais de transações de pagamento de um banco Europeu.
Baker, Cohanier, e Leo (2017).	Examinaram falhas que possibilitavam a detecção de fraudes no sistema de informações de trocas comerciais, que causaram um colapso financeiro em um banco Europeu, de modo identificar oportunidades de controle adequadas.
Belle, Baesens, e Weerdt (2022).	Propuseram um método de detecção de fraude em cartões de crédito com base em rede de aprendizado de representação, considerando a estrutura relacional das transações a partir da análise de dados reais.
Byers et al. (2021).	Analisaram a presença de <i>outliers</i> em dados de preços futuros de <i>commodities</i> por meio de um algoritmo de <i>machine learning</i> , que também é utilizado para investigar o impacto desses <i>outliers</i> nas medidas de riscos das <i>commodities</i> .
Canhoto (2021).	Promoveu uma investigação acerca do uso de algoritmos de <i>machine learning</i> para detectar transações financeiras suspeitas, além de prever e evitar ações de lavagem de dinheiro e financiamento de terrorismo.
Carmona, Climent, e Momparler (2019).	Exploraram o uso do <i>Extreme Gradient Boosting</i> , um método de <i>machine learning</i> , para aperfeiçoar a precisão de predição de falências bancárias nos EUA a partir da identificação de variáveis chave que podem antecipar e prever padrões bancários e, ainda, prevenir crises financeiras.
Carneiro et al. (2017).	Apresentaram o desenvolvimento e implementação de um sistema de pontuação de riscos que promove a detecção de fraudes de cartão de crédito com base em métodos de <i>machine learning</i> em uma empresa de varejo <i>online</i> .

Continua na próxima página.

Artigos	Breve Resumo
Carta, Fenu, Recupero, e Saia (2019).	Apresentaram uma técnica inovadora que combina <i>machine learning</i> e estatística para maximizar a capacidade de detectar transações de cartão de crédito fraudulentas em um <i>e-commerce</i> .
Charakopoulos e Karakasidis (2022).	Propuseram o uso de um método computacional gráfico para identificar pontos de mudanças e anomalias em séries temporais do mercado financeiro.
K. Chen, Yadav, Khan, e Zhu (2020).	Usaram algoritmos de <i>machine learning</i> combinados com modelos de pontuação de crédito para definir um modelo híbrido de detecção de fraude em cartões de crédito passível de sinalizar a ocorrência de divergências entre a situação atual de uma pessoa e a sua situação em um futuro próximo.
L. Chen et al. (2022).	Apresentaram uma abordagem de aprendizado multivariado com base em técnicas estatísticas para identificar fraudes em pedidos de empréstimos financeiros.
Chu, Wang, Lin, e Li (2022).	Implementaram um processo operacional criado a partir de uma simulação do método estatístico de Monte Carlo para evitar adulteração de dados e detectar fraudes em uma rede <i>blockchain</i> de consórcio.
Climent et al. (2019).	Aplicaram um método de <i>machine learning</i> para construir um modelo de classificação que identifica e monitora um conjunto de indicadores e variáveis relevantes que podem auxiliar na predição e prevenção de falência bancária na Zona Euro.
Dinh, Powell, e Vo (2021).	Implementaram um modelo estrutural que calcula distâncias padrão com base no mercado de modo a investigar e prever o surgimento de dificuldades financeiras em bancos e mercados financeiros asiáticos.
Druică et al. (2018).	Analisaram a conformidade dos dois primeiros dígitos dos saldos bancários anuais de um banco na Romênia com a Lei de Benford com base em testes estatísticos.

Continua na próxima página.

Artigos	Breve Resumo
Forough e Momtazi (2021).	Apresentaram um modelo que combina modelagem sequencial de dados com algoritmos de redes neurais para obter uma melhor performance na detecção de fraudes em cartões de crédito.
Gianini et al. (2020).	Desenvolveram um modelo de gerenciamento de regras para detecção de fraudes de cartão de crédito a partir de um índice de potência baseado em Teoria dos Jogos que avalia e atualiza as regras e a performance de suas aplicações.
González-Sánchez (2021).	Analisou a existência de <i>ouliers</i> em mercados financeiros e investigou se tais desvios provocam o processo de escala de tempo e quais os efeitos comportamentais decorrentes dessa possível correlação.
Gu, Lin, e Wu (2022).	Propuseram um modelo de análise de transações que envolvem troca de <i>criptomoedas</i> a partir de uma técnica de <i>machine learning</i> para detectar anomalias e situações de lavagem de dinheiro.
Herrera, Climent, Carmona, e Momparler (2022).	Propuseram um modelo de análise dos dados financeiros do <i>Euro Interbank</i> para com base em técnicas de <i>machine learning</i> para detectar anomalias e identificar se as informações foram manipuladas.
Houdek (2020).	Propôs uma revisão de literatura em que é possível explorar vantagens e desvantagens de métodos que possam revelar comportamentos fraudulentos e desonestos em organizações, assim como os prós e contras de cada um desses métodos.
James, Menzies, e Chan (2021).	Apresentaram um modelo baseado em técnicas estatísticas para avaliar o impacto da pandemia de COVID-19 em ativos financeiros, especialmente no mercado de <i>criptomoedas</i> , identificando inconsistências entre comportamento extremo e errático, que se traduzem em anomalias.
Jurgovsky et al. (2018).	Apresentaram uma técnica de redes de memória de curto prazo longa para compreender o comportamento de compra individual de cada titular de cartão de crédito de modo a aperfeiçoar a detecção de fraudes em futuras transações.

Continua na próxima página.

Artigos	Breve Resumo
Kou et al. (2021).	Apresentaram um modelo de <i>machine learning</i> para identificar e prever falências bancárias a partir de dados provenientes de transações e demonstrações financeiras.
Kristóf e Virág (2022).	Implementaram técnicas de <i>machine learning</i> para construir modelos de detecção e prevenção de falência bancária com base em informações financeiras de bancos da União Europeia.
León (2020).	Apresentou um método de investigação de redes de pagamento de modo a detectar mudanças atípicas nos recursos de conexão, os quais desviam do que é considerado uma estrutura de rede normal, sinalizando uma rede anômala.
León et al. (2020).	Abordaram o uso de um método de <i>machine learning</i> para compreender o comportamento de sistemas de pagamento de instituições financeiras da Colômbia que serve como ferramenta para monitorar e detectar anomalias.
Lin et al. (2018).	Investigaram as manipulações contábeis feitas por diretores de empresas chamados de <i>fat cats</i> e seus impactos em contraste com o modelo de distribuição de números significativos da Lei de Benford.
Q. Liu, Wang, Zhang, e Zheng (2021).	Aplicaram técnicas de <i>machine learning</i> para identificar manipulações e anomalias em sistemas financeiros, utilizando indicadores extraídos de relatórios de classificação de analistas e notícias financeiras.
Lucas et al. (2020).	Propuseram um modelo de <i>machine learning</i> que quantifica a similaridade entre sequências de transações observadas e sequências de transações fraudulentas de cartão de crédito de uma empresa europeia, de modo a aperfeiçoar o processo de detecção de fraudes.
Mamede e Malaquias (2017).	Analisaram a diferença de padrões de comportamento em ações de fundos de cobertura brasileira, de modo a identificar e compreender anomalias provocadas pelo chamado <i>Monday Effect</i> ^a que fazem com que os retornos de investimentos sejam anormais.

Continua na próxima página.

^aEm tradução livre: Efeito segunda-feira.

Artigos	Breve Resumo
Misra et al. (2020).	Propuseram um modelo de <i>machine learning</i> com dois estágios, em que primeiro são extraídos recursos do conjunto de dados para promover o aprendizado de atributos transacionais significativos para depois identificar se há fraudes em transações de cartão de crédito.
Mustafa, Lin, e Kakinaka (2022).	Implementaram um modelo com base na técnica de redes neurais de <i>machine learning</i> capaz de reconhecer padrões estruturais no mercado financeiro de modo a identificar mudanças que se traduzem em anomalias.
Nikkel (2020).	Apresentou uma reflexão acerca do reconhecimento de <i>Fintech Forensics</i> como uma ferramenta forense digital para detectar ações criminosas que abusam das tecnologias financeiras para cometer fraude, lavagem de dinheiro e extorsão.
Nissim (2022).	Propôs uma revisão acerca de pesquisas que utilizaram métodos que envolvem <i>big data</i> ou <i>machine learning</i> para promover <i>insights</i> relevantes acerca de dados contábeis.
Papík e Papílková (2022).	Propuseram um modelo de desempenho superior ao relatado na literatura anterior de mineração de dados e <i>machine learning</i> para identificar fraudes em demonstrações contábeis de instituições dos Estados Unidos.
Parnes (2020).	Promoveu a investigação de anomalias no índice de mercado S&P 500 em diferentes aspectos econômicos para detectar padrões de comportamento e, ainda, examina desvios entre o índice de mercado S&P 500 e nove indicadores financeiros americanos.
Patil, Nemade, e Soni (2018).	Propuseram um modelo de <i>framework</i> analítico e auto adaptativo com o objetivo de identificar fraudes em cartão de crédito e, ainda, compararam a performance de diversos algoritmos de detecção de fraude de modo a aprimorar o modelo.
Pourhabibi et al. (2020).	Apresentaram uma revisão de literatura para consolidar o conhecimento existente acerca do uso de <i>Graph-Based Anomaly Detection techniques</i> no contexto de detecção de fraudes financeiras a partir de uma mapeamento de tendências de pesquisas, métodos e desafios.

Continua na próxima página.

Artigos	Breve Resumo
R.B. e K.R. (2021).	Utilizaram diferentes algoritmos de <i>machine learning</i> para identificar a melhor performance de detecção de transações fraudulentas feitas com cartões de crédito.
Rocha-Salazar, Segovia-Vargas, e del Mar Camacho-Miñano (2021).	Apresentaram um modelo de detecção de transações financeiras suspeitas relacionadas à lavagem de dinheiro e financiamento de terrorismo a partir de um indicador de anormalidade com base em variáveis que possibilitam análises comportamentais das transações.
Rtayli e Enneya (2020).	Desenvolveram um novo modelo híbrido que combina métodos e técnicas de <i>machine learning</i> para aperfeiçoar a eficiência de predição de transações fraudulentas de cartão de crédito.
Rubio et al. (2020).	Desenvolveram um <i>autoencoder</i> para um sistema de pagamentos bancários do Equador para compreender padrões de pagamento e detectar possíveis anomalias a partir de redes neurais.
Sá et al. (2018).	Apresentaram o uso de um algoritmo classificador de rede bayesiano como método de detecção de fraudes em transações de cartão de crédito de modo a classificar e distinguir transações legítimas de transações fraudulentas.
Sabeti e Heijmans (2021).	Desenvolveram um <i>autoencoder</i> com base em métodos de <i>machine learning</i> que reconhece padrões em um sistema de pagamento varejista que não estão em conformidade com o comportamento esperado, sendo considerados anomalias.
Semeels (2022).	Apresentou três formas de detectar atividades suspeitas de negociação de <i>non-fungible tokens</i> (NFTs), direitos de propriedade digital que surgiram como uma nova classe de ativos financeiros.
Severino e Peng (2021).	Apresentaram uma análise comparativa de como o uso diversos modelos de <i>machine learning</i> podem contribuir para a detecção de fraudes em políticas de seguro de propriedade.

Continua na próxima página.

Artigos	Breve Resumo
Shi et al. (2018).	Conduziram um teste estatístico para avaliar se os índices de preços de relatórios financeiros de países com economia não madura estavam em conformidade com os dígitos significativos da Lei de Benford ou se possuíam valores anômalos.
K. Singh e Best (2019).	Demonstraram a aplicação de técnicas de visualização que promovem a investigação de transações financeiras a partir da detecção de padrões que representam potenciais atividades de lavagem de dinheiro.
Stripling et al. (2018).	Apresentaram uma abordagem de <i>machine learning</i> que promove uma detecção de anomalias baseada em isolamento de dados de uma empresa europeia de seguros a partir da computação condicional de anomalias consideradas úteis para a identificação de fraudes.
Subudhi e Panigrahi (2020).	Propuseram um novo modelo híbrido que aplica algoritmos genéticos com técnicas de agrupamento para aprimorar o processo de identificação de fraudes em seguros de automóveis.
Tardelli, Avvenuti, Tesconi, e Cresci (2021).	Apresentaram dois métodos de detecção automática de desinformação financeira em uma rede social por meio de classificação e regressão para estimar a quantidade de conteúdo inorgânico nas discussões financeiras.
Vandervorst, Verbeke, e Verdonck (2022).	Propuseram um modelo de análise baseado em estimativas de densidade condicional para auxiliar companhias de seguro na detecção de fraudes e <i>outliers</i> , que se estende, também, a dados multivariados.
Y. Wang e Xu (2018).	Apresentaram um modelo de identificação de fraude em seguro de automóveis que combina o uso de métodos de mineração de dados com a experiência de análise de especialistas envolvidos com os dados para aumentar a precisão de detecção.
Wei, Peng, e Wu (2021).	Apresentaram um sistema que demonstra que a educação financeira pode colaborar para o aumento da eficiência de detecção de fraudes em sistemas financeiros com base em um modelo de regressão estatística.

Continua na próxima página.

Artigos	Breve Resumo
Wever, Shah, e O’Leary (2022).	Projetaram sistemas de alerta precoce para detectar riscos sistêmicos em sistemas financeiros de modo contínuo e automatizado a partir de inteligências artificiais e avanços tecnológicos.
Yan et al. (2020).	Propuseram um método que combina <i>machine learning</i> , estatística e simulações computacionais para promover a detecção de fraudes em seguros de automóveis.
Yankol-Schalck (2022).	Apresentou uma ferramenta baseada em técnicas de <i>machine learning</i> para prever se um pedido de seguro é real ou fraudulento com base nas informações financeiras disponíveis em uma empresa de seguros de automóveis.
Yoon, Liu, Chiu, e Vasarhelyi (2021).	Apresentaram um sistema de auditoria contínua para ser implementado em tempo real em demonstrações financeiras, buscando prevenir a existência de alarmes e fraudes.
Yuan, Chi, Zhou, e Yin (2022).	Apresentaram um modelo híbrido com duas técnicas de <i>machine learning</i> capaz de identificar e prever possíveis inadimplências em instituições financeiras.
Zhai, Cao, Yao, Ding, e Li (2017).	Apresentaram um modelo híbrido composto por técnicas de mineração de dados capazes de identificar padrões e comportamentos anômalos em trocas comerciais feitas em um mercado de ações americano.
W. Zhang, Xie, Wang, Yang, e Li (2022).	Propuseram um modelo que combina estatística generativa com algoritmos de <i>machine learning</i> para detectar fraudes em avaliações de plataformas de <i>e-commerce</i> que podem afetar o comportamento do consumidor e, conseqüentemente, as vendas de produtos.
G. Zhang, Atasoy, e Vasarhelyi (2022).	Apresentam um <i>framework</i> de controle e monitoramento em conjunto com técnicas de <i>machine learning</i> para avaliar riscos, violações de controles internos e identificar anomalias em um sistema de folha de pagamento da área da saúde.
Y. Zhang, Hu, Wang, e Zhang (2022).	Implementaram um modelo de detecção de fraudes em sistemas financeiros da China com base em <i>machine learning</i> e na técnica de representação de saco-de-palavras.
Zoričák, Gnip, Drotár, e Gazda (2020).	Apresentaram uma ferramenta de predição de falências bancárias que, a partir da detecção de atributos financeiros relevantes, consegue identificar quais empresas são saudáveis e quais não são conforme dados de anormalidade financeira.

Tabela 3.2: Breve resumo dos artigos selecionados.

Após a seleção desses artigos, como muito bem indicado nos artigos de Jabbour (2013, p. 145) e Lage Junior e Godinho Filho (2010, p. 14), foi realizado o segundo passo do método de pesquisa, em que os artigos foram classificados dentro de categorias de análise com base em códigos de distinção. O terceiro passo foi feito de modo a sintetizar o conhecimento científico já existente para viabilizar uma visão mais abrangente acerca da área de estudo. O quarto passo, em consonância com o terceiro, foi feito para possibilitar um entendimento visual dos resultados e tendências da literatura atual em detecção de anomalias financeiras. Por fim, o quinto passo concluiu a revisão por meio da discussão dos resultados e proposição de uma agenda de pesquisa para futuras investigações. Todos os passos necessários a essa pesquisa podem ser visualmente compreendidos por meio do *graphical abstract* apresentado na Figura 3.1, inspirado no estudo desenvolvido por Al-Hashedi e Magalingam (2021, p. 5) e Pourhabibi et al. (2020, p. 3).

3.3 CLASSIFICAÇÃO E CODIFICAÇÃO

Após a seleção e leitura dos artigos da presente base de estudos, foi construído um *framework* de classificação a partir de códigos para categorizar os artigos de modo pertinente ao tema de pesquisa da revisão. A estrutura dele é composta por 10 categorias de classificação, numeradas de 1 a 10, que são combinadas com uma codificação apresentada em forma de letras (A, B, C, D, e assim por diante) justapostas aos números, que representam subcategorias de classificação para promover uma análise mais aprofundada acerca das tendências de pesquisa no contexto de métodos e técnicas de detecção de anomalias em sistemas financeiros. Na Tabela 3.3, é apresentado o *framework* de classificação. Considerando esse sistema de classificação, os artigos foram analisados e incorporados à estrutura, sendo atribuído a cada um deles códigos referentes a cada categoria de análise. Vale destacar que um artigo pode receber mais de um código em uma determinada categoria, uma vez que pode se encaixar em mais de um subtópico de classificação.

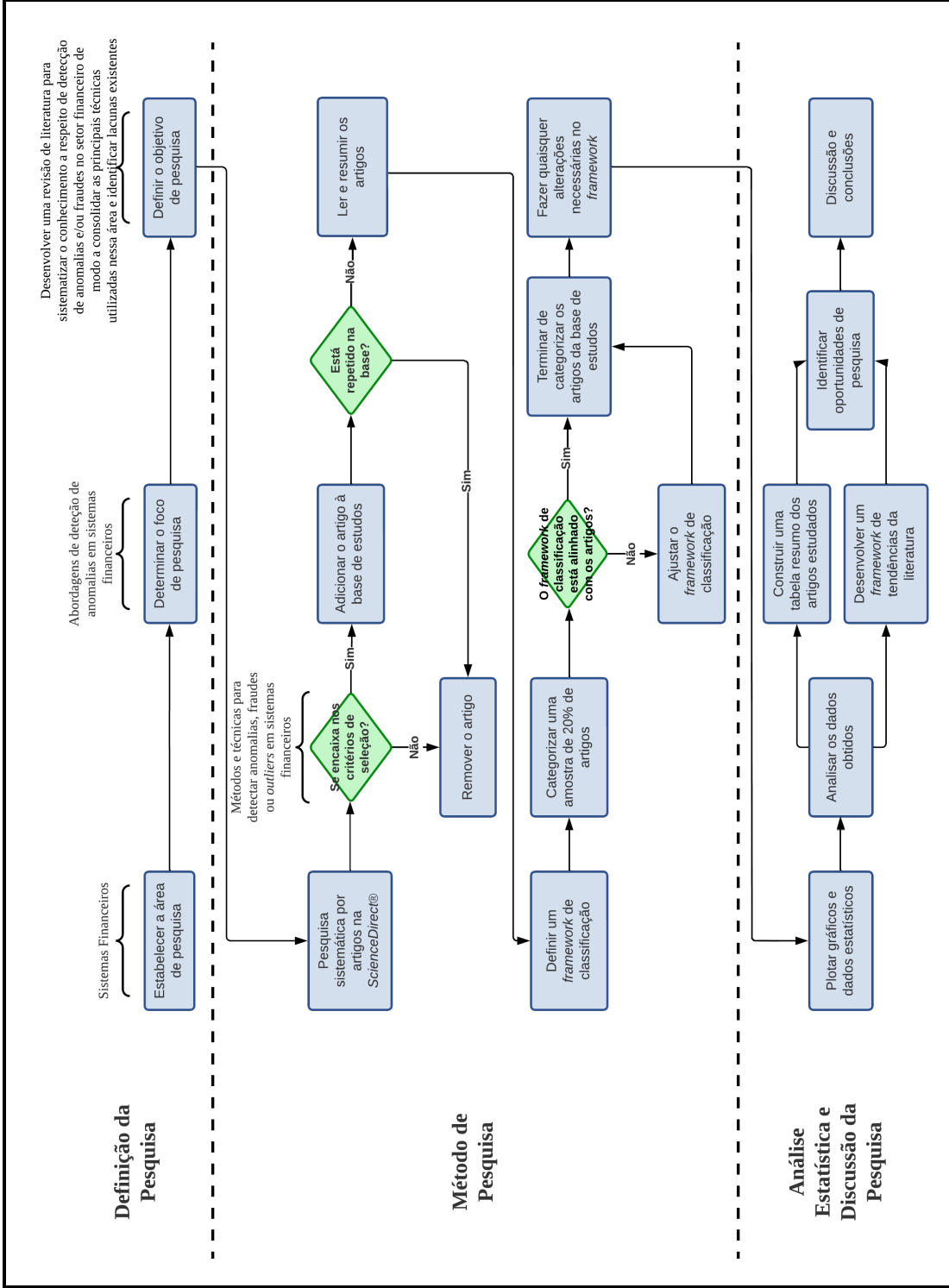


Figura 3-1: Graphical abstract da revisão de literatura.

Notas: Baseado nos estudos de Pourhabibi et al. (2020, p. 3), Al-Hashedi e Magalingam (2021, p. 5) e Pinto e Sobreiro (2022, p. 7).

Classificação	Significado	Códigos
1	Contexto Econômico.	1A - Economia madura. 1B - Economia não madura. 1C - Não informado. 1D - Não se aplica.
2	Localização Geográfica de Pesquisa.	2A - América. 2B - Europa. 2C - Ásia. 2D - África. 2E - Oceania. 2F - Todos os continentes. 2G - Outros. 2H - Não se aplica.
3	Método de Pesquisa.	3A - Qualitativo. 3B - Quantitativo. 3C - Conceitual. 3D - Modelagem. 3E - <i>Survey</i> . 3F - Outras abordagens quantitativas. 3G - Estudo de caso único. 3H - Múltiplos estudos de caso.
4	Área de Aplicação.	4A - Banco. 4B - Atuarial. 4C - Mercado financeiro. 4D - Seguro social. 4E - Contabilidade financeira. 4F - Outros. 4G - Não se aplica.
5	Foco.	5A - Cartão de crédito. 5B - Seguro. 5C - Benefícios de bem-estar financeiro. 5D - Transações financeiras. 5E - Declarações financeiras. 5F - Ações e <i>commodities</i> . 5G - Outros. 5H - Não se aplica.

Continua na próxima página.

Classificação	Significado	Códigos
6	Base de Dados Utilizada.	6A - Real. 6B - Sintética/Artificial. 6C - Outros. 6D - Não se aplica.
7	Disponibilidade.	7A - Acesso restrito. 7B - Acesso irrestrito. 7C - Outros. 7D - Não se aplica.
8	Técnica Utilizada.	8A - <i>Machine learning</i> . 8B - Estatística. 8C - Estatística Bayesiana. 8D - Lei de Benford. 8E - Métodos gráficos (visuais). 8F - Simulações computacionais. 8G - Outros. 8H - Não se aplica.
9	Número de Estratégias.	9A - Uma. 9B - Duas. 9C - Três. 9D - Quatro ou mais. 9E - Outros. 9F - Não se aplica.
10	Área do Periódico.	10A - Ciências da saúde. 10B - Ciências da vida. 10C - Ciências físicas e engenharia. 10D - Ciências sociais e humanas.

Tabela 3.3: *Framework* de classificação e codificação.

3.3.I CONTEXTO ECONÔMICO

A Classificação I envolve o Contexto Econômico em que os artigos estão inseridos, considerando os países dos dados que são utilizados para desenvolver as pesquisas. A codificação foi feita por meio das letras A-D. As codificações “IA - Economia madura” e “IB - Economia não madura” foram subcategorias escolhidas com base no trabalho de Jabbour (2013, pp. 145–146), com o objetivo de identificar se o tipo da economia possui alguma relação com o desenvolvimentos de atividades fraudulentas e criminosas que possam levar à detecção de anomalias nos sistemas financeiros, considerando que economias de mercado maduras possuem estruturas econômicas mais estáveis, enquanto economias ainda em desenvolvimento, normalmente, passam por diversas mudanças estruturais (Borgersen e King, 2014, p. 102) que podem viabilizar tais ações. Essa categoria correlaciona as perspectivas econômicas das regiões com as interferências provocadas nos sistemas financeiros que causam as anomalias, de modo que possibilita identificar se há uma relação entre o desenvolvimento de mais ou menos estudos em países de economia madura ou não maduras ao longo dos últimos anos (Henriques et al., 2018, p. 146) ao sinalizar a existência de incidentes onde preocupações financeiras estão alinhadas com detecção de anomalias (M. C. R. d. C. Ferreira et al., 2016, p. 8). Ela é relevante, uma vez que estratégias de autorregulação utilizadas em diferentes cenários podem ser dotadas de diferentes abordagens, o que se traduz em diferentes respostas comportamentais, que podem facilitar ou não o surgi-

mento de anomalias que possam alterar contextos econômicos (Florack, Keller, e Palcu, 2013, p. 127).

3.3.2 LOCALIZAÇÃO GEOGRÁFICA DE PESQUISA

Em consonância com a categorização anterior, foi estruturada a Classificação 2, que sinaliza a Localização Geográfica de Pesquisa ao indicar os continentes relacionados aos países dos dados. Elencar a localização geográfica onde os artigos foram desenvolvidos possibilita não somente uma análise correlata com fatores econômicos provenientes da primeira classificação, mas como, também, permite a examinação de efeitos combinados entre o conhecimento geográfico e a atividade de pesquisa geográfica de maneira mais detalhada (Hohberger e Wilden, 2022, p. 706). Isso, porque os estudos desenvolvidos em diferentes localidades exibem características particulares (Boubakri, Guedhami, e Saffar, 2016, p. 364; Henriques et al., 2018, p. 146) que refletem nos recursos disponíveis, no nível de comunicação e no grau de prosperidade e atividade financeira das organizações (T.-K. Chen, 2016, p. 323), afetando também habilidades de controle e monitoramento de dados (Boubaker, Derouiche, e Lasfer, 2015, p. 24).

Logo, a análise geográfica complementa a primeira classificação de um modo mais específico (M. C. R. d. C. Ferreira et al., 2016, p. 9), buscando identificar se as anomalias financeiras podem estar relacionadas com regiões ou países específicos, podendo ser considerado algo endêmico. Essa classificação mostra-se pertinente, uma vez que países ao redor do globo possuem diferentes níveis de maturidade em seus sistemas econômicos, financeiros e legais (Nazário et al., 2017, p. 118), o que pode refletir no combate às anomalias quando elas são geradas a fim de criar ganhos ou prejuízos financeiros, bem como no aumento ou diminuição da discussão acadêmica perante a existência de anomalias financeiras. A codificação foi feita considerando as letras A-H, visando identificar nichos ou grupos de pesquisas específicos a respeito da temática, que podem viabilizar percepções acerca de como as anomalias são vistas e tratadas conforme o local em que são identificadas, bem como se esse fator influencia nas abordagens acadêmicas desenvolvidas.

Os artigos, nessas duas categorias, classificados como “IC - Não informado” e “2G - Outros” foram aqueles que utilizaram algum tipo de dado para a construção da pesquisa, mas não especificaram a origem geográfica deles, como, por exemplo, o estudo desenvolvido por K. Chen et al. (2020, p. 3) que adotou um conjunto de dados com transações de cartão de crédito selecionados dos repositórios *Kaggle*¹ e *UCI*² de *machine learning* para apresentar seu método de pesquisa, mas não explicita os países dos dados escolhidos. Já, os artigos classificados como “1D - Não se aplica” na categoria de contexto econômico e “2H - Não se aplica” na categoria de localização geográfica de pesquisa, não se encaixam nas formas de classificação previamente estabelecidas. Os artigos assim classificados, por exemplo, não apresentam dados que remetem a fontes geográficas de estudo, como a revisão de literatura desenvolvida por Pourhabibi et al. (2020, pp. 2–3), que utilizou artigos dentro do tema de detecção de fraudes, publicados em periódicos internacionais para promover uma análise aprofundada acerca de técnicas de detecção de anomalias baseadas em gráficos nessa área.

3.3.3 MÉTODO DE PESQUISA

A Classificação 3, estruturada na escala de codificação de A-H, sinaliza o Método de Pesquisa utilizado pelos autores no desenvolvimento de suas pesquisas a respeito de anomalias. A ideia por trás dessa classificação é a de reconhecer quais os métodos mais utilizados dentro do campo de detecção de anomalias de modo a identificar as tendências de pesquisa. Considerando

¹Kaggle website: <https://www.kaggle.com/>.

²UCI Machine Learning website: <https://archive.ics.uci.edu/ml/index.php>.

que a identificação de padrões relaciona-se, muitas vezes, com simulações computacionais diversas por meio de algoritmos, estatísticas e distribuições, como, por exemplo, com o uso de *machine learning*, estatística Bayesiana e Lei de Benford, é esperado que a maioria dos artigos sejam classificados como “3B - Quantitativo”, pois a adoção de modelos quantitativos promove análises promissoras e rápidas que podem mapear mais facilmente o objeto de estudo e evidenciar percepções relevantes para a pesquisa (Ciaccio et al., 2019, p. 120).

É oportuno destacar que a classificação “3C - Conceitual” foi utilizada para os artigos que promovem uma revisão de literatura, uma vez que tais artigos são consonantes com os objetivos de uma pesquisa conceitual, que se pauta na observação, investigação e análise de métodos e conceitos para identificar lacunas de pesquisa (Soulé, Michonneau, Michel, e Bockstaller, 2021, p. 2). No entanto, tal classificação foi criada também para identificar a existência de pesquisas conceituais que estabeleçam práticas comuns a serem desenvolvidas nos estudos de detecção de padrões anômalos de modo a compreender melhor se há um padrão estrutural de realização das pesquisas que permita que haja comparações mais significantes entre elas.

3.3.4 ÁREA DE APLICAÇÃO

A Classificação 4 refere-se à Área de aplicação na qual as pesquisas desenvolvem os métodos de detecção de anomalias, com uma escala de codificação que vai de A-G. Essas áreas foram escolhidas com base nas tendências de estudos desenvolvidas nos artigos e relacionam os diversos setores do ramo financeiro onde podem ocorrer fraudes (Azevedo et al., 2021, pp. 1–2; Yan et al., 2020, pp. 115–116; Sá et al., 2018, pp. 21–22), adulteração de dados (Ausloos et al., 2021, pp. 2–3; Druică et al., 2018, pp. 77–79; Shi et al., 2018, pp. 879–880), desvio de capital (K. Singh e Best, 2019, pp. 1–2; Ardizzi et al., 2018, pp. 105–106; Badal-Valero et al., 2018, pp. 24–25), entre outras ações que podem indicar anormalidades nos sistemas.

Essa classificação possui como objetivo principal identificar as diversas ferramentas ou técnicas usadas para identificar as anomalias considerando uma área específica, de modo que as expertises advindas delas, bem como a abordagem considerada, possam ser aplicadas em outras áreas do conhecimento. Isso, porque com o avanço da comunicação e das tecnologias no setor financeiro, foi criado um mundo altamente conectado que oferece uma vasta gama de informações facilmente acessadas por meio de contas anônimas, o que se tornou um incentivo para que fraudadores manipulassem os diversos sistemas financeiros, ocultando suas atividades maliciosas entre as grandes quantidades de dados (Hooi et al., 2017, p. 1; Pourhabibi et al., 2020, p. 1). Desse modo, identificar as abordagens utilizadas separadas por áreas específicas podem servir de base e inspiração para aplicação em novos sistemas e, subsequentemente, evitar a disseminação de atos ilegais, seja nos sistemas financeiros já em vigor, seja naqueles entrando no mercado por meio da transformação digital (Nikkel, 2020, p. 1).

3.3.5 FOCO

A Classificação 5 indica os Focos de abordagem, dentro das áreas de aplicação, escolhidos pelos pesquisadores para desenvolver e estudar as técnicas de identificação de anomalias. Essa classificação visa afunilar ainda mais o entendimento acerca das abordagens de detecção de anomalias escolhidas para focos de pesquisa e áreas de aplicação específicos, de modo a atuar como suporte para que futuros pesquisadores consigam identificar mais facilmente como as diferentes técnicas de identificação de anomalias são escolhidas e se correlacionam com as diferentes áreas financeiras com base em suas características e necessidades. A escala de codificação abrange as letras A-H. Os artigos que receberam a codificação “5G - Outros” abrangem setores de estudo que não se enquadram nas sub categorizações propostas, como, por exemplo, o estudo proposto por Arboleda et al. (2018, pp. 14–17) que, além de demonstrações financeiras,

abordou operações de venda provenientes de um armazém de dados.

3.3.6 BASE DE DADOS UTILIZADA

A Classificação 6 está relacionada à Base de Dados Utilizada para a construção dos métodos de detecção de anomalias, considerando uma escala que vai de A-D. Ela é relevante para identificar quais artigos utilizaram uma base com dados reais para validar seu modelo de estudos e quais utilizaram dados artificiais ou simulados, uma vez que o uso de dados reais mesmo quando combinados com dados sintéticos produzem uma versão melhor do modelo de detecção (Shtanko e Kulik, 2021, p. 712). Além disso, os mercados financeiros geram grandes volumes de dados e, portanto, a análise dos dados utilizados pode revelar informações úteis e valiosas para o processo de tomada de decisão (H. Wang e Weigend, 2004, p. 457).

3.3.7 DISPONIBILIDADE

A Classificação 7, Disponibilidade, também considera os dados utilizados para as pesquisas, sinalizando se possuem acesso restrito ou irrestrito a partir da escala A-D. Os códigos “6C - Outros” da classificação 6 e “7C - Outros” da classificação 7 são utilizados, respectivamente, quando os autores não especificam se a base de dados é real ou artificial, e não informam a origem dos dados, não sendo possível identificar se tais dados são públicos ou privados. Já, os códigos “6D - Não se aplica” e “7D - Não se aplica”, quando utilizados, sinalizam que o artigo não se encaixa nessas formas de classificação. Isso ocorre, por exemplo, com artigos definidos como revisão de literatura, método adotado pelos autores Al-Hashedi e Magalingam (2021, pp. 5–6), Pourhabibi et al. (2020, pp. 2–3) e Houdek (2020, p. 1), que não possuem uma base com dados de validação de técnicas e métodos de análise.

3.3.8 TÉCNICA UTILIZADA

A Classificação 8 se refere à Técnica Utilizada para detectar as mais diversas formas de anomalias nos sistemas financeiros, com base em uma escala de codificação que vai de A-H. A codificação “8G - Outros” é usada quando os artigos a serem classificados não correspondem às opções de classificação propostas, mas, ainda assim, apresentam técnicas passíveis de identificar anomalias, e a codificação “8H - Não se aplica” é escolhida quando o artigo não é compatível com essa forma de classificação.

3.3.9 NÚMERO DE ESTRATÉGIAS

Considerando a variedade de técnicas existentes para análise na presente revisão de literatura, a Classificação 9 foi estruturada seguindo a escala A-F de modo a verificar o Número de Estratégias utilizadas nos artigos para detectar anomalias, ou seja, a quantidade de técnicas empregadas em um único modelo de estudos. Essa classificação foi escolhida de modo a viabilizar análises acerca da tendência do uso de técnicas combinadas, uma vez que, normalmente, modelos híbridos são mais precisos do que aqueles com processos únicos (K. Chen et al., 2020, p. 3) e a combinação de métodos pode aperfeiçoar o procedimento de predição a ser desenvolvido (Sankarrao, Ghose, e Rathinsamy, 2021, p. 2).

3.3.10 ÁREA DO PERIÓDICO

Finalmente, a Classificação 10, estruturada na escala de codificação A-D, foi feita de modo a identificar as áreas de maior incidência de estudos dentro do tema de detecção de ano-

malias em sistemas financeiros, considerando as Áreas dos Periódicos da *ScienceDirect*[®]. As categorias são: Ciências da saúde (medicina e odontologia; profissões de enfermagem e saúde; farmacologia, toxicologia e ciências farmacêuticas; e ciência veterinária e medicina veterinária), Ciências da vida (ciências agrícolas e biológicas; bioquímica, genética e biologia molecular; ciência ambiental; imunologia e microbiologia; e neurociência), Ciências físicas e engenharia (engenharia química; química; ciência da computação; ciências da terra e planetárias; energia; engenharia; ciência de materiais; matemática; e física e astronomia), Ciências sociais e humanas (artes e humanidades; negócios, gestão e contabilidade; ciências da decisão; economia, econometria e psicologia das finanças; e ciências sociais). É oportuno destacar que como a área de finanças é um campo de estudo das ciências sociais, espera-se que os artigos selecionados se encaixem majoritariamente na codificação “IOB - Ciências sociais e humanas”.

CAPÍTULO 4

RESULTADOS E DISCUSSÕES

*“Queremos ter certezas e não dúvidas,
resultados e não experiências, mas nem
mesmo percebemos que as certezas só
podem surgir através das dúvidas e os
resultados somente através das
experiências”*
—Carl G. Jung.

Os métodos de IA em conjunto com métodos estatísticos e de análise de dados aplicados em *Big Data* e conjuntos de dados de sistemas financeiros originam diversas abordagens de como lidar com informações advindas da área de finanças, considerando, muitas vezes, situações em tempo real e uma complexidade de dados que a habilidade de análise humana sozinha não é capaz de atender. Isso se estende ainda mais no contexto de constantes transformações digitais e, conseqüentemente, aumento da geração de dados dentro da 4^a Revolução Industrial. Essa situação cria, portanto, uma série de desafios para as empresas de modo que elas precisam se organizar para compreender seus dados produzidos para identificar padrões e prever determinadas circunstâncias, o que pode ser feito por meio da criação de modelos computacionais que quando aplicados podem auxiliar na tomada de decisões. Esse ciclo de análise, modelagem e produção de dados pode ser visto na Figura 4.1.

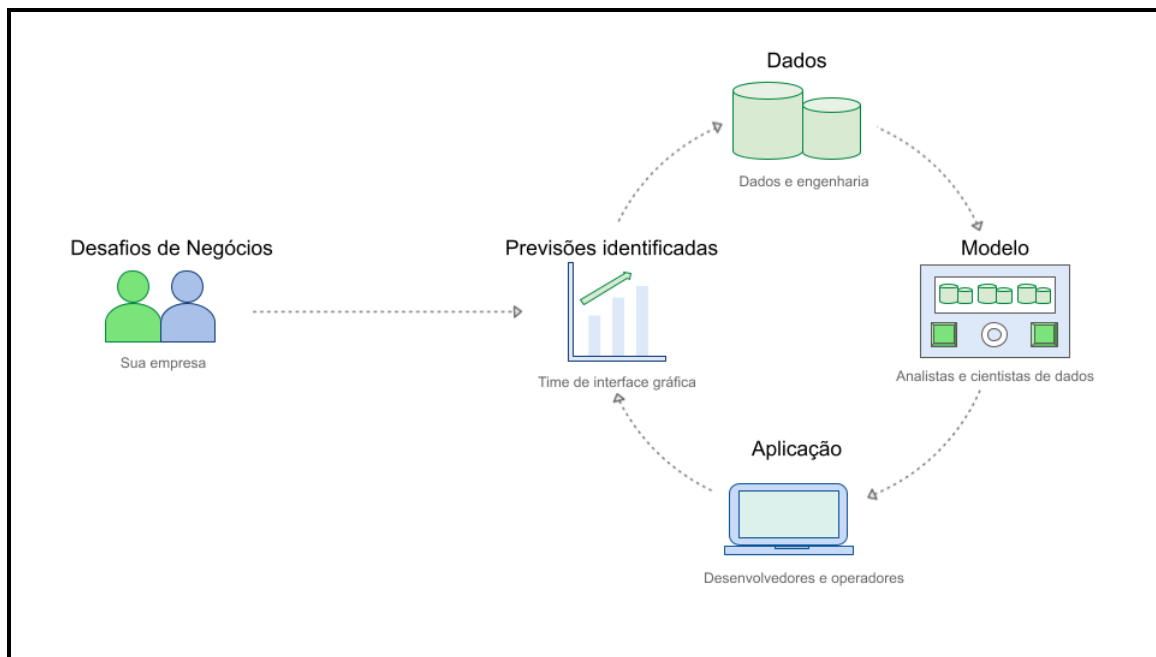


Figura 4.1: Ciclo de implementação de modelos de análise de dados.

Notas: Adaptado de Services (2022, p. 1)

No entanto, existem muitos cenários nos quais podem ser aplicadas várias técnicas para construir um modelo de detecção de anomalias. Por isso, é importante entender quais pesquisas estão em voga e quais algoritmos e técnicas são mais pertinentes a cada caso. Assim, os gestores serão capazes de identificar como tomar suas decisões de modo mais estratégico a partir de uma análise de dados mais precisa, evitando custos desnecessários e gerando mais valor aos seus clientes, uma vez que seus dados, bem como os dados produzidos por eles estarão mais seguros.

A partir do *framework* de classificação proposto, os 74 artigos selecionados foram revisados, analisados e categorizados segundo suas especificações, como apresentado na Tabela 4.1. Essa classificação procura viabilizar o entendimento e análise de modo inteligível acerca dos conhecimentos e técnicas desenvolvidos nos *papers* selecionados. Com base nisso, a presente revisão de literatura do tema de detecção de anomalias em sistemas financeiros pode delinear as tendências de estudo e identificar as oportunidades de pesquisas futuras conforme um sistema de classificação que propicia o reconhecimento de pontos de paridade e de diferença entre as pesquisas.

Artigos Seleccionados	1	2	3	4	5	6	7	8	9	10
Al-Hashedi e Magalingam (2021).	1D	2H	3C	4G	5H	6D	7D	8H	9F	10C
Andor e Bohák (2017).	1A	2A	3B	4C	5F	6A	7B	8B	9A	10D
Arboleda et al. (2018).	1B	2A	3G	4E	5E/5G	6A	7A	8D	9A	10C
Ardizzi et al. (2018).	1A	2B	3B/3D	4A	5D	6A	7A/7B	8B	9A	10D
Arévalo et al. (2022).	1B	2A	3B	4A	5D	6A	7A	8A	9A	10D
Aslam et al. (2022).	1A	2A	3B	4B	5B	6A	7B	8A/8B/8C	9C	10D
Ausloos et al. (2021).	1A	2A	3B	4C	5F	6A	7B	8B/8G	9B	10D
Azevedo et al. (2021).	1B	2A	3B/3G	4D	5C/5E	6A	7B	8B/8D	9B	10C
Badal-Valero et al. (2018).	1A	2B	3B	4A	5D	6A	7C	8A/8D	9B	10C
Baestens et al. (2021).	1C	2B	3B	4A	5A	6A	7A	8A	9A	10C
Baker et al. (2017).	1A	2B	3G	4A/4E	5F	6A	7B	8H	9A	10D
Belle et al. (2022).	1A	2A/2B	3B	4A	5A	6A	7A	8A/8C	9B	10C
Byers et al. (2021).	1C	2G	3B	4C	5F	6A	7B	8A	9A	10C
Canhoto (2021).	1A	2B	3G	4A	5D	6A	7A	8A/8G	9B	10D
Carmona et al. (2019).	1A	2A	3A	4A	5E	6A	7B	8A	9A	10D
Carneiro et al. (2017).	1C	2G	3B/3G	4A	5A	6A	7A	8A	9C	10C
Carta et al. (2019).	1C	2B	3B/3D	4A	5A	6A	7B	8A/8B	9B	10C
Charakopoulos e Karakasidis (2022).	1C	2G	3B	4C	5F	6A/6B	7B	8E	9A	10C
K. Chen et al. (2020).	1C	2G	3B/3D	4A	5A	6C	7B	8A/8B	9B	10C
L. Chen et al. (2022).	1B	2C	3B	4A	5D	6A	7A	8B	9A	10D
Chu et al. (2022).	1C	2G	3B	4F	5G	6A	7A	8B	9A	10D
Climent et al. (2019).	1A	2B	3B	4A/4E	5E	6A	7B	8A	9A	10D
Dinh et al. (2021).	1B	2C	3B	4A/4C	5E/5F	6A	7A	8G	9A	10D
Druicá et al. (2018).	1A	2B	3B	4A	5D/5E	6A	7B	8B/8D	9B	10D
Forough e Momtazi (2021).	1B/1C	2A/2B	3B/3D	4A	5A	6A	7B	8A	9B	10C
Gianini et al. (2020).	1C	2G	3B	4A	5A	6A	7A	8A/8G	9B	10C

Continua na próxima página.

Artigos Seleccionados	1	2	3	4	5	6	7	8	9	10
González-Sánchez (2021).	1A/1B	2F	3B	4C	5F	6A	7B	8B	9A	10D
Gu et al. (2022).	1D	2H	3B	4F	5D/5G	6A	7B	8A	9A	10D
Herrera et al. (2022).	1A	2B	3G	4A/4C	5D	6A	7B	8A	9A	10D
Houdek (2020).	1D	2H	3C/3H	4G	5H	6D	7D	8H	9F	10D
James et al. (2021).	1C	2G	3B	4C	5G	6A	7B	8B	9A	10C
Jurgovsky et al. (2018).	1C	2G	3B	4A	5A	6A	7C	8A	9A	10C
Kou et al. (2021).	1B	2C	3B	4A	5D/5E	6A	7A	8A	9A	10C
Kristóf e Virág (2022).	1A	2B	3B	4A	5D	6A	7B	8A	9A	10D
León (2020).	1B	2A	3B	4A	5D	6A/6B	7A	8A/8G	9B	10D
León et al. (2020).	1B	2A	3B	4A	5D	6A	7A	8A	9A	10D
Lin et al. (2018).	1B	2C	3B	4E	5E	6A	7B	8B/8D	9B	10D
Q. Liu et al. (2021).	1B	2C	3B	4E	5E	6A	7B	8A	9A	10D
Lucas et al. (2020).	1A	2B	3B	4A	5A	6A	7A	8A	9A	10C
Mamede e Malaquias (2017).	1B	2A	3B	4C	5F	6A	7B	8B	9A	10D
Misra et al. (2020).	1C	2B	3B/3D	4A	5A	6A	7B	8A	9A	10C
Mustafa et al. (2022).	1A	2A	3B	4C	5F	6A	7B	8A	9A	10D
Nikkel (2020).	1A	2B	3A	4A	5A/5D	6A	7B	8G	9A	10C
Nissim (2022).	1D	2H	3C	4G	5H	6D	7D	8H	9F	10D
Papík e Papíková (2022).	1A	2A	3B	4E	5E	6A	7A	8A	9A	10D
Parnes (2020).	1A	2A	3A/3B	4C	5F	6A	7B	8A/8B/8E	9A	10D
Patil et al. (2018).	1A	2B	3B	4A	5A	6A	7C	8A/8F	9B	10C
Pourhabibi et al. (2020).	1D	2H	3C	4G	5H	6D	7D	8H	9F	10C
R.B. e K.R. (2021).	1C	2B	3A	4A	5A	6A	7B	8A	9A	10A
Rocha-Salazar et al. (2021).	1B	2A	3B	4A	5D	6A	7A	8A	9A	10C

Continua na próxima página.

Após a classificação, os dados categorizados obtidos foram compilados estatisticamente, por meio do uso de um *script*¹ desenvolvido na linguagem de programação *Python*, versão 3.7, mediante a *Anaconda Distribution*, da *Anaconda Inc*[®]. De maneira mais precisa, esse *script* automatizou os cálculos das estatísticas descritivas para cada categoria. Além disso, foram construídos gráficos, a partir de tais dados, que evidenciaram análises e discussões dos resultados também de modo visual. Os gráficos foram divididos em dois para cada categoria, sinalizando os dados estatísticos referentes à contagem de códigos isoladamente e aos agrupamentos de códigos que apareceram no momento da classificação de cada artigo, com exceção das classificações 9 e 10, em que não houve mais de um código indicado para um único artigo e, portanto, os dados estatísticos mostram apenas a contagem de códigos isolados.

4.1 CONTEXTO ECONÔMICO

Essa categoria de classificação foi estruturada de modo a relacionar os artigos selecionados de acordo com suas características econômicas. Os países que possuem economias maduras, normalmente, possuem sistemas mais estruturados, com leis mais rígidas e maior pressão coercitiva para promover regulação e fiscalização (Raj, Agrahari, e Srivastava, 2020, pp. 2–3). No entanto, os países com economia não madura, por mais que tenham diferentes infraestrutura físicas, financeiras e legais, estão cada vez mais investindo em tecnologia da informação (Arnott, Jirachiefpattana, e O'Donnell, 2007, p. 2078) e criando sistemas de suporte à decisão que podem auxiliar tanto os setores públicos quanto os privados em termos econômicos (Chaudhry, Li, Xu, e Zhang, 2007, p. 1987). Desse modo, identificar o contexto econômico em que os modelos de detecção de anomalias são aplicados nos artigos é de suma importância não só para verificar as diferenças estruturais dos países em questão, mas como também identificar as semelhanças e os possíveis avanços tecnológicos no que diz respeito à identificação de fraudes, desvios de capital ou adulteração em sistemas financeiros.

Os dados estatísticos referentes às classificações dos artigos podem ser observados por meio da Figura 4.2. A codificação mais utilizada foi “1A - Economia madura”, contemplando 38% dos artigos, o que vai de encontro com a pesquisa de Raj et al. (2020, p. 2) sobre os avanços de economias mais maduras e, devido a isso, o número de pesquisas nesses países acaba sendo maior. No entanto, é necessário observar que a porcentagem de classificação da codificação “1B - Economia não madura” também é significativa, abrangendo 29% dos artigos. Desse modo, é possível verificar que a busca por modelos e técnicas de detecção de anomalias eficientes não se limita apenas a países de economia madura e, sim, configura uma semelhança entre as duas formas de economia. Porém, para ter-se análises comparativas nesse nível, seria necessário que pesquisas futuras combinassem um conjunto de dados testes que envolvesse tanto países de economia madura quanto países de economia não madura, como feito em apenas um artigo dentre os selecionados (González-Sánchez, 2021, p. 7). Tem-se, então, a primeira lacuna identificada nesta pesquisa, a saber:

- L₁ Desenvolver pesquisas que considerem países de economia madura e não madura para identificar semelhanças e divergências em suas abordagens de detecção de anomalias e administração de sistemas de decisão.

Além disso, uma análise significativa que pode ser aprofundada é se há a existência de uma correlação entre o contexto econômico da pesquisa e o comportamento dos indivíduos, ou seja, se há fatores econômicos que propiciam o aumento de ações criminosas que sinalizem

¹Nota do autor: Sequência de códigos.

Artigos Seleccionados	1	2	3	4	5	6	7	8	9	10
Rtayli e Enneya (2020).	1C	2B/2G	3B/3D	4A	5A	6A/6B	7A/7B	8A/8G	9C	10C
Rubio et al. (2020).	1B	2A	3B	4A	5D	6A	7C	8A	9A	10D
Sá et al. (2018).	1B	2A	3B/3G	4A	5A	6A	7A	8A/8B	9B	10C
Sabetti e Heijmans (2021).	1A	2A	3B	4E	5D	6A/6B	7D	8A	9A	10D
Serneels (2022).	1D	2H	3A	4F	5G	6A	7B	8G	9C	10D
Severino e Peng (2021).	1B	2A	3B	4B	5B	6A	7A	8A	9A	10C
Shi et al. (2018).	1B	2A/2B/2C	3B	4E	5E	6A	7B	8B/8D	9B	10C
K. Singh e Best (2019).	1C	2G	3A/3G	4A	5D	6A	7A	8E/8F	9B	10D
Stripling et al. (2018).	1C	2B	3B	4B	5B	6A	7A	8A	9A	10C
Subudhi e Panigrahi (2020).	1A	2A	3B	4B	5B	6A	7B	8A	9B	10C
Tardelli et al. (2021).	1A	2A	3B/3G	4C	5F	6A	7B	8A	9B	10C
Vandervorst et al. (2022).	1A	2B	3B	4B	5B	6A	7A	8B	9A	10C
Y. Wang e Xu (2018).	1B	2C	3B	4B	5B	6A	7A	8A/8G	9B	10C
Wei et al. (2021).	1B	2C	3A/3B	4F	5G	6A	7A	8B	9A	10D
Wever et al. (2022).	1A	2C	3A	4F	5G	6A	7B	8A/8G	9B	10D
Yan et al. (2020).	1C	2G	3B/3D	4B	5B	6A/6B	7A/7B	8A/8B/8F	9D	10B
Yankol-Schalck (2022).	1A	2B	3B	4B	5B	6A	7A	8A	9A	10D
Yoon et al. (2021).	1C	2G	3A	4E	5E	6A	7A	8G	9A	10D
Yuan et al. (2022).	1B	2C	3B	4E	5E	6A	7B	8A	9B	10D
Zhai et al. (2017).	1A	2A	3B	4C	5F	6A	7B	8A/8B	9B	10C
W. Zhang et al. (2022).	1C	2G	3B	4F	5G	6A	7B	8A/8B	9B	10C
G. Zhang et al. (2022).	1A	2E	3A	4E	5D	6A	7A	8A/8G	9B	10D
Y. Zhang et al. (2022).	1B	2C	3B	4E	5E	6A	7B	8A/8G	9B	10D
Zoričák et al. (2020).	1A	2B	3B	4A/4E	5E	6A	7B	8A/8B	9C	10D

Tabela 4.1: Artigos categorizados com base no *framework* de classificação e codificação.

anomalias financeiras e, conseqüentemente, ocasionam o aumento da pesquisa por modelos que possam identificar esses *outliers* e, assim, viabilizar uma maior segurança para os sistemas financeiros e para os indivíduos que deles usufruem.

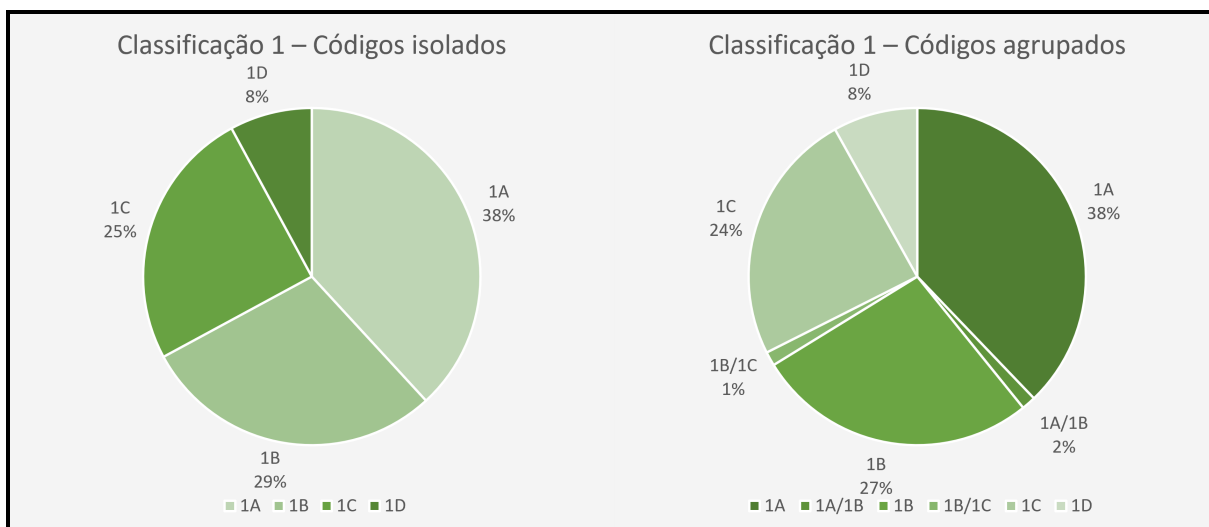


Figura 4.2: Contexto econômico.

Notas: 1A - Economia madura; 1B - Economia não madura; 1C - Não informado; e 1D - Não se aplica.

Vale destacar que uma lacuna observada é o número expressivo de artigos que não indicam a origem dos dados utilizados para aplicar e testar a funcionalidade de seus modelos de detecção, sendo 25% de artigos classificados na categoria “1C - Não informado”, que abrange os estudos os quais não explicitam a localização dos dados e, portanto, não possibilitam uma análise acerca do contexto econômico de onde o desenvolvimento da pesquisa está inserido. Isso configura, portanto, a segunda lacuna oportuna para pesquisas futuras, a saber:

L₂ Sinalizar a origem dos dados utilizados para pesquisa, de modo que seja possível relacioná-los com os seus respectivos contextos econômicos.

4.2 LOCALIZAÇÃO GEOGRÁFICA DE PESQUISA

Ao produzir um estudo acerca de um tema, a análise de como os fatores de localização geográfica nos quais a pesquisa está inserida é fundamental para a aquisição de informações que podem ser úteis para diferentes formas de investigação (Lozano, Schreiber, e Brynielsson, 2017, p. 18) e, conseqüentemente, tal contato mais aprofundado com as informações referentes à posição geográfica de onde o estudo está sendo conduzido pode ser de extrema importância para a compreensão das diferentes formas de manifestação de um tema de pesquisa ao redor do mundo. Isso ocorre, uma vez que análises acerca da localização possibilitam o entendimento acerca de possíveis padrões comportamentais, arranjos espaciais, agrupamentos, além de prováveis relacionamentos que podem decorrer de fenômenos geográficos (Pick, Turetken, Deokar, e Sarkar, 2017, p. 1).

Verificar e ponderar as questões acerca da localização geográfica das pesquisas desenvolvidas, nesse contexto, pode viabilizar a atribuição de informações sobre os possíveis objetos de interesse, que podem mostrar-se como importantes características a serem consideradas por um tomador de decisões (Hess, Rubin, e West, 2004, p. 198), visto que tais particularidades

podem influenciar na interpretação das situações em estudo. Com base nisso, os artigos selecionados envolvendo métodos de detecção de anomalias em sistemas financeiros diversos e as fraudes e ações criminosas decorrentes de tais anormalidades podem ser analisados conforme as perspectivas da posição geográfica em que eles foram produzidos. É oportuno destacar que a perspectiva utilizada para análise pode ser um indicativo de diferentes formas de desvios do padrão, como, por exemplo, no estudo desenvolvido por Jurgovsky et al. (2018, p. 236) em que as transações fraudulentas podem ser analisadas como anomalias no comportamento do consumidor ou, então, como um conjunto de transações que divergem da classe de transações genuínas. Desse modo, o ponto de vista quando combinado com a análise da localização geográfica das pesquisas produzidas sobre anomalias sinaliza a necessidade de identificar correlações entre localidades e os índices de ações criminosas, ou, então, entre culturas e o modo de posicionar-se acerca de questões como segurança de sistemas ou punição e justiça.

Os dados estatísticos provenientes da classificação dos artigos selecionados foram apresentados na Figura 4.3. A maioria dos estudos foram classificados como “2A - América” e “2B - Europa”, de modo que esses dois continentes desenvolvem mais pesquisas sobre como identificar anomalias em seus sistemas financeiros, abrangendo 59% dos artigos revisados. Inclusive, dentre esses artigos, há dois que desenvolvem pesquisas considerando países de ambos os continentes (Forough e Momtazi, 2021, pp. 6–7; Shi et al., 2018, p. 880). Considerando isso, cabe aqui a discussão de quais seriam as possíveis causas desse número expressivo de artigos sobre anomalias financeiras em países Americanos e Europeus, de modo a promover uma ilustração das circunstâncias que possivelmente propiciam ou justificam tais dados estatísticos, como os anteriormente abordados padrões culturais e, subsequentemente, posicionamentos acerca da situação.

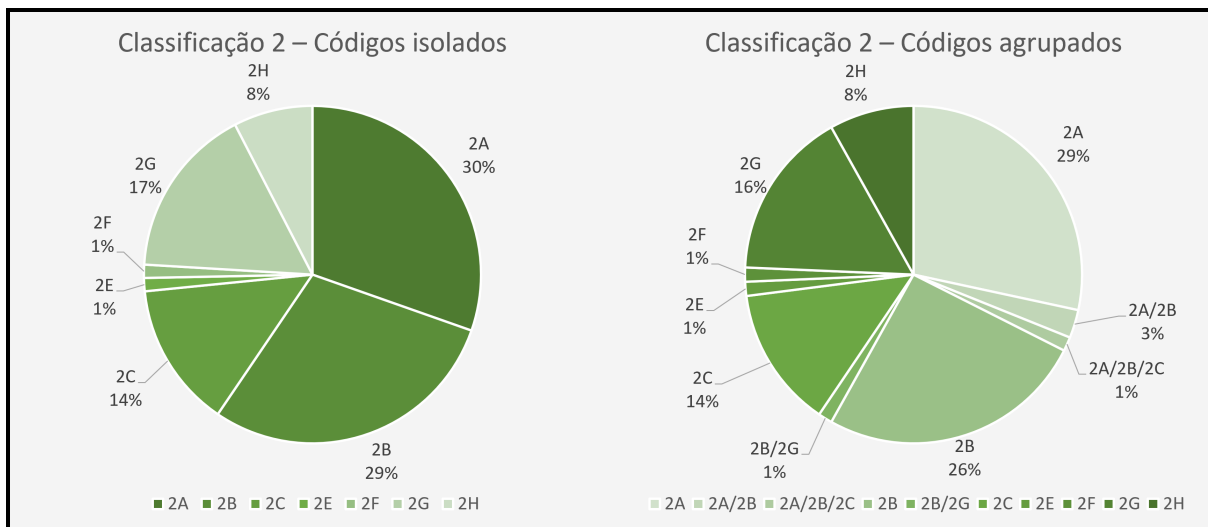


Figura 4.3: Localização geográfica de pesquisa.

Notas: 2A - América; 2B - Europa; 2C - Ásia; 2D - África; 2E - Oceania; 2F - Todos os continentes; 2G - Outros; e 2H - Não se aplica.

As dimensões culturais são fortemente utilizadas para encontrar as razões para a ocorrência de variações de resultados entre países, uma vez que indivíduos que possuem valores, crenças e preferências divergentes tendem a agir de maneira diferente (Tung e Quaddus, 2002, p. 190). Além disso, culturas distintas despertam perspectivas dissimilares que, a partir dos valores, crenças e normas, propiciam diferentes comportamentos humanos, e geram, portanto, atitudes e práticas particulares à cada cultura (Kankanhalli, Tan, Wei, e Holmes, 2004, pp. 183–184). Do mesmo modo, as divergentes culturas podem explicar diferenças significativas em termos de desenvolvimento do setor financeiro e das questões a ele relacionadas M. A. Khan, Gu, Khan,

e Meyer (2021, p. 8). Esses efeitos culturais podem ser observados, por exemplo, nos diferentes discursos utilizados pelos autores para abordar a existência de anomalias nos sistemas financeiros considerando América e Europa.

Quando se trata de América, Azevedo et al. (2021, p. 2) abordou a corrupção e o desvio de dinheiro como um problema tradicional e expressivo no país dos dados da pesquisa, de modo que nem programas sociais estão livres de ações criminosas, e Parnes (2020, pp. 292–293) afirmou que o estudo das anomalias em conjunto com análises de indicadores macroeconômicos podem auxiliar não só no entendimento da temática de pesquisa, mas como também na compreensão de eventos passados. Com relação a Europa, Baker et al. (2017, p. 20) discorreu sobre a necessidade de examinar as falhas e desvios que provocaram a falência de um banco internacionalmente conhecido por suas expertises e lucratividade alta para poder gerar direcionamentos a pesquisas futuras, e Ardizzi et al. (2018, p. 106) abordou a importância do uso de ferramentas de detecção de anomalias em conjunto com as premissas de uma unidade de inteligência estratégica Italiana. Desse modo, o discurso de pesquisas feitas em países da América reforça que tais problemas estão instaurados nos países e são constantes, enquanto o discurso de países Europeus constrói argumentos que reforçam as estratégias e expertises locais, sendo menos destrutivos com relação à existência de anomalias nos sistemas financeiros. Tais abordagens podem ser derivadas de efeitos culturais, assim como a frequência de pesquisas entre os continentes.

Ainda com relação aos efeitos culturais, pode-se verificar que as menores categorias de classificação, em termos estatísticos, foram “2E - Oceania” e “2F - Todos os continentes”, com apenas um artigo classificado em cada (González-Sánchez, 2021, p. 7; G. Zhang et al., 2022, p. 2). Isso significa que não há muitos artigos que desenvolvam estudos abrangentes, considerando uma maior diversidade de países e regiões ao redor do mundo. A base de dados utilizada nos artigos normalmente inclui contextos nacionais isolados em que são feitas análises sobre a efetividade de métodos de detecção de anomalias em sistemas financeiros específicos de um local, sem que haja nenhuma forma de comparação geográfica, como nos estudos desenvolvidos por Lin et al. (2018, p. 159), Druicã et al. (2018, p. 77) e Mamede e Malaquias (2017, pp. 47–48). Além disso, não houve nenhum artigo classificado como “2D - África”, o que apenas ressalta a escassez de pesquisas sobre anomalias não só de modo comparativo, mas como também pesquisas isoladas em todos os continentes. Países na África, por exemplo, possuem uma vasta população que não possui acesso a serviços financeiros, mas com os avanços tecnológicos no campo dos telefones celulares essa situação pode mudar (Zaffar et al., 2019, p. 62) e, portanto, estudar correlações entre o comportamento populacional e a detecção de anomalias nos sistemas financeiros pode indicar diferentes percepções acerca desta temática. Desse modo, é possível identificar a terceira e a quarta lacuna de pesquisa, a saber:

L₃ A falta de artigos que usem dados da África e Oceania, ou, ainda, que considerem todos os continentes em uma só pesquisa.

L₄ Caracterizar o comportamento humano em diferentes culturas em consonância com a existência de anomalias em sistemas financeiros.

É oportuno destacar que, entender a relação entre a localização geográfica e os efeitos culturais decorrentes dela pode auxiliar em uma compreensão mais aprofundada do por quê América e Europa obtiveram mais estudos dentro da temática em questão, como, também, identificar as causas para não haver tantos estudos em outros continentes. Essas lacunas passam a ser uma importante oportunidade para pesquisas futuras, de modo a responder se há a existência de uma correlação de causa e efeito entre culturas diversas e a adoção de práticas criminosas que resultem em anomalias, ou, então, entre as culturas e a forma de prevenir e abordar tais ocorrências.

Cada contexto e pergunta de pesquisa relaciona-se com métodos diferentes e apropriados de análise (Lo, Rey-Martí, e Botella-Carrubi, 2020, p. 221). Uma análise do método de pesquisa dos artigos desenvolvidos no campo de detecção de anomalias financeiras possibilita a identificação de métodos comuns utilizados nessa temática, como, também, aqueles que não são muito explorados e, assim, configuram uma lacuna de pesquisa que poderia ser estudada e desenvolvida em pesquisas futuras. Na Figura 4.4 são apresentados os dados estatísticos de classificação e codificação dos artigos selecionados, de modo a ser possível visualizar os diferentes métodos de pesquisa escolhidos para essas pesquisas.

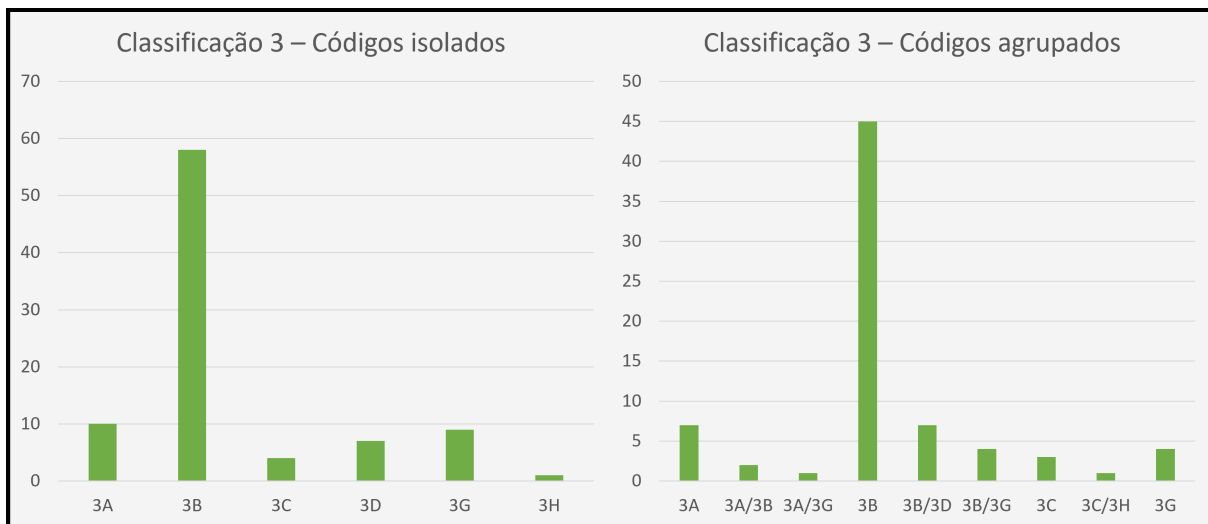


Figura 4.4: Método de pesquisa.

Notas: 3A - Qualitativo; 3B - Quantitativo; 3C - Conceitual; e 3D - Modelagem; 3G - Estudo de caso único; e 3H - Múltiplos estudos de caso.

A codificação mais adotada na classificação foi a “3B - Quantitativo”, de modo que os artigos desenvolvem métodos de pesquisa que envolvam algoritmos e abordagens numéricas para explorar quantidades de dados. Esse resultado é consonante com os estudos de Bose, Piramuthu, e Shaw (2011, p. 557), que abordou como o desenvolvimento da tecnologia de informação e das ciências de decisão propiciam o uso de modelos quantitativos de pesquisa, uma vez que eles viabilizam a automatização e diminuição de operações manuais durante o processo de detecção de fraudes, por exemplo. Desse modo, o método quantitativo de pesquisa mostra-se como o mais relevante nessa temática, seja ele utilizado de forma isolada, ou de modo combinado com outros métodos, como os estudos de (Forough e Momtazi, 2021, p. 4; K. Chen et al., 2020, p. 3; Misra et al., 2020, pp. 258–259; Rtayli e Enneya, 2020, p. 3; Yan et al., 2020, p. 118; Carta et al., 2019, p. 16; Ardizzi et al., 2018, pp. 108–109) que, além de receber a codificação 3B, foram codificados também como “3D - Modelagem”, ao desenvolver um novo modelo de detecção a partir de técnicas quantitativas.

No entanto, por mais que o método quantitativo seja predominante, pode-se observar a partir dos resultados obtidos, que faltam artigos conceituais para a área de conhecimento, de modo que seja possível padronizar práticas comuns. Dentre os artigos selecionados, aqueles categorizados como “3C - Conceitual” foram revisões de literatura (Al-Hashedi e Magalingam, 2021, pp. 5–6; Houdek, 2020, p. 1; Pourhabibi et al., 2020, pp. 2–3), já que esse método de pesquisa viabiliza a abstração de aspectos relevantes de um domínio, além de identificar requisitos e objetivos de pesquisa (Maass e Storey, 2021, p. 2), assim como pesquisas conceituais. Porém, os resultados indicaram que, além desses três artigos de revisão de literatura, não há artigos que

utilizem modelos conceituais. Esse modo de pesquisa é criticamente importante para a academia, além de se mostrar como uma necessidade para os praticantes, uma vez que define métodos, ferramentas e teorias relevantes para a construção e desenvolvimento de pesquisas ao criar um sistema de implementação de práticas globais que garante aplicações de qualidade (Trujillo et al., 2021, p. 1). Desse modo, isso caracteriza a quinta lacuna de pesquisa, a saber:

L₅ A falta de estudos conceituais na literatura que facilitem a comparação entre artigos dentro de uma mesma temática e abordagem de pesquisa.

Essa lacuna, no entanto, não é isolada, e sinaliza a existência da sexta lacuna de pesquisa relacionado a essa temática, a saber:

L₆ A falta de um indicador específico para verificar a efetividade dos modelos e técnicas desenvolvidos para detecção de anomalias financeiras, como também, de uma base de dados pública para testes e mensuração padronizados.

Se houvesse a construção de referências teóricas, de um indicador e de uma base de dados pública com conjuntos de dados padronizados de diferentes tamanhos, seria possível que os autores aplicassem testes em seus modelos de detecção de modo a viabilizar uma comparação concreta da eficiência e precisão de identificação de anomalias. Isso possibilitaria, portanto, a mensuração e comparação de soluções acadêmicas dentro de um padrão de pesquisa mais plausível e real.

4.4 ÁREA DE APLICAÇÃO

Essa categoria de classificação foi criada de modo a identificar as áreas financeiras em que as fraudes, desvios de capital e adulteração de dados são realizados, de modo a compreender melhor quais setores mais são afetados e quais são menos afetados de acordo com a quantidade de pesquisas realizadas nessas áreas. Os dados estatísticos referentes a essa classificação podem ser visualizados por meio da Figura 4.5.

A codificação mais recorrente dentre os artigos selecionados foi a “4A - Banco”, englobando 34 artigos que representam 43,03% do total, sendo assim mais da metade. Para Patil et al. (2018, p. 385), a área bancária é um setor muito importante na contemporaneidade, uma vez que quase todos os indivíduos precisam lidar com questões bancárias, seja física ou virtualmente, porém essa virtualização dos sistemas tornou o banco muito vulnerável a ataques diversos. Nikkel (2020, pp. 1–2) afirmou que esses novos sistemas financeiros facilitam a transferência de dinheiro entre indivíduos, pagamentos de serviços e substituem a necessidade de dinheiro físico e, por causa disso, criminosos abusam dessas tecnologias financeiras para promover fraudes, extorsões e lavagem de dinheiro. Isso justifica a grande necessidade de pesquisas por métodos e técnicas de detecção de anomalias nessa área financeira, de modo a identificar e mitigar tais atividades criminosas.

É oportuno destacar que Climent et al. (2019, p. 885) em sua pesquisa discorreu sobre como os bancos são uma peça importante para a economia dos países, e crises em seus sistemas podem desestabilizar os mercados financeiros e, ainda, prejudicar o crescimento econômico global. Para Canhoto (2021, p. 441), como eles mantêm registros da identidade e do comportamento de seus clientes, é possível utilizar esse conjunto de dados para detectar e prevenir atividades criminosas. Nesse contexto, Carmona et al. (2019, p. 304) afirmou que essa abordagem de prevenção auxilia no processo de confiança desses clientes com os bancos, mantendo a reputação evitando a falência bancária (Carmona et al., 2019, p. 304). Desse modo, o número expressivo

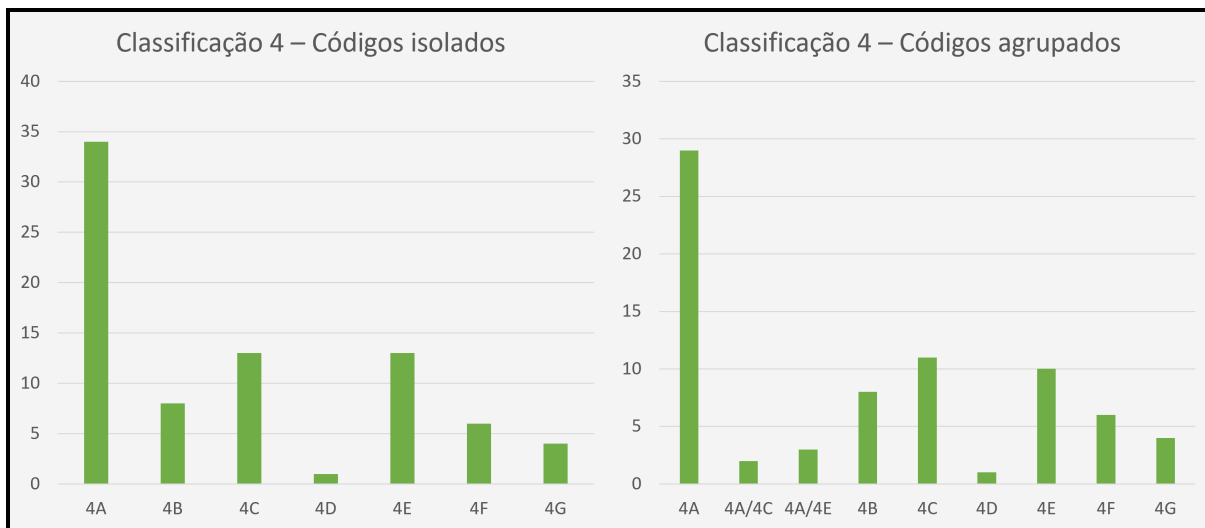


Figura 4.5: Área de aplicação.

Notas: 4A - Banco; 4B - Atuarial; 4C - Mercado financeiro; 4D - Seguro social; 4E - Contabilidade financeira; 4F - Outros; e 4G - Não se aplica.

encontrado de modelos de detecção de anomalias nos sistemas bancários explica uma grande necessidade de segurança nesse setor que, se ameaçado, pode interferir na economia de forma global.

Considerando tais questões de confiança, cabe aqui ressaltar que os resultados significativos acerca de pesquisas na área bancária podem relacionar-se com as dimensões culturais abordadas por Tung e Quaddus (2002, p. 190), causando variações nos resultados conforme o comportamento dos clientes em diferentes localizações. Picoto e Pinto (2021, p. 620) em seus estudos identificou que países que mais utilizam de *mobile banking*², por exemplo, são aqueles que, normalmente, ainda não desenvolveram adequadamente seus sistemas de banco, o que pode influenciar na proteção dos dados e viabilizar ações criminosas como fraudes financeiras. Logo, observa-se a sétima lacuna de pesquisa, a saber:

L₇ Identificar se os efeitos culturais podem influenciar também as áreas de aplicação, levando determinados países a precisarem de sistemas de prevenção mais rígidos do que outros.

Por outro lado, a menor codificação de classificação utilizada foi “4D - Seguro social”, representando apenas 1,26% dos artigos. Para Azevedo et al. (2021, p. 2), corrupções e fraudes nessa área podem prejudicar o recebimento de dinheiro por parte dos beneficiários. Desse modo, para pesquisas futuras dentro dessa temática, tem-se a oitava lacuna, a saber:

L₈ Identificar o por quê de seguro social ser uma área menos explorada e se isso está relacionado com efeitos culturais ou se esse setor apenas representa uma baixa prioridade quando se trata de desenvolver sistemas de detecção de anomalias.

4.5 Foco

A categoria de Foco aprofunda as áreas de aplicação de modo a identificar de maneira mais precisa na qual concentram-se anormalidades financeiras. Os resultados estatísticos podem

²Serviço oferecido por instituições financeiras para que os cliente possam acessar serviços bancários a partir de dispositivos móveis.

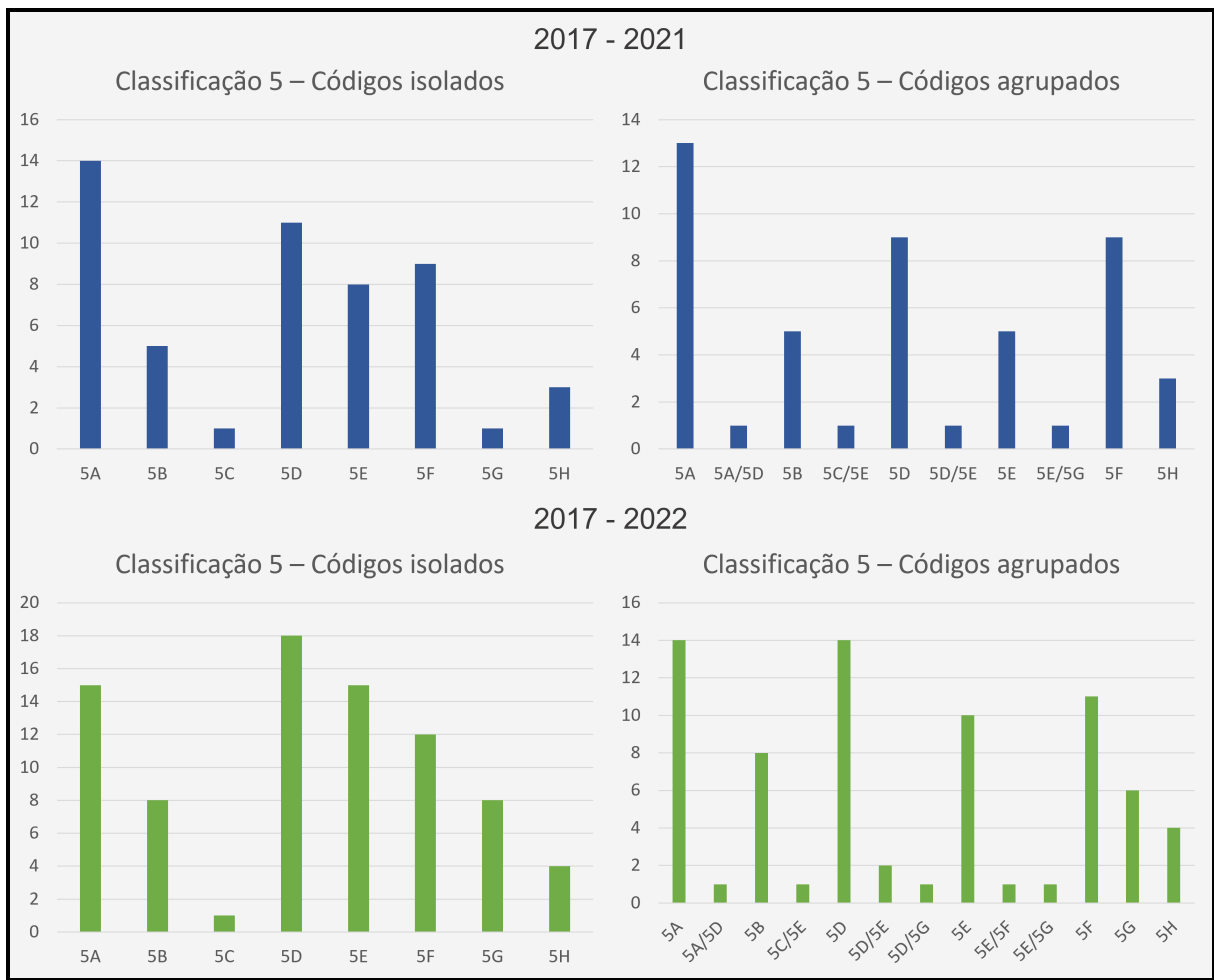


Figura 4.6: Foco.

Notas: 5A - Cartão de crédito; 5B - Seguro; 5C - Benefícios de bem-estar financeiro; 5D - Transações financeiras; 5E - Declarações financeiras; 5F - Ações e *commodities*; 5G - Outros; e 5H - Não se aplica.

ser observados por meio da Figura 4.6, que contém uma comparação entre dados de 2017 até julho de 2021 e os dados após adição de artigos de agosto de 2021 a outubro de 2022 à base de literatura. Os números expressivos de artigos classificados como “5A - Cartão de crédito”, “5D - Transações financeiras” e “5E - Declarações financeiras” justificam o resultado alto encontrado para as classificações de banco na subseção “Área de aplicação”, explicitando ainda mais a necessidade de uma segurança elevada no setor bancário.

A codificação que mais se destacava nessa classificação até julho de 2021 era a “5A - Cartão de crédito”, representando 26,92% do total de artigos. Isso é consonante com os resultados da pesquisa de Rubio et al. (2020, p. 3) de que, quando se trata de contextos financeiros, as aplicações de modelos de detecção de anomalias estão principalmente concentradas em transações fraudulentas de cartão de crédito. No entanto, por mais que a codificação “5A” não seja a mais recorrente após a adição dos novos artigos à base de revisão, devido as porcentagens das codificações “5D” e “5E” que cresceram, sua aparição ainda é expressiva, representando 18, 52% do total. Para Forough e Momtazi (2021, p. 1) e Jurgovsky et al. (2018, p. 234), esse resultado ocorre devido ao crescimento da forma eletrônica de pagamentos que viabiliza as fraudes em cartão de crédito ao criar oportunidades para os fraudadores realizarem as ações criminosas (Misra et al., 2020, p. 254). Em conjunto a isso, o grande volume de transações financeiras *online* tornam inviável a verificação manual por anomalias (Gianini et al., 2020, p. 549), o que gera uma série de contratempos quando se trata do desenvolvimento de métodos de detecção de dados anormais.

Para Forough e Momtazi (2021, p. 9), um dos maiores desafios ao identificar fraudes de cartão de crédito é a necessidade de um modelo que seja capaz de trabalhar com dados em tempo real. Carta et al. (2019, p. 13) apresentaram a necessidade de ter-se essa mesma perspectiva de tempo real, uma vez que, atualmente, a dinâmica comercial permite o consumo de bens e serviços por meio quaisquer aparelhos conectados à *internet*. Lucas et al. (2020, p. 393), por outro lado, afirmaram que o problema encontra-se no fato de que muitos estudos adotam as transações de cartão de crédito como eventos isolados, ao invés de relacioná-las como sequências. Carneiro et al. (2017, p. 91) sinalizaram, ainda, outro desafio: cada vez mais os fraudadores estão manipulando os dados de modo a aparentar estarem de acordo com os dados referentes ao comportamento legítimo do detentor do cartão. Logo, as opiniões divergentes, na verdade, convergem para a grande necessidade de pesquisas nessa temática e explicam o resultado obtido na presente revisão de literatura.

Além disso, uma classificação que vale destacar é a “5B - Seguro”, que foi pouco utilizada, contemplando apenas 9,88% dos artigos totais. Segundo Stripling et al. (2018, p. 13), os fraudadores veem a fraude de seguros como de baixo risco e, ao mesmo tempo, como uma forma de receber altos valores monetários, sendo considerada por eles como um crime mais “seguro” de se cometer. Ainda, Severino e Peng (2021, p. 1) afirmaram que o mercado de seguros é muito lucrativo e movimenta uma grande quantidade de dinheiro ao ano. Para Y. Wang e Xu (2018, p. 87), o desenvolvimento otimista desse mercado possibilitará a entrada de ainda mais capital nessa forma de indústria. Desse modo, a falta de um número expressivo de pesquisas com esse enfoque sinaliza a nona lacuna, a saber:

L₉ Por que as anomalias traduzidas em fraudes de seguro não recebem tanta atenção quando se trata do estudo de anomalias financeiras?

4.6 BASE DE DADOS UTILIZADA

A análise acerca do tipo de base de dados utilizada para promover testes de eficiência dos modelos de detecção de anomalias desenvolvidos nos artigos visa a compreensão de quais seriam os melhores dados para mensuração: os reais ou os simulados artificialmente. Os resultados estatísticos desta categoria são apresentados na Figura 4.7.

A codificação mais utilizada foi a “6A - Real”, contemplando 88% dos artigos selecionados. Esse resultado reforça o resultado obtido por Pourhabibi et al. (2020, p. 8) de que os estudos de pesquisa acerca de modelos de detecção de fraude, em sua maioria, usam dados do mundo real como sua plataforma de teste, enquanto os dados sintéticos servem apenas para simular cenários e circunstâncias mais específicas. Para Severino e Peng (2021, p. 1), dados reais configuram uma vantagem expressiva quando comparados com dados simulados, tanto em termos de avaliação quanto em questões de aplicabilidade prática no processo de tomada de decisão. Os eventos que acontecem no mundo real não podem ser considerados monótonos (Andor e Bohák, 2017, p. 47), porque sofrem variações que causam um alto grau de desequilíbrio de dados (Carta et al., 2019, p. 21) e, por isso, normalmente, precisam de muitos ajustes e tratamentos (Badal-Valero et al., 2018, p. 28). Isso é necessário uma vez que os custos da não identificação de uma anomalia ou fraude são muito altos quando se trata de um conjunto de dados reais (Sá et al., 2018, p. 25). No entanto, por mais que haja diversos desafios, os resultados indicam que o uso de dados reais ainda é o mais utilizado para mensurar a efetividade dos modelos e técnicas de detecção de anomalias financeiras.

Porém, é oportuno ressaltar que a quantidade de dados utilizados nos testes varia de artigo para artigo, não há uma quantidade padronizada que sirva de base para comparação e, desse modo, não há como saber se uma técnica que é muito muito efetiva em um banco de dados pe-

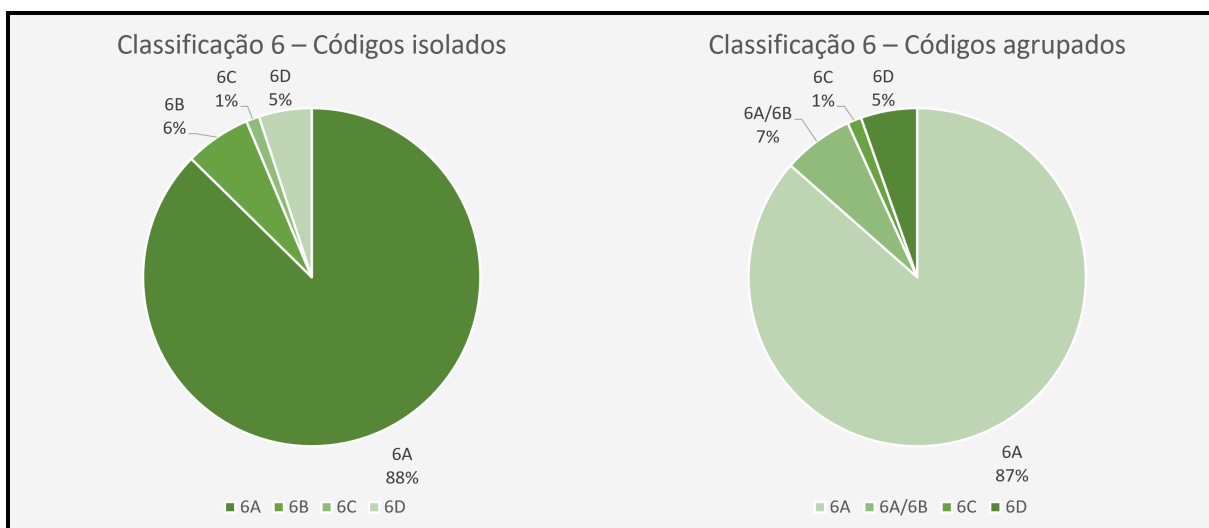


Figura 4.7: Base de dados utilizada.

Notas: 6A - Real; 6B - Sintética/Artificial; 6C - Outros; e 6D - Não se aplica.

queno seria viável para o processamento necessário de grandes bancos de dados. Quando se trata de processar e analisar um grande volume de dados que, normalmente, não possuem natureza e tamanho estruturados, como, por exemplo, o chamado *Big Data*, não é possível utilizar métodos tradicionais (Trujillo et al., 2021, p. 2). Os algoritmos existentes para a detecção de anomalias são limitados por problemas de eficiência e dimensionalidade, principalmente ao processar bancos de dados de sequência de alta dimensão (H. Liu, Lin, He, e Cai, 2010, p. 270). Isso se agrava ainda mais quando se trata do processamento de dados em tempo real, uma vez que é necessária habilidade para interpretar e administrar os 5V's do *Big Data* que são continuamente gerados, de modo que seja possível fornecer um suporte informacional adequado para os gestores (Fertier et al., 2020, p. 1). Cabe aqui sinalizar, então, a décima lacuna de pesquisa, a saber:

L₁₀ Desenvolver modelos de pesquisa que levem em consideração problemas de dimensionalidade e sejam, assim, capazes de trabalhar com dados volumosos e variados em tempo real.

Por mais que muitos sistemas que auxiliam no processo de decisão tenham sido desenvolvidos, a maioria das atividades foram direcionadas para a construção de modelos que não envolvem controle de dados em tempo real (Chaturvedi et al., 1993, p. 213). Dentre os artigos selecionados, apenas três desenvolveram técnicas de detecção de anomalias em sistemas financeiros em tempo real (Forough e Momtazi, 2021, p. 10; Rubio et al., 2020, p. 15; Patil et al., 2018, p. 385) e puderam identificar padrões em transações recorrentes. Além disso, muitos estudos selecionam conjuntos de dados do *UCI Machine Learning Repository*³, que possuem dimensionalidades baixas, com menos de 100 recursos, enquanto circunstâncias referentes ao mundo real contêm dados altamente dimensionados, com milhares de recursos (Tsai e Chen, 2014, p. 80), como o artigo de K. Chen et al. (2020, p. 3). Essa análise revela as dificuldades decorrentes do processo de decisão em tempo real (Chaturvedi et al., 1993, p. 214), uma vez que os modelos de detecção de anomalias financeiros desenvolvidos não necessariamente são capazes de suprir a necessidade de identificação de fraudes, desvios ou adulteração de dados para que os sistemas de suporte à decisão consigam fazer as escolhas mais adequadas, conseguindo detectar e prevenir ações criminosas. Com base nisso, pesquisas futuras podem abordar os problemas de dimensionalidade de modo a desenvolver técnicas que viabilizem o processamento de um grande volume de dados.

³Repositório de dados para serem aplicados algoritmos de *machine learning*.

A disponibilidade dos dados utilizados para pesquisa pode viabilizar uma maior compreensão do estudo a ser desenvolvido. Desse modo, essa classificação foi feita de modo a entender se os dados utilizados para testar os modelos de detecção de anomalias são restritos ou não e como isso pode influenciar na interpretação dos resultados. Os dados estatísticos relacionados a essa classificação podem ser vistos na Figura 4.8.

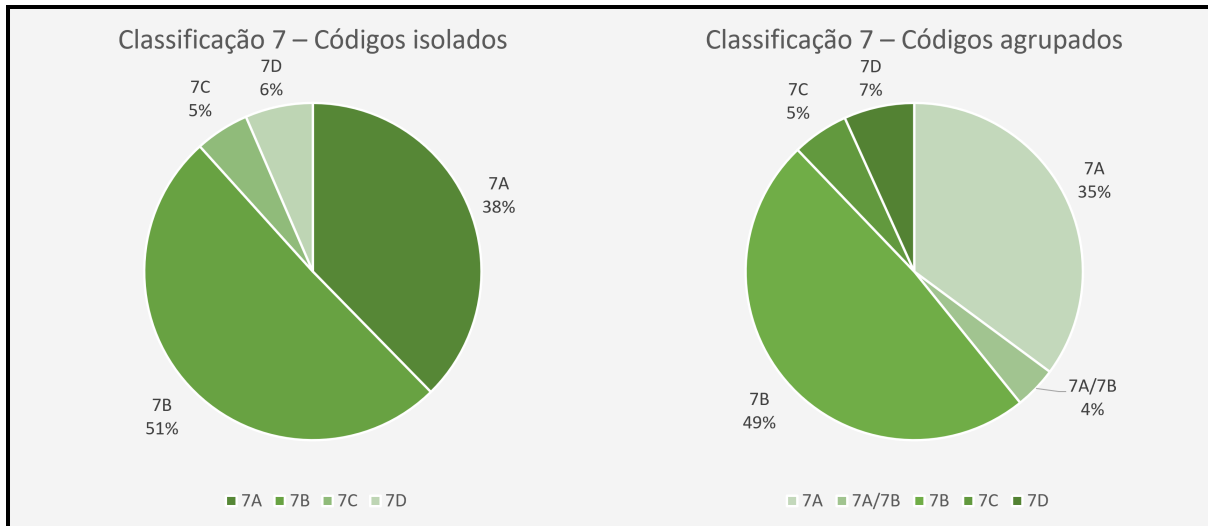


Figura 4.8: Disponibilidade.

Notas: 7A - Acesso restrito; 7B - Acesso irrestrito; 7C - Outros; e 7D - Não se aplica.

A codificação mais frequente foi a “7B - Acesso irrestrito”, utilizada para mais da metade dos artigos. No entanto, a codificação “7A - Acesso restrito” também contemplou uma porcentagem significativa dos artigos, ou seja, 38%. Desse modo, por mais que a maioria dos artigos dentre os selecionados tenham utilizado dados disponíveis para verificar a consistência do modelo desenvolvido, ainda há um número expressivo de artigos que não disponibilizam seus dados, principalmente devido à necessidade de manter dados privados para poupar a exposição de empresas, por exemplo. Porém, as preocupações com a privacidade intensificam os desafios de integração, compartilhamento e análise de dados (Wimmer, Yoon, e Sugumaran, 2016, p. 51). Para Canhoto (2021, p. 444) é complicado obter acesso aos dados financeiros das empresas por causa da natureza secreta que envolve esse setor. Pourhabibi et al. (2020, p. 8) afirmaram em sua pesquisa que existem regulações que proíbem a transmissão e distribuição de dados confidenciais, sejam eles pessoais ou financeiros, dificultando o compartilhamento de dados e impulsionando o uso de medidas de avaliação matemática para detectar fraudes. Um exemplo disso é a pesquisa de K. Singh e Best (2019, pp. 3–4) que, por mais que utilize dados reais, eles foram obtidos a partir de transações bancárias de um grande entidade financeira e, desse modo, as informações e especificidades foram retiradas da pesquisa devido a confidencialidade.

Essa situação coincide com a sexta lacuna já discutida na subseção “Método de pesquisa”, isto é, não há uma base padronizada que permita verificar a eficiência dos modelos de detecção de anomalias desenvolvidos. Porém, a não disponibilidade dos dados é outro fator que prejudica a mensuração, caracterizando a décima primeira lacuna de pesquisa, a saber:

L₁₁ A falta de dados disponíveis para a compreensão do desenvolvimento dos modelos de detecção nas pesquisas.

Essa lacuna faz com que não seja possível interpretar plenamente a técnica estudada, como também, não seja possível realizar comparações entre pesquisas e modelos, inviabili-

zando a compreensão da efetividade de identificação de anomalias nos sistemas financeiros, seja pela falta de dados, seja por diferenças de tamanho entre os dados. Desse modo, é crítica a necessidade de uma base de dados padronizadas que possibilite que tal comparação seja feita, principalmente considerando que, em muitas pesquisas, os dados utilizados para testagem não são sequer disponibilizados, o que pode tornar a evolução dessa temática de pesquisa imprecisa e deficiente. Pesquisas futuras, portanto, devem considerar a estruturação de teorias e ferramentas que viabilizem um processo de pesquisa mais pertinente e real.

4.8 TÉCNICA UTILIZADA

Essa categoria de classificação foi estruturada de modo a identificar as tendências de técnicas utilizadas nos modelos de detecção de anomalias em sistemas financeiros, como também, sinalizar quais técnicas estão sendo menos utilizadas. Isso é feito de modo que seja possível identificar não só oportunidades de pesquisa, mas as técnicas mais pertinentes e efetivas para esse campo de pesquisa de acordo com a frequência de uso, considerando que a investigação de padrões comportamentais comuns e incomuns na área financeira é de extrema importância para o controle eficiente e seguro do funcionamento dos sistemas (León et al., 2020, p. 1). Os dados estatísticos referentes ao processo de classificação e codificação dos artigos foi apresentado na Figura 4.9.

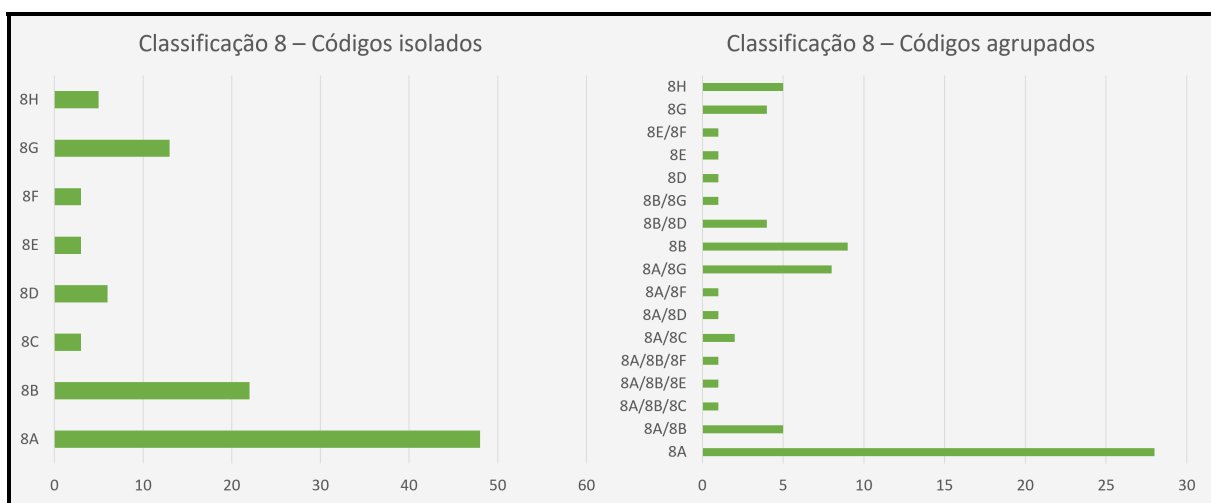


Figura 4.9: Técnica utilizada.

Notas: 8A - *Machine learning*; 8B - Estatística; 8C - Estatística Bayesiana; 8D - Lei de Benford; 8E - Métodos gráficos (visuais); 8F - Simulações computacionais; 8G - Outros; e 8H - Não se aplica.

A codificação mais frequente entre os artigos foi a “8A - *Machine learning*”, contemplando 46,60% dos artigos. A escolha dessa técnica pode estar relacionada com o fato de que métodos de pesquisa envolvendo *machine learning* foram e são amplamente utilizados quando se trata de problemas envolvendo decisões financeiras (Kuzey, Uyar, e Delen, 2014, p. 128) e, ainda, devido a grande quantidade de *frameworks* de códigos e algoritmos disponíveis abertamente para desenvolver os modelos (A. I. Khan e Al-Badi, 2020, p. 573), o que viabiliza o acesso dos autores às bibliotecas de programação das técnicas a serem modeladas. Os resultados de pesquisa de Delen (2010, pp. 501–502) confirmaram a superioridade de métodos de *machine learning* para a produção de melhores formas de prevenir situações. A capacidade de aprendizado advinda do processo de *machine learning* viabiliza a adaptação e aprimora a performance dos sistemas (Ali,

Shamsuddin, e Ismail, 2012, p. 568), o que justifica a escolha dos autores para a construção de modelos de detecção de desvios nos padrões de dados.

Para León et al. (2020, p. 1), as técnicas de *machine learning* lideram o desenvolvimento de ferramentas e técnicas automatizadas capazes de controlar e supervisionar sistemas de pagamento e as diversas infraestruturas referentes ao mercado financeiro. Carmona et al. (2019, p. 308) discutiram sobre o fato de que abordagens que utilizam técnicas de *machine learning* são importantes para as áreas de detecção de fraudes e de eventos anômalos. Canhoto (2021, p. 442), nessa mesma linha de raciocínio, afirmou que explorar o potencial de técnicas de *machine learning* pode ser de grande auxílio à luta global contra o crime, uma vez que sinalizam anomalias e *outliers* presentes dentro de um padrão de dados (Baesens et al., 2021, p. 6). Carneiro et al. (2017, p. 91), por outro lado, discutiram sobre como, por mais que os algoritmos de *machine learning* aperfeiçoem os métodos de detectar fraude, raramente são relatados o uso de técnicas que são realmente práticas e relevantes a essa temática de pesquisa.

Além disso, os autores discordam sobre qual método de *machine learning* é o mais eficiente para detectar anomalias. Para Y. Wang e Xu (2018, p. 87), as técnicas de redes neurais possuem uma performance superior a outros modelos de *machine learning* muito utilizados, como, florestas aleatórias e máquina de vetores de suporte. Forough e Momtazi (2021, p. 2) afirmaram que, quando se trata de aplicações da vida real, o melhor método a ser utilizado é o de florestas aleatórias. Badal-Valero et al. (2018, p. 33), porém, discutiram sobre como o uso de florestas aleatórias, na verdade, é a melhor abordagem apenas quando considera-se a aplicação de técnicas de balanceamento, e redes neurais possuem uma performance superior apenas quando envolvem conjuntos de dados desequilibrados. De todo modo, os resultados indicam que as técnicas de *machine learning* são passíveis de detectar atividades suspeitas em tempo real, e isso é um fator importante para a redução de perdas econômicas em empresas (Yan et al., 2020, p. 11).

É oportuno destacar que nos resultados obtidos dessa classificação são apresentadas diversas combinações de códigos que incluem a codificação “8A - *Machine learning*”, o que sinaliza a importância do uso dessa técnica combinada mesmo quando os autores optam por usar outros modelos de detecção. Porém, cabe aqui ressaltar as técnicas desenvolvidas com base na Lei de Benford de detecção de desvios em dados, simbolizada pela codificação “8D - Lei de Benford”, que foi utilizada por 5,83% dos artigos selecionados. Ela normalmente é desenvolvida em um modelo híbrido com técnicas estatísticas que aprimoram seus resultados (Azevedo et al., 2021, p. 3; Druică et al., 2018, p. 78, Shi et al., 2018, pp. 879–880), sendo uma importante ferramenta para a detecção de fraudes financeiras quando se trata da observação da conformidade de números que ocorrem naturalmente nos sistemas de empresas (Bhattacharya, Xu, e Kumar, 2011, pp. 576–577). Um argumento acerca da não confiabilidade de detecção de fraude pela Lei de Benford é baseado no fato de que potenciais criminosos podem estar cientes da necessidade de os dados estarem distribuídos conforme ela e, por causa disso, é importante que autores ao desenvolverem seus modelos de detecção considerem métodos alternativos que possam complementar a análise de distribuição (Shi et al., 2018, p. 879). A Lei de Benford, portanto, quando utilizada corretamente, pode ser considerada uma ferramenta poderosa para identificar anomalias e atividades suspeitas (Azevedo et al., 2021, p. 11).

No entanto, tal técnica não só não é frequente dentre os últimos artigos que envolvem abordagens de detecção de anomalias em sistemas financeiros, mas, como também, seu uso foi pouco combinado com *machine learning*, a técnica mais frequente, amplamente utilizada em conjunto com as demais técnicas em análise. A codificação agrupada “8A/8D” foi utilizada em apenas um artigo, em que quatro modelos de classificação de *machine learning* foram associados com a conformidade de distribuição de números da Lei de Benford para identificar padrões de desvio de dinheiro que pudessem sinalizar atividades criminosas de lavagem de dinheiro (Badal-Valero et al., 2018, p. 24). Os resultados obtidos indicam que o modelo híbrido consegue

identificar anomalias e, ainda, avaliar a conformidade com a distribuição de números significativos da Lei de Benford, considerando os algoritmos de *machine learning* mais apropriados para cada situação (Badal-Valero et al., 2018, pp. 31–33). Desse modo, é possível identificar a décima segunda lacuna, a saber:

L₁₂ Quais as vantagens de combinar técnicas diversas com *machine learning* no processo de detecção de anomalias?

4.9 NÚMERO DE ESTRATÉGIAS

O número de estratégias refere-se à quantidade de técnicas utilizadas em um mesmo artigo para promover o desenvolvimento de um modelo de detecção de anomalias em sistemas financeiros que seja eficaz e relevante para os estudos dessa temática. Essa classificação reforça os argumentos já apresentados na subseção “Técnica utilizada”, e esclarece a quantidade de artigos que realizam modelos híbridos, combinando técnicas de detecção. Os resultados estatísticos obtidos dessa classificação foram apresentados na Figura 4.10.

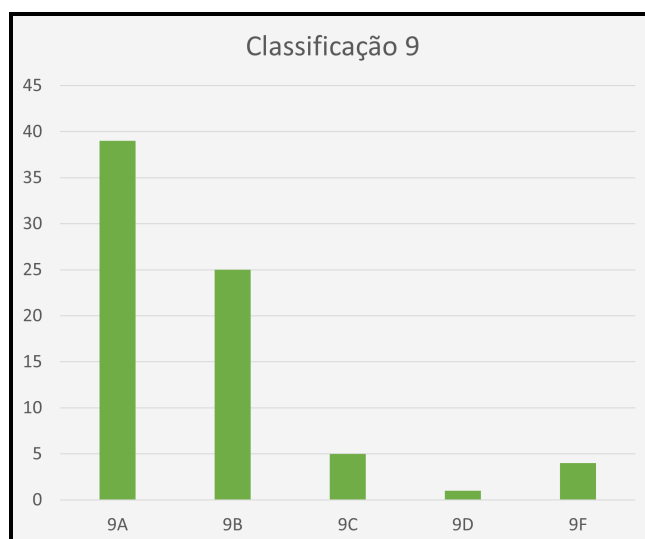


Figura 4.10: Número de estratégias.

Notas: 9A - Uma; 9B - Duas; 9C - Três; 9D - Quatro ou mais; e 9F - Não se aplica.

Modelos que integram duas ou mais técnicas possuem o potencial de aumentar a qualidade da performance de predição (K. C. Lee, Han, e Kwon, 1996, p. 63; Liang, Chandler, e Han, 1990, p. 392). Isso, porque abordagens híbridas que combinam diferentes técnicas são capazes de superar limitações que ocorreriam ao utilizar uma técnica isolada e, ainda, conseguem unir as vantagens de cada uma de modo a aprimorar o modelo desenvolvido (Kundu, Pal, e Mandal, 2020, p. 1; Ince e Trafalis, 2006, pp. 1060–1061). Considerando a temática de detecção de anomalias, como a grande quantidade de dados dos sistemas financeiros impossibilita a detecção manual de anormalidades (Gianini et al., 2020, p. 549), a combinação de técnicas pode atuar também de forma estratégica ao promover a articulação e colaboração humano-computador, em um modelo híbrido com potencial de aumentar ainda mais a capacidade de tomadores de decisão humanos (Bayrak, McComb, Cagan, e Kotovsky, 2021, p. 1). Desse modo, o número de estratégias utilizadas em um único modelo pode aumentar as chances de sucesso de detecção de anomalias financeiras, tanto em termos computacionais, quanto em termos de gestão humana do sistema.

Considerando isso, a maior parte dos artigos foi classificado com a codificação “9A - Uma”, contemplando 52,70%, o que significa que é muito frequente que os autores utilizem apenas uma estratégia para construção de seus modelos de estudo, não considerando a utilização de técnicas combinadas. No entanto, a codificação “9B - Duas” também considera uma porcentagem expressiva dos artigos selecionados, de 33,78%, e os autores de tais artigos revelam em seus estudos os benefícios provenientes da utilização de modelos híbridos. Para Baesens et al. (2021, p. 1), a investigação de atividades suspeitas, quando se considera, por exemplo, a perspectiva de *machine learning*, pode ser vista como um problema de classificação binária e, por causa disso, muitas técnicas e modelos podem ser aplicados. Carneiro et al. (2017, p. 93), nessa mesma linha de raciocínio, afirmam que, para que o modelo seja mais efetivo, é possível combinar soluções com técnicas supervisionadas e não supervisionadas. Segundo Jurgovsky et al. (2018, p. 244), isso ocorre, porque combinações podem aperfeiçoar a precisão do modelo de detecção.

Além disso, um modelo híbrido é capaz de cobrir um espectro mais completo de atividades disruptivas (Zhai et al., 2017, p. 27) de modo a aprimorar o processo de predição de padrões e anomalias (Rtayli e Enneya, 2020, p. 2). Com a combinação de técnicas a precisão é aumentada, porque se trata de mais de um modelo de reconhecimento de anormalidades em um único processo de detecção (K. Chen et al., 2020, p. 3). Desse modo, os modelos e técnicas podem se complementar de forma estratégica (Baesens et al., 2021, p. 2) para combater fraudes, desvios, adulteração e inconsistência de dados, ao alcançar performances mais estáveis e adequadas à investigação de anomalias (Zhai et al., 2017, p. 38). Isso caracteriza a décima terceira lacuna de pesquisa, a saber:

L₁₃ A falta de mais modelos híbridos de detecção de anomalias financeiras que explorem combinações diversificadas de técnicas de modo a identificar quais modelos são mais eficientes para a detecção de anomalias em sistemas financeiros.

Assim, pesquisas futuras podem examinar mais a condução de estudos que envolvam a mistura de técnicas e modelos, visando aumentar os níveis de precisão da detecção estrategicamente, e esclarecendo, ainda mais, os benefícios advindos do uso de modelos híbridos no setor financeiro.

4.10 ÁREA DO PERIÓDICO

O estudo do campo de Finanças faz parte da área de conhecimento de ciências sociais aplicadas, porém essa classificação visa compreender as categorias nas quais os artigos referentes à aplicação de modelos de detecção de anomalias em sistemas financeiros estão inseridos e se, realmente, por serem artigos que englobam questões financeiras, estes contemplam a categoria de Ciências Sociais. Isso, porque a análise de citação de quando estuda-se um campo base específico revela que é mais provável ao conduzir uma pesquisa científica que a citação de artigos e periódicos dela esteja contida, principalmente, dentro desse campo base (Holsapple, Johnson, Manakyan, e Tanner, 1995, p. 360). Uma comparação entre os dados estatísticos provenientes das classificações de 2017 até julho 2021 e os dados obtidos ao adicionar os artigos de agosto de 2021 até outubro 2022 pode ser vista na Figura 4.II.

A partir dos resultados, é possível observar que, por mais que a quantidade de artigos com a codificação “10D - Ciências sociais e humanas” tenha sido expressiva até 2021, contemplando 40% dos artigos selecionados, essa categoria não era a mais frequente. A codificação mais presente foi a “10C - Ciências físicas e engenharia”, abrangendo 56% dos artigos. No entanto, ao adicionar os artigos mais recentemente produzidos, esse resultado se reverte. A codificação

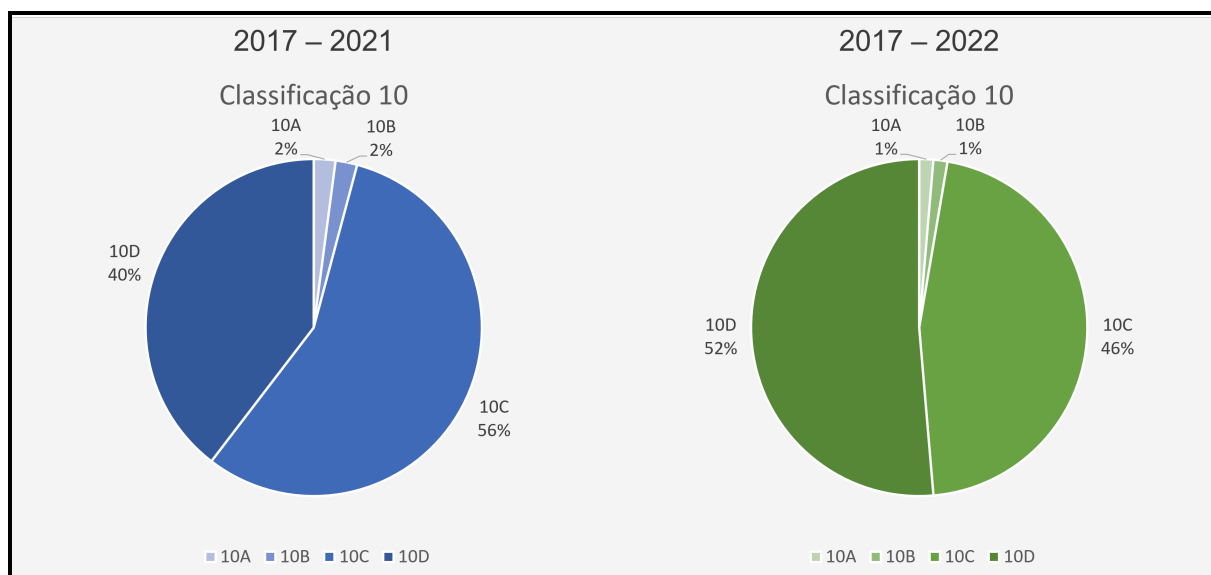


Figura 4.11: Área do periódico.

Notas: 10A - Ciências da saúde; 10B - Ciências da vida; 10C - Ciências físicas e engenharia; e 10D - Ciências sociais e humanas.

“10D”, com aproximadamente 52%, supera em 4 artigos a codificação “10C”, que contempla aproximadamente 46% dos artigos.

Desse modo, é interessante observar que por mais que finanças se encontre dentro do campo de ciências sociais, os estudos acerca de anomalias financeiras estão sendo desenvolvidos e publicados, também, em periódicos relacionados a ciências físicas. Isso pode estar acontecendo, uma vez que os artigos selecionados para esta revisão de literatura promovem a construção de modelos, muitas vezes, quantitativos que detectam dados que divergem de um padrão de normalidade, utilizando-se de simulações computacionais (K. Singh e Best, 2019, p. 4; Patil et al., 2018, p. 387), *machine learning* (Severino e Peng, 2021, p. 3; Climent et al., 2019, p. 888), Lei de Benford (Arboleda et al., 2018, pp. 13–14; Druică et al., 2018, p. 77), estatística (Ardizzi et al., 2018, p. 108; Mamede e Malaquias, 2017, pp. 49–50), entre outras técnicas que, normalmente, estão relacionadas à computação e à engenharia. No entanto, apesar de essa temática de detecção de anomalias possuir tais características, os estudos ainda possuem enfoque na área financeira, o que caracteriza a necessidade de desenvolver cada vez mais estudos nessa área, principalmente considerando questões relacionadas a ciências sociais. Desse modo, tem-se a décima quarta lacuna de pesquisa, a saber:

L₁₄ Aprofundar uma análise acerca do por quê não é utilizada uma visão mais humanista para tratar de anormalidades em finanças, além de identificar padrões de publicação dentro das categorias de Ciências Sociais e Ciências Físicas, de modo a investigar correlações entre as áreas de conhecimento.

Além disso, apesar das baixas porcentagens, as codificações “10A - Ciências da saúde” e “10B - Ciências da vida”, apresentaram artigos classificados (R.B. e K.R., 2021, p. 35; Yan et al., 2020, p. 115), mesmo se tratando de categorias de periódico improváveis considerando a temática em análise. Desse modo, pode-se identificar aqui, a décima quinta lacuna de pesquisa, a saber:

L₁₅ Desenvolver pesquisas para compreender se há a necessidade de aprofundar o estudo de detecção de anomalias financeiras nas áreas de Ciências da saúde e da vida.

Pesquisas futuras podem contemplar o desenvolvimento de artigos dentro desses setores de estudos, considerando que anomalias podem se manifestar de formas variadas em sistemas financeiros diversos, incluindo aqueles relacionados às áreas de saúde e do campo das ciências biológicas.

CAPÍTULO 5

CONCLUSÃO

“O importante é não parar de
questionar.”
—Albert Einstein.

Comportamentos disruptivos e anômalos em sistemas financeiros provenientes de desvios, adulteração e fraude de dados prejudicam gravemente o mercado de capitais e a economia em geral, devido a propagação de informações enganosas que geram consequências negativas a diferentes *stakeholders* e inviabilizam a alocação eficiente de recursos econômicos (Hou, Wang, e Ma, 2021, p. 97), dificultando uma tomada de decisão mais aprimorada. Com base nisso, as anomalias, por serem inconsistências na execução regular de dados (Conti e Naldi, 2008, p. 420), podem causar ineficiência de performance e manutenção, propiciando a ocorrência de potenciais erros nos sistemas (Mukherjee, Gamble, e Parkinson, 1997, p. 232).

A atual pandemia, nesse contexto, modifica rapidamente os cenários tecnológicos, afetando as práticas de negócios e o comportamento dos dados (Shaikh et al., 2020, p. 2), o que aumenta a ameaça digital, podendo envolver um leque de atores, como cibercriminosos e vítimas, sendo estas pessoas ou sistemas organizacionais (Wessels et al., 2021, p. 1). Por isso, é preciso que os modelos de negócios estejam em constante manutenção (Karki e Porras, 2021, p. 2), criando vantagem competitiva e aprimorando o valor que a organização tem a oferecer aos *stakeholders* (Marikyan et al., 2022, p. 1). Desse modo, é criticamente importante o estudo sobre ferramentas que possam identificar tais anormalidades e que, assim, possam desenvolver e aprimorar o aspecto funcional de tomada de decisão acerca delas.

Com base nisso, objetivo principal desse estudo foi verificar, analisar e sistematizar artigos dos últimos seis anos acerca de métodos e técnicas de detecção de anomalias utilizados em sistemas financeiros, visando compreender e indicar uma base de conhecimento científico existente, como também, lacunas para pesquisas futuras, de modo a viabilizar a criação de uma agenda de pesquisas. Essas informações estão sintetizadas na Figura 5.1. Para isso, foi utilizado o método de pesquisa em cinco passos proposto por Jabbour (2013, p. 145) e Lage Junior e Godinho Filho (2010, p. 14): pesquisa sistemática de artigos, desenvolvimento de uma estrutura de classificação, classificação dos artigos pré selecionados, delineamento de um *framework* de resultados, e discussão de resultados para propor sugestões para pesquisas futuras.

Desse modo, usando 10 categorias, foi definido um *framework* de classificação de 74 artigos identificados como relevantes nas pesquisas na base de dados *ScienceDirect*[®] por meio de codificações para sistematizar, de maneira padronizada, o conhecimento existente acerca da temática em estudo. Ao adotar esse método de pesquisa, foram superadas as limitações encontradas em literaturas anteriores e foram reveladas combinações de características presentes nos artigos que possibilitam que futuros pesquisadores possam aprimorar seus estudos de maneira mais precisa e embasada (Mallinguh e Zéman, 2020, p. 163), viabilizando a evolução da base de conhecimento acerca de métodos e abordagens de detecção de anomalias e, a partir disso, oferecendo benefícios através de uma lente financeira/econômica.

Logo, este trabalho faz várias contribuições para o desenvolvimento da área de estudos acerca da detecção de anomalias em sistemas financeiros. A partir do *framework* de classificação criado e dos resultados estatísticos obtidos foi possível identificar 15 lacunas de pesquisa principais que estão distribuídas ao longo do texto. Cada uma dessas lacunas pode ser considerada importante para o avanço científico em pesquisas futuras. No entanto, o método dos cinco passos utilizado, ao definir as tendências de pesquisa, possibilitou verificar três lacunas que mais se destacam, por não estarem presentes ou não serem levadas em consideração na maioria dos artigos, mas possuem características pertinentes à evolução dos modelos de detecção de anomalias. Elas são a falta de uma base de dados comum, em que seja possível utilizar dados padronizados para promover a comparação e mensuração efetiva dos modelos de detecção; a necessidade de promover estudos que testem seus modelos em bases de dados de tamanhos diferentes, buscando mitigar os problemas decorrentes da dimensionalidade de dados; e a falta de indicadores de mensuração da efetividade da técnica escolhida para detecção de padrões de anormalidade, de modo a aprimorar ainda mais o processo comparativo entre os modelos.

Essas três lacunas são mais relevantes para o presente trabalho, uma vez que, com o crescimento dos negócios digitais e, conseqüentemente, com o aumento de transações financeiras, movimentações bancárias, fluxo de demonstrações contábeis, entre outros, as abordagens de detecção de anomalias, bem como os estudos desenvolvidos para testar, adaptar e aprimorar as técnicas, devem ser capazes de acompanhar a transformação digital (Konopik et al., 2022, p. 1). Isso, porque o rápido crescimento dessas tecnologias no setor financeiro contribuem para a grande acumulação de dados em tempo real fluindo nas redes (Habeeb et al., 2019, p. 289), de modo que são necessários métodos de detecção de anomalias financeiras eficientes e escaláveis a fim de trabalhar grandes conjuntos de dados, considerando volume, velocidade, variedade e veracidade (Fertier et al., 2020, p. 1) para que a chamada maldição da dimensionalidade seja superada eficientemente.

Assim, pode-se inferir a partir deste estudo como torna-se criticamente importante desenvolver pesquisas que considerem as 3 lacunas mencionadas. Isso, porque quando a dimensão dos dados reais é muito maior do que a amostra escolhida para um modelo de detecção, a eficiência se torna limitada (J.-T. Zhang et al., 2021, p. 144) e a não comparação entre modelos dificulta a interpretação de quais realmente são mais eficientes e quais são mais restritos (H. Liu et al., 2010, p. 270), assim como, a inexistência de indicadores que possibilitem essa mensuração. Para Zimek, Schubert, e Kriegel (2012, p. 364), as dificuldades decorrentes da maldição da dimensionalidade têm servido de inspiração para o desenvolvimento de muitos estudos, mas deixam questões não resolvidas e seus resultados não são suficientemente detalhados, assim como mostram os resultados do presente trabalho. Por mais que os métodos utilizados para cada conjunto de dados sejam considerados difíceis de generalizar, muitos não foram totalmente explorados para serem identificados como técnicas eficientes, o que torna os resultados das pesquisas menos adequados à evolução do conhecimento sobre como quebrar a maldição (Gupta, Gao, Aggarwal, e Han, 2014, p. 2263) e como aprofundar as capacidades técnicas dos modelos de detecção.

Desse modo, a base de dados comum e a aplicação dos modelos em conjuntos de dados de tamanhos diferentes poderiam intervir nos empecilhos mencionados, uma vez que, ter uma

base padronizada para testar a precisão e eficiência de um modelo de detecção, faz com que seja possível explorar e mensurar não somente a capacidade de identificação de anomalias, como também a habilidade do modelo de atuar em diferentes volumes de dados. Os indicadores que poderiam advir dessa mensuração sinalizariam informações relevantes, como características positivas e negativas dos modelos, o potencial de detecção, bem como, a habilidade de lidar, por exemplo, com dados de alta dimensionalidade ou em tempo real, de modo que estudos futuros podem utilizar-se desses benefícios para aprimorar ainda mais a forma como se lida com anomalias na contemporaneidade, consolidando as melhores técnicas para mitigar atos ilegais e erros transacionais ou de registro, possibilitando uma evolução saudável dos sistemas financeiros no mundo digital.

No entanto, é importante atentar-se que, por mais que essas lacunas sejam sanadas e os indicadores apontem dados relevantes ao desenvolvimento de futuros modelos de detecção, não há um único algoritmo que seja considerado o melhor para quebrar a maldição da dimensionalidade e analisar e identificar os dados com mais precisão (Chatterjee e Ahmed, 2022, p. 12). Isso, porque os sistemas financeiros são muito diversos e existem diversas técnicas que podem ser indicadas para variadas aplicações e modelos de negócio. Assim, o benefício de sanar as lacunas seria afunilar as técnicas existentes em cada área de aplicação financeira, considerando volumes de dados variados e bases de dados comuns para testes e mensuração a partir de indicadores, de modo que futuros pesquisadores, bem como, gestores de organizações, possam escolher modelos e técnicas de detecção de anomalias que julguem pertinentes para se aprofundar ou administrar dados e colher informações relevantes para tomadas de decisão de modo estratégico.

Com base nisso, a batalha contra a maldição da dimensionalidade considerando os sistemas financeiros ainda não se mostra em um nível satisfatório (D. Li, Wang, Wang, e Yao, 2008, p. 7778). As implicações decorrentes disso para o sistema financeiro são diversas e prejudiciais ao desenvolvimento dessa área de pesquisa e à aplicação direta de modelos nas organizações. Isso porque modelos de negócios precisam de alta manutenção (Karki e Porras, 2021, p. 2) para que possam melhorar sua tomada de decisão e, conseqüentemente, melhorar o valor que a organização pode oferecer aos *stakeholders* (Marikyan et al., 2022, p. 1). Logo, é de suma importância que futuras pesquisas atentem para as dificuldades impostas pela maldição da dimensionalidade, para que os estudos não sejam realizados em vão e possam de fato agregar *insights* relevantes quando se trata de análise de dados de alta dimensão e detecção de anomalias, fraudes e *outliers*.

Portanto, os autores de trabalhos futuros devem explorar mais profundamente métodos conhecidos e adaptados a dados de alta dimensão e em tempo real. Eles devem propor novos métodos que possam questionar os desafios impostos pela maldição da dimensionalidade e pelas constantes transformações digitais nos sistemas financeiros. Além disso, na medida do possível, devem disponibilizar as bases de dados e algoritmos utilizados em simulações ou testes computacionais para outros pesquisadores realizarem testes comparativos, ou, então, aderir ao uso das bases de dados comuns para efetuar os testes de eficiência. Assim, os estudos nessa área podem encontrar um ponto comum de análise e aprofundamento, consolidando as melhores técnicas para mitigar atos ilícitos ou erros nas transações financeiras, contribuindo para uma evolução adequada dos sistemas financeiros (Pinto e Sobreiro, 2022, p. 19). Outra solução conveniente para promover os testes comparativos de eficiência de técnicas e modelos para superar a maldição e as lacunas dessa área de pesquisa, seria comunidade acadêmica disponibilizar um *big data* que possa servir como referência de desempenho para diversos estudos, como comumente ocorre na área de pesquisa operacional (Taillard, 1993, p. 278). Assim, além de vários estudos partindo do mesmo ponto, também será possível verificar a consistência de resultados de estudos anteriores e implementar soluções em diferentes áreas do conhecimento, identificando os melhores modelos e técnicas para cada cenário.

Além disso, vale ressaltar que dentre os artigos analisados na base de revisão de literatura, não houve nenhum estudo acerca dos pagamentos no modelo PIX, considerando o cenário

financeiro brasileiro. Por mais que o PIX seja recente, tendo sido criado em 2020, e sua popularização tenha sido rápida (A. R. Ferreira, 2022, p. 107), ainda não é uma tecnologia financeira muito explorada em estudos científicos, principalmente, considerando que essa forma de pagamento é exclusiva do Brasil e foi utilizada uma base de artigos internacional. Porém, o uso do PIX tem sido cada vez mais comum nos estabelecimentos comerciais brasileiros, inclusive nos *e-commerces* (Ávila et al., 2022, p. 12017), o que sinaliza uma necessidade de estruturar e estudar condições de segurança para o usuário, bem como criar alertas para atos criminosos envolvendo esse sistema de pagamento (Lima, 2020, p. 1). O banco central brasileiro, inclusive, reforça a necessidade de as instituições financeiras atentarem-se a questões de segurança, de modo a prevenir fraudes, lavagem de dinheiro ou financiamento ao terrorismo (Banco Central, 2020b, p. 1). Desse modo, considerando o contexto brasileiro, estudos futuros poderiam desenvolver pesquisas e modelos para detectar anomalias de modo a proteger os usuários e organizações dessas ações criminosas.

Outro ponto importante a se considerar a partir deste trabalho se pauta nas técnicas utilizadas para a construção dos modelos de detecção. *Machine Learning* foi a técnica mais utilizada, porém uma outra técnica que aparece bastante quando se trata desses modelos é a Lei de Benford, mas não foi uma técnica tão escolhida pelos autores da base de revisão de literatura. Isso porque, como se trata de uma lei da natureza em que os números seguem um padrão de distribuição, os criminosos, tendo conhecimento disso, podem remodelar os dados de modo que se encaixem dentro da Lei (Shi et al., 2018, p. 879). No entanto, caso sejam utilizados métodos complementares em conjunto com a Lei, como abordagens estatísticas ou até mesmo as técnicas de *machine learning*, ela pode ser considerada uma ferramenta de grande relevância para a detecção de anomalias e atividades suspeitas (Azevedo et al., 2021, p. 11). Desse modo, estudos futuros podem, também, desenvolver pesquisas que levem mais a Lei de Benford em consideração.

Consequentemente, o desenvolvimento e evolução científica dos artigos considerando tais lacunas é criticamente importante para o aprimoramento dos sistemas de tomada de decisão, principalmente em termos financeiros. Logo, o *framework* proposto e as lacunas advindas da análise estatística dos resultados obtidos, indicam que essa área de estudos ainda pode ser muito explorada, de modo aprimorar os modelos de detecção de anomalias, aumentando a precisão de resultados e propiciando a construção de estruturas que promovam o suporte necessário ao processo de tomada de decisão.

No entanto, é oportuno destacar que essa revisão de literatura possui algumas limitações, dentre as quais pode-se destacar o fato de que para fazer o levantamento dos artigos só foi utilizada a base de dados *ScienceDirect*[®]. Além disso, é possível que alguns estudos relevantes tenham sido omitidos, considerando as palavras-chave e os filtros de pesquisas utilizados. Desse modo, pesquisas futuras podem explorar a detecção de anomalias em outros sistemas financeiros que não apareceram ou foram pouco explorado entre os estudos analisados neste trabalho, como *blockchain*, que pode criar recursos para as organizações ao alterar o modo como as partes envolvidas se conectam na troca financeira digital (Marikyan et al., 2022, p. 1), ou então, considerar a interconexão entre sistemas físicos e digitais (Wessels et al., 2021, p. 1).

Ademais, a técnica dos cinco passos, por buscar identificar traços comuns a partir da classificação, leitura, entendimento e percepção acerca das pesquisas, torna-se complexa de ser completamente automatizada ou quantificada, a não ser quando se considera os resultados proporcionados pelo método. Desse modo, pesquisas futuras podem complementar as análises realizadas com base em outras perspectivas, principalmente no caso de pesquisas quantitativas, como no estudo desenvolvido por Alamsyah, Rahardjo, e Kuspriyanto (2013, p. 3), que discorre sobre as conexões existentes dentro de sua temática de estudo, criando uma taxonomia para reconhecer os limites dessa temática e reforçar o mapeamento de esforços e interesses de investigação ao desenvolver uma revisão de literatura.

De todo modo, com a quantidade de artigos selecionada e com os resultados obtidos, pes-

quisas futuras podem se basear nesta revisão de literatura para identificar as lacunas de pesquisas mais pertinentes a serem desenvolvidas e analisadas. Com isso, gestores poderão encontrar nas pesquisas científicas um guia para administrar seus sistemas financeiros de maneira mais eficiente, alocando recursos para implementar modelos de detecção de anomalias pertinentes ao cenário em que se encontram de modo a garantir a segurança dos dados produzidos, diminuir possíveis custos relacionados a fraude ou desvios de dinheiro, bem como gerar valor aos seus clientes e consumidores. Assim, a administração das finanças dentro das organizações passa a ser mais do que apenas uma gestão de tarefas, processos ou planejamento financeiro, passa a ser uma gestão estratégica de dados a partir da análise e controle de informações que estão, muitas vezes, ocultas, mas são determinantes para o funcionamento e crescimento organizacional de maneira correta. Administrar é analisar, compreender e proteger.

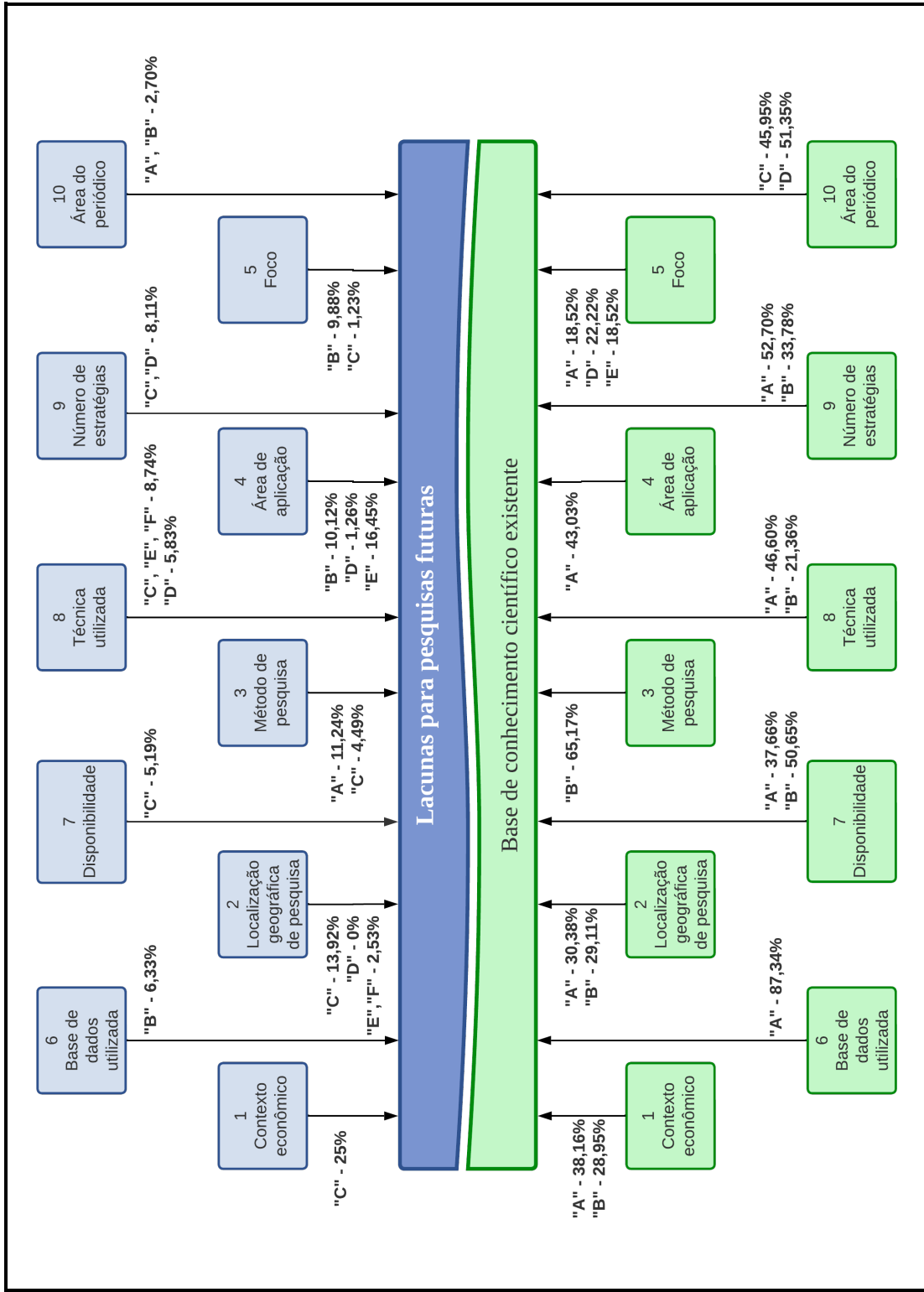


Figura 5.1: Sugestões de pesquisas futuras.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abreu, E. S. d., Kimura, H., e Sobreiro, V. A. (2019, January). [What is going on with studies on banking efficiency? Research in International Business and Finance](#), 47(1), 195–219.
(Citado 1 vez na página 30.)
- Abu, F., Gholami, H., Saman, M. Z. M., Zakuan, N., Sharif, S., e Streimikiene, D. (2021, May). [Pathways of lean manufacturing in wood and furniture industries: a bibliometric and systematic review](#). *European Journal of Wood and Wood Products*, 79(4), 753–772.
(Citado 1 vez na página 30.)
- Agi, M. A., e Jha, A. K. (2022, May). [Blockchain technology in the supply chain: An integrated theoretical perspective of organizational adoption](#). *International Journal of Production Economics*, 247(1), 108458. (Online first.)
(Citado 1 vez na página 15.)
- Ahmed, A., Deokar, A., e Lee, H. C. B. (2021, September). [Vulnerability disclosure mechanisms: A synthesis and framework for market-based and non-market-based disclosures](#). *Decision Support Systems*, 148(1), 113586. (Online first.)
(Citado 1 vez na página 29.)
- Ahmed, N. A. (2022). [What is Artificial Intelligence? \(AI\)](#). (Publicado online.)
(Citado 1 vez na página 11.)
- Ahmed, S., Alshater, M. M., Ammari, A. E., e Hammami, H. (2022, October). [Artificial intelligence and machine learning in finance: A bibliometric review](#). *Research in International Business and Finance*, 61(1), 101646. (Online first.)
(Citado 3 vezes nas páginas 10, 12, e 21.)
- Ain, N., Vaia, G., DeLone, W. H., e Waheed, M. (2019, October). [Two decades of research on business intelligence system adoption, utilization and success - A systematic literature review](#). *Decision Support Systems*, 125(1), 113113. (Online first.)
(Citado 1 vez na página 29.)
- Alamsyah, A., Rahardjo, B., e Kuspriyanto. (2013, June). [Social network analysis taxonomy based on graph representation](#). In R. R. M. A. Magetsari (Ed.), *The 5th indonesia international conference on innovation, entrepreneurship, and small business (iicies 2013)* (pp. 1–11). Bandung, West Java, Indonesia: IICIES.
(Citado 1 vez na página 76.)
- Albayati, H., Kim, S. K., e Rho, J. J. (2020, August). [Accepting financial transactions using blockchain technology and cryptocurrency: A customer perspective approach](#). *Technology in Society*, 62(1), 101320. (Online first.)
(Citado 1 vez na página 2.)
- Al-Hashedi, K. G., e Magalingam, P. (2021, May). [Financial fraud detection applying Data Mining techniques: A comprehensive review from 2009 to 2019](#). *Computer Science Review*, 40(1), 100402. (Online first.)
(Citado 8 vezes nas páginas 4, 29, 34, 43, 44, 49, 53, e 60.)
- Ali, W., Shamsuddin, S. M., e Ismail, A. S. (2012, June). [Intelligent web proxy caching approaches based on machine learning techniques](#). *Decision Support Systems*, 53(3), 565–579.
(Citado 1 vez na página 67.)
- Amani, F. A., e Fadlalla, A. M. (2017, February). [Data Mining applications in accounting: A review of the literature and organizing framework](#). *International Journal of Accounting Information Systems*, 24(1), 32–58.
(Citado 1 vez na página 29.)
- Amini, S., Elmore, R., Özde Öztekin, e Strauss, J. (2021, October). [Can machines learn capital structure dynamics? Journal of Corporate Finance](#), 70(1), 102073. (Online first.)

- (Citado 1 vez na página 22.)
- Amui, L. B. L., Jabbour, C. J. C., de Sousa Jabbour, A. B. L., e Kannan, D. (2017, January). [Sustainability as a dynamic organizational capability: A systematic review and a future agenda toward a sustainable transition](#). *Journal of Cleaner Production*, 142(1), 308–322.
- (Citado 1 vez na página 5.)
- Andor, G., e Bohák, A. (2017, August). [Identifying events in financial time series - A new approach with bipower variation](#). *Finance Research Letters*, 22(1), 42–48.
- (Citado 3 vezes nas páginas 34, 53, e 64.)
- Ané, T., Ureche-Rangau, L., Gambet, J.-B., e Bouverot, J. (2008, October). [Robust outlier detection for Asia–Pacific stock index returns](#). *Journal of International Financial Markets, Institutions and Money*, 18(4), 326–343.
- (Citado 1 vez na página 20.)
- Arboleda, F. J. M., Guzman-Luna, J. A., e Torres, I.-D. (2018, October). [Fraud detection-oriented operators in a data warehouse based on forensic accounting techniques](#). *Computer Fraud & Security*, 2018(10), 13–19.
- (Citado 6 vezes nas páginas 23, 24, 34, 48, 53, e 71.)
- Ardizzi, G., Franceschis, P. D., e Giammatteo, M. (2018, December). [Cash payment anomalies and money laundering: An econometric analysis of Italian municipalities](#). *International Review of Law and Economics*, 56(1), 105–121.
- (Citado 8 vezes nas páginas 4, 5, 34, 48, 53, 59, 60, e 71.)
- Aremu, O. O., Hyland-Wood, D., e McAree, P. R. (2020, March). [A machine learning approach to circumventing the curse of dimensionality in discontinuous time series machine data](#). *Reliability Engineering & System Safety*, 195(1), 106706. (Online first.)
- (Citado 1 vez na página 26.)
- Arévalo, F., Barucca, P., Téllez-León, I.-E., Rodríguez, W., Gage, G., e Morales, R. (2022, March). [Identifying clusters of anomalous payments in the salvadorian payment system](#). *Latin American Journal of Central Banking*, 3(1), 100050. (Online first.)
- (Citado 2 vezes nas páginas 34 e 53.)
- Arnott, D., Jirachiefpattana, W., e O'Donnell, P. (2007, January). [Executive information systems development in an emerging economy](#). *Decision Support Systems*, 42(4), 2078–2084.
- (Citado 1 vez na página 55.)
- Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., e Shams, T. (2022, December). [Insurance fraud detection: Evidence from artificial intelligence and machine learning](#). *Research in International Business and Finance*, 62(1), 101744. (Online first.)
- (Citado 2 vezes nas páginas 34 e 53.)
- Audibert, J., Michiardi, P., Guyard, F., Marti, S., e Zuluaga, M. A. (2022, December). [Do deep neural networks contribute to multivariate time series anomaly detection?](#) *Pattern Recognition*, 132(1), 108945. (Online first.)
- (Citado 2 vezes nas páginas 20 e 21.)
- Ausloos, M., Ficcadenti, V., Dhesi, G., e Shakeel, M. (2021, July). [Benford's laws tests on S&P 500 daily closing values and the corresponding daily log-returns both point to huge non-conformity](#). *Physica A: Statistical Mechanics and its Applications*, 574(1), 125969. (Online first.)
- (Citado 3 vezes nas páginas 34, 48, e 53.)
- Ávila, C. R. D., Silva, R. F., e Bender, A. (2022, February). [Análise das vendas das lojas de vestuário da área central de Francisco Beltrão: Um estudo após a implantação do PIX / Analysis of sales in clothing stores in the central area of Francisco Beltrão: A study after the implementation of the PIX](#). *Brazilian Journal of Development*, 8(2), 12007–12024.
- (Citado 2 vezes nas páginas 19 e 76.)
- Azevedo, C. d. S., Gonçalves, R. F., Gava, V. L., e Spinola, M. d. M. (2021, April). [A Benford's law based methodology for fraud detection in social welfare programs: Bolsa Familia analysis](#). *Physica A: Statistical Mechanics and its Applications*, 567(1), 125626. (Online first.)
- (Citado 11 vezes nas páginas 22, 23, 24, 25, 34, 48, 53, 59, 62, 68, e 76.)
- Badal-Valero, E., Alvarez-Jareño, J. A., e Pavía, J. M. (2018, January). [Combining Benford's law and machine learning to detect money laundering. An actual Spanish court case](#). *Forensic Science International*, 282(1), 24–34.
- (Citado 8 vezes nas páginas 23, 25, 35, 48, 53, 64, 68, e 69.)
- Baesens, B., Höppner, S., e Verdonck, T. (2021, January). [Data Engineering for fraud detection](#). *Decision Support Systems*, 150(1), 113492. (Online first.)
- (Citado 6 vezes nas páginas 13, 20, 35, 53, 68, e 70.)
- Bajaj, Y., Kashiramka, S., e Singh, S. (2020, August). [Application of capital structure theories: A systematic review](#). *Journal of Advances in Management Research*, 18(2), 173–199.
- (Citado 1 vez na página 30.)
- Baker, C. R., Cohanier, B., e Leo, N. J. (2017, August). [Breakdowns in internal controls in bank trading information](#)

- systems: The case of the fraud at Société Générale. *International Journal of Accounting Information Systems*, 26(1), 20–31.
(Citado 3 vezes nas páginas 35, 53, e 59.)
- Banco Central. (2020a). *Thread: Você sabia que fazer um PIX será muito seguro?* (Publicado online.)
(Citado 1 vez na página 19.)
- Banco Central. (2020b). *O que é PIX?* (Publicação no site do Banco Central.)
(Citado 3 vezes nas páginas 2, 18, e 76.)
- Barboza, F., Kimura, H., Sobreiro, V. A., e Basso, L. F. C. (2016). *Credit risk: From a systematic literature review to future directions*. *Corporate Ownership and Control*, 13(3), 326–346.
(Citado 1 vez na página 5.)
- Barroso, M., e Laborda, J. (2022). *Digital transformation and the emergence of the Fintech sector: Systematic literature review*. *Digital Business*, 2(2), 100028. (Online first.)
(Citado 4 vezes nas páginas 3, 4, 5, e 13.)
- Bayrak, A. E., McComb, C., Cagan, J., e Kotovsky, K. (2021, May). *A strategic decision-making architecture toward hybrid teams for dynamic competitive problems*. *Decision Support Systems*, 144(1), 113490. (Online first.)
(Citado 1 vez na página 69.)
- Begenau, J., Farboodi, M., e Veldkamp, L. (2018, August). *Big Data in finance and the growth of large firms*. *Journal of Monetary Economics*, 97(1), 71–87.
(Citado 2 vezes nas páginas 12 e 13.)
- Belle, R. V., Baesens, B., e Weerdt, J. D. (2022, September). *CATCHM: A novel network-based credit card fraud detection method using node representation learning*. *Decision Support Systems*, 163(1), 113866. (Online first.)
(Citado 2 vezes nas páginas 35 e 53.)
- Belli, L., e Zingales, N. (2022, November). *Data protection and artificial intelligence inequalities and regulations in Latin America*. *Computer Law & Security Review*, 47(1), 105761. (Online first.)
(Citado 1 vez na página 11.)
- Benford, F. (1938, March). *The law of anomalous numbers*. *Proceedings of the American Philosophical Society*, 78(4), 551–572.
(Citado 1 vez na página 23.)
- Bhattacharya, S., Xu, D., e Kumar, K. (2011, February). *An ANN-based auditor decision support system using Benford's law*. *Decision Support Systems*, 50(3), 576–584.
(Citado 1 vez na página 68.)
- Bollaert, H., de Silanes, F. L., e Schwienbacher, A. (2021, June). *Fintech and access to finance*. *Journal of Corporate Finance*, 68(1), 101941. (Online first.)
(Citado 1 vez na página 2.)
- Borgersen, T.-A., e King, R. M. (2014, September). *Structural origins of debt-sustainability in mature and transition economies: Domar, Balassa-Samuelson and Maastricht*. *Structural Change and Economic Dynamics*, 30, 101–119.
(Citado 1 vez na página 46.)
- Bose, I., Piramuthu, S., e Shaw, M. J. (2011, February). *Quantitative methods for detection of financial fraud*. *Decision Support Systems*, 50(3), 557–558.
(Citado 1 vez na página 60.)
- Boubaker, S., Derouiche, I., e Lasfer, M. (2015, December). *Geographic location, excess control rights, and cash holdings*. *International Review of Financial Analysis*, 42(1), 24–37.
(Citado 1 vez na página 47.)
- Boubakri, N., Guedhami, O., e Saffar, W. (2016, June). *Geographic location, foreign ownership, and cost of equity capital: Evidence from privatization*. *Journal of Corporate Finance*, 38(1), 363–381.
(Citado 1 vez na página 47.)
- Bouzguenda, K. (2018, April). *Emotional intelligence and financial decision making: Are we talking about a paradigmatic shift or a change in practices?* *Research in International Business and Finance*, 44(1), 273–284.
(Citado 1 vez na página 31.)
- Byers, J., Popova, I., e Simkins, B. (2021, March). *Robust estimation of conditional risk measures using machine learning algorithm for commodity futures prices in the presence of outliers*. *Journal of Commodity Markets*, 24(1), 100174. (Online first.)
(Citado 3 vezes nas páginas 5, 35, e 53.)
- Cabral, C., e Dhar, R. L. (2019a, June). *Ecotourism research in India: From an integrative literature review to a future research framework*. *Journal of Ecotourism*, 19(1), 23–49.
(Citado 1 vez na página 30.)
- Cabral, C., e Dhar, R. L. (2019b, June). *Skill development research in India: A systematic literature review and future research agenda*. *Benchmarking: An International Journal*, 26(7), 2242–2266.

- (Citado 1 vez na página 5.)
- Canhoto, A. I. (2021, July). [Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective](#). *Journal of Business Research*, 131(1), 441–452.
- (Citado 5 vezes nas páginas 35, 53, 61, 66, e 68.)
- Carmona, P., Climent, F., e Momparler, A. (2019, May). [Predicting failure in the U.S. banking sector: An extreme gradient boosting approach](#). *International Review of Economics & Finance*, 61(1), 304–323.
- (Citado 4 vezes nas páginas 35, 53, 61, e 68.)
- Carneiro, N., Figueira, G., e Costa, M. (2017, March). [A Data Mining based system for credit-card fraud detection in e-tail](#). *Decision Support Systems*, 95(1), 91–101.
- (Citado 8 vezes nas páginas 1, 20, 21, 35, 53, 64, 68, e 70.)
- Carta, S., Fenu, G., Recupero, D. R., e Saia, R. (2019, June). [Fraud detection for e-commerce transactions by employing a prudential Multiple Consensus model](#). *Journal of Information Security and Applications*, 46(1), 13–22.
- (Citado 4 vezes nas páginas 36, 53, 60, e 64.)
- Chandola, V., Banerjee, A., e Kumar, V. (2009, July). [Anomaly detection](#). *ACM Computing Surveys*, 41(3), 1–58.
- (Citado 1 vez na página 21.)
- Chang, V., Baudier, P., Zhang, H., Xu, Q., Zhang, J., e Arami, M. (2020, September). [How Blockchain can impact financial services – The overview, challenges and recommendations from expert interviewees](#). *Technological Forecasting and Social Change*, 158(1), 120166. (Online first.)
- (Citado 3 vezes nas páginas 14, 15, e 16.)
- Charakopoulos, A., e Karakasidis, T. (2022, October). [Backward Degree a new index for online and offline change point detection based on complex network analysis](#). *Physica A: Statistical Mechanics and its Applications*, 604(1), 127929. (Online first)
- (Citado 2 vezes nas páginas 36 e 53.)
- Chatterjee, A., e Ahmed, B. S. (2022, August). [IoT anomaly detection methods and applications: A survey](#). *Internet of Things*, 19(1), 100568. (Online first.)
- (Citado 2 vezes nas páginas 20 e 75.)
- Chaturvedi, A. R., Hutchinson, G. K., e Nazareth, D. L. (1993, September). [Supporting complex real-time decision making through machine learning](#). *Decision Support Systems*, 10(2), 213–233.
- (Citado 2 vezes nas páginas 26 e 65.)
- Chaudhry, S., Li, H., Xu, L., e Zhang, H. (2007, January). [Decision support systems in emerging economies](#). *Decision Support Systems*, 42(4), 1987–1988.
- (Citado 1 vez na página 55.)
- Chen, K., Yadav, A., Khan, A., e Zhu, K. (2020). [Credit fraud detection based on hybrid Credit Scoring Model](#). *Procedia Computer Science*, 167(1), 2–8.
- (Citado 7 vezes nas páginas 36, 47, 49, 53, 60, 65, e 70.)
- Chen, L., Jia, N., Zhao, H., Kang, Y., Deng, J., e Ma, S. (2022, December). [Refined analysis and a hierarchical multi-task learning approach for loan fraud detection](#). *Journal of Management Science and Engineering*, 7(4), 589–607.
- (Citado 2 vezes nas páginas 36 e 53.)
- Chen, M., Kanade, T., Pomerleau, D., e Rowley, H. A. (1999, January). [Anomaly detection through registration](#). *Pattern Recognition*, 32(1), 113–128.
- (Citado 1 vez na página 4.)
- Chen, T.-K. (2016, July). [Does geography matter in a geographically small and culturally homogeneous country? Firm location and corporate credit risk](#). *International Review of Economics & Finance*, 44(1), 323–348.
- (Citado 1 vez na página 47.)
- Chen, X., Teng, L., e Chen, W. (2022, July). [How does FinTech affect the development of the digital economy? Evidence from China](#). *The North American Journal of Economics and Finance*, 61(1), 101697. (Online first.)
- (Citado 1 vez na página 8.)
- Chen, Y., e Xu, J. (2023, March). [Digital transformation and firm cost stickiness: Evidence from China](#). *Finance Research Letters*, 52(1), 103510. (Online first.)
- (Citado 1 vez na página 8.)
- Cherchi, E., e Guevara, C. A. (2012, February). [A Monte Carlo experiment to analyze the curse of dimensionality in estimating random coefficients models with a full variance-covariance matrix](#). *Transportation Research Part B: Methodological*, 46(2), 321–332.
- (Citado 1 vez na página 26.)
- Cho, T.-Y., e Chen, Y.-S. (2021, July). [The impact of financial technology on China's banking industry: An application of the metafrontier cost Malmquist productivity index](#). *The North American Journal of Economics and Finance*, 57(1), 101414. (Online first.)
- (Citado 2 vezes nas páginas 2 e 14.)

- Chu, X., Wang, R., Lin, Y., e Li, Y. (2022, August). [Building trust in online trade-in programs with a blockchain-enabled system](#). *Transportation Research Part E: Logistics and Transportation Review*, 164(1), 102833. (Online first.)
(Citado 2 vezes nas páginas 36 e 53.)
- Ciaccio, E. J., Wan, E. Y., Saluja, D. S., Acharya, U. R., Peters, N. S., e Garan, H. (2019, September). [Addressing challenges of quantitative methodologies and event interpretation in the study of atrial fibrillation](#). *Computer Methods and Programs in Biomedicine*, 178(1), 113–122.
(Citado 1 vez na página 48.)
- Climent, F., Momparler, A., e Carmona, P. (2019, August). [Anticipating bank distress in the Eurozone: An extreme gradient boosting approach](#). *Journal of Business Research*, 101(1), 885–896.
(Citado 6 vezes nas páginas 1, 20, 36, 53, 61, e 71.)
- Coimbra, L. G. A. M., Ikeno, F. H., Ikeda, E. A., e Rodrigues, G. M. (2022, April). [Simplificando Pagamentos-Digitais: prevenção à lavagemde dinheiro e antifraude](#). *Revista do Laboratório de InovaçõesFinanceiras e Tecnológicas*, 4(4), 157–183.
(Citado 3 vezes nas páginas 17, 18, e 19.)
- Conti, P., e Naldi, M. (2008, December). [Detection of anomalous bids in procurement auctions](#). *Decision Support Systems*, 46(1), 420–428.
(Citado 1 vez na página 73.)
- Cruz, A. F. d., Kimura, H., e Sobreiro, V. A. (2019, January). [What Do We Know About Corporate Cash Holdings? A Systematic Analysis](#). *Journal of Corporate Accounting & Finance*, 30(1), 77–143.
(Citado 1 vez na página 30.)
- Cumming, D., Groh, A. P., e Johan, S. (2018, May). [Same rules, different enforcement: Market abuse in Europe](#). *Journal of International Financial Markets, Institutions and Money*, 54(1), 130–151.
(Citado 1 vez na página 32.)
- Damaso, O. (2021, May). [Open banking](#). Disponível em https://www.bcb.gov.br/conteudo/home-ptbr/TextosApresentacoes/Open_Banking-2021-05-27-AWS-V03.pdf (Relatório publicado online.)
(Citado 2 vezes nas páginas 8 e 18.)
- Daragmeh, A., Lentner, C., e Sági, J. (2021, December). [FinTech payments in the era of COVID-19: Factors influencing behavioral intentions of “Generation X” in Hungary to use mobile payment](#). *Journal of Behavioral and Experimental Finance*, 32(1), 100574. (Online first.)
(Citado 1 vez na página 2.)
- Dehbi, S., Lamrani, H. C., Belgnaoui, T., e Lafou, T. (2022). [Big Data Analytics and Management control](#). *Procedia Computer Science*, 203(1), 438–443.
(Citado 1 vez na página 13.)
- Delen, D. (2010, November). [A comparative analysis of machine learning techniques for student retention management](#). *Decision Support Systems*, 49(4), 498–506.
(Citado 1 vez na página 67.)
- Dinh, D. V., Powell, R. J., e Vo, D. H. (2021, June). [Forecasting corporate financial distress in the Southeast Asian countries: A market-based approach](#). *Journal of Asian Economics*, 74(1), 101293. (Online first.)
(Citado 2 vezes nas páginas 36 e 53.)
- Druică, E., Oancea, B., e Vălsan, C. (2018, December). [Benford’s law and the limits of digit analysis](#). *International Journal of Accounting Information Systems*, 31(1), 75–82.
(Citado 8 vezes nas páginas 23, 25, 36, 48, 53, 59, 68, e 71.)
- Duftschnid, G., e Miksch, S. (2001, April). [Knowledge-based verification of clinical guidelines by detection of anomalies](#). *Artificial Intelligence in Medicine*, 22(1), 23–41.
(Citado 1 vez na página 4.)
- Elisa, U., Annalisa, C., Armin, S., Massimo, C., e Alexander, B. (2021, July). [Digital technologies and the changing entrepreneurial finance landscape: Status quo and perspectives for future research](#). *Technological Forecasting and Social Change*, 168(1), 120768. (Online first.)
(Citado 2 vezes nas páginas 9 e 10.)
- Elsaify, M., e Hasan, S. (2021, October). [Data exchanges among firms](#). *Digital Business*, 1(2), 100010. (Online first.)
(Citado 2 vezes nas páginas 3 e 31.)
- Febraban. (2021). [Pesquisa FEBRABAN deTecnologia Bancária 2021](#). (Relatório publicado online.)
(Citado 1 vez na página 19.)
- Fernández-Martínez, J. L., e Fernández-Muñiz, Z. (2020, May). [The curse of dimensionality in inverse problems](#). *Journal of Computational and Applied Mathematics*, 369(1), 112571. (Online first.)
(Citado 1 vez na página 27.)
- Fernandez-Vidal, J., Perotti, F. A., Gonzalez, R., e Gasco, J. (2022, November). [Managing digital transformation: The view from the top](#). *Journal of Business Research*, 152(1), 29–41.

- (Citado 1 vez na página 9.)
- Ferraz, D., Falguera, F. P. S., Mariano, E. B., e Hartmann, D. (2021, January). [Linking Economic Complexity, Diversification, and Industrial Policy with Sustainable Development: A Structured Literature Review](#). *Sustainability*, 13(3), 1265.
- (Citado 1 vez na página 5.)
- Ferreira, A. R. (2022, September). [Arranjo PIX: regulação e concorrência em pagamentos digitais](#). *Revista da Procuradoria-Geral do Banco Central*, 16(1), 100–113.
- (Citado 4 vezes nas páginas 2, 9, 18, e 76.)
- Ferreira, M. C. R. d. C., Sobreiro, V. A., Kimura, H., e Barboza, F. L. d. M. (2016, April). [A systematic review of literature about finance and sustainability](#). *Journal of Sustainable Finance & Investment*, 6(2), 112–147.
- (Citado 3 vezes nas páginas 30, 46, e 47.)
- Fertier, A., Barthe-Delanoë, A.-M., Montarnal, A., Truptil, S., e Bénaben, F. (2020, June). [A new emergency decision support system: The automatic interpretation and contextualisation of events to model a crisis situation in real-time](#). *Decision Support Systems*, 133(1), 113260. (Online first.)
- (Citado 3 vezes nas páginas 25, 65, e 74.)
- Feuerriegel, S., e Gordon, J. (2018, August). [Long-term stock index forecasting based on text mining of regulatory disclosures](#). *Decision Support Systems*, 112(1), 88–97.
- (Citado 1 vez na página 27.)
- Florack, A., Keller, J., e Palcu, J. (2013, October). [Regulatory focus in economic contexts](#). *Journal of Economic Psychology*, 38(1), 127–137.
- (Citado 1 vez na página 47.)
- Forough, J., e Momtazi, S. (2021, February). [Ensemble of deep sequential models for credit card fraud detection](#). *Applied Soft Computing*, 99(1), 106883. (Online first.)
- (Citado 8 vezes nas páginas 37, 53, 58, 60, 63, 64, 65, e 68.)
- Frömmel, M., Han, X., Li, Y., e Vigne, S. A. (2021, June). [Low liquidity beta anomaly in China](#). *Emerging Markets Review*, 50(1), 100832. (Online first.)
- (Citado 1 vez na página 32.)
- Gan, L., Wang, H., e Yang, Z. (2020, April). [Machine learning solutions to challenges in finance: An application to the pricing of financial products](#). *Technological Forecasting and Social Change*, 153(1), 119928. (Online first.)
- (Citado 1 vez na página 21.)
- Ghasemaghaei, M., Ebrahimi, S., e Hassanein, K. (2018, March). [Data analytics competency for improving firm decision making performance](#). *The Journal of Strategic Information Systems*, 27(1), 101–113.
- (Citado 1 vez na página 27.)
- Ghoddusi, H., Creamer, G. G., e Rafizadeh, N. (2019, June). [Machine learning in energy economics and finance: A review](#). *Energy Economics*, 81(1), 709–727.
- (Citado 1 vez na página 21.)
- Gianini, G., Fossi, L. G., Mio, C., Caelen, O., Brunie, L., e Damiani, E. (2020, January). [Managing a pool of rules for credit card fraud detection by a Game Theory based approach](#). *Future Generation Computer Systems*, 102(1), 549–561.
- (Citado 6 vezes nas páginas 1, 2, 37, 53, 63, e 69.)
- González-Sánchez, M. (2021, January). [Is there a relationship between the time scaling property of asset returns and the outliers? Evidence from international financial markets](#). *Finance Research Letters*, 38(1), 101510. (Online first.)
- (Citado 4 vezes nas páginas 37, 54, 55, e 59.)
- Goodell, J. W., Kumar, S., Lim, W. M., e Pattnaik, D. (2021, December). [Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis](#). *Journal of Behavioral and Experimental Finance*, 32(1), 100577. (Online first.)
- (Citado 2 vezes nas páginas 10 e 21.)
- Grüne, L. (2021). [Overcoming the curse of dimensionality for approximating Lyapunov functions with deep neural networks under a small-gain condition](#). *IFAC-PapersOnLine*, 54(9), 317–322.
- (Citado 1 vez na página 26.)
- GSMA. (2017). [State of the Industry Report on Mobile Money](#). (Report published by GSMA Mobile Money)
- (Citado 1 vez na página 2.)
- Gu, Z., Lin, D., e Wu, J. (2022, October). [On-chain analysis-based detection of abnormal transaction amount on cryptocurrency exchanges](#). *Physica A: Statistical Mechanics and its Applications*, 604(1), 127799. (Online first.)
- (Citado 2 vezes nas páginas 37 e 54.)
- Guo, Y., e Liang, C. (2016, December). [Blockchain application and outlook in the banking industry](#). *Financial Innovation*, 2(1), 1–12.
- (Citado 1 vez na página 3.)

- Gupta, M., Gao, J., Aggarwal, C. C., e Han, J. (2014, September). [Outlier Detection for Temporal Data: A Survey](#). *IEEE Transactions on Knowledge and Data Engineering*, 26(9), 2250–2267. (Citado 1 vez na página 74.)
- Habeeb, R. A. A., Nasaruddin, F., Gani, A., Hashem, I. A. T., Ahmed, E., e Imran, M. (2019, April). [Real-time big data processing for anomaly detection: A Survey](#). *International Journal of Information Management*, 45(1), 289–307. (Citado 1 vez na página 74.)
- Hanafizadeh, P., e Kim, S. (2020, September). [Digital Business: A new forum for discussion and debate on digital business model and digital transformation](#). *Digital Business*, 1(1), 100006. (Citado 3 vezes nas páginas 1, 2, e 31.)
- Hanafizadeh, P., e Marjaie, S. (2021, October). [Exploring banking business model types: A cognitive view](#). *Digital Business*, 1(2), 100012. (Citado 1 vez na página 3.)
- Handley, L. (2019, January). [Nearly three quarters of the world will use just their smartphones to access the internet by 2025](#). (Report published by World Advertising Research Center.) (Citado 1 vez na página 2.)
- Henrique, B. M., Sobreiro, V. A., e Kimura, H. (2019, June). [Literature review: Machine learning techniques applied to financial market prediction](#). *Expert Systems with Applications*, 124(1), 226–251. (Citado 1 vez na página 29.)
- Henriques, I. C., Sobreiro, V. A., e Kimura, H. (2018, May). [Science and technology park: Future challenges](#). *Technology in Society*, 53(1), 144–160. (Citado 3 vezes nas páginas 30, 46, e 47.)
- Henriques, I. C., Sobreiro, V. A., Kimura, H., e Mariano, E. B. (2020, December). [Two-stage DEA in banks: Terminological controversies and future directions](#). *Expert Systems with Applications*, 161(1), 113632. (Online first.) (Citado 1 vez na página 30.)
- Herrera, R., Climent, F., Carmona, P., e Momparler, A. (2022, March). [The manipulation of Euribor: An analysis with machine learning classification techniques](#). *Technological Forecasting and Social Change*, 176(1), 121466. (Online first.) (Citado 2 vezes nas páginas 37 e 54.)
- Hess, R. L., Rubin, R. S., e West, L. A. (2004, November). [Geographic information systems as a marketing information system technology](#). *Decision Support Systems*, 38(2), 197–212. (Citado 1 vez na página 57.)
- Hilal, W., Gadsden, S. A., e Yawney, J. (2022, May). [Financial Fraud: A review of anomaly detection techniques and recent advances](#). *Expert Systems with Applications*, 193(1), 116429. (Online first.) (Citado 2 vezes nas páginas 20 e 21.)
- Hohberger, J., e Wilden, R. (2022, June). [Geographic diversity of knowledge inputs: The importance of aligning locations of knowledge inputs and inventors](#). *Journal of Business Research*, 145(1), 705–719. (Citado 1 vez na página 47.)
- Holsapple, C. W., Johnson, L. E., Manakyan, H., e Tanner, J. (1995, August). [An empirical assessment and categorization of journals relevant to DSS research](#). *Decision Support Systems*, 14(4), 359–367. (Citado 1 vez na página 70.)
- Hooi, B., Shin, K., Song, H. A., Beutel, A., Shah, N., e Faloutsos, C. (2017, November). [Graph-Based Fraud Detection in the Face of Camouflage](#). *ACM Transactions on Knowledge Discovery from Data*, 11(4), 1–26. (Citado 1 vez na página 48.)
- Horita, F. E., Albuquerque, J. P., Marchezini, V., e Mendiondo, E. M. (2017, May). [Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil](#). *Decision Support Systems*, 97(1), 12–22. (Citado 1 vez na página 25.)
- Hosseini, S., e Ivanov, D. (2020, December). [Bayesian networks for supply chain risk, resilience and ripple effect analysis: A literature review](#). *Expert Systems with Applications*, 161(1), 113649. (Online first.) (Citado 1 vez na página 29.)
- Hou, X., Wang, T., e Ma, C. (2021, September). [Economic policy uncertainty and corporate fraud](#). *Economic Analysis and Policy*, 71(1), 97–110. (Citado 1 vez na página 73.)
- Houdek, P. (2020, October). [Detecting fraud beyond cooked books](#). *Organizational Dynamics*, 49(4), 100734. (Online first.) (Citado 7 vezes nas páginas 4, 5, 29, 37, 49, 54, e 60.)
- Huang, Y., e Kou, G. (2014, August). [A kernel entropy manifold learning approach for financial data analysis](#). *Decision Support Systems*, 64(1), 31–42.

- (Citado 1 vez na página 25.)
- Höppner, S., Baesens, B., Verbeke, W., e Verdonck, T. (2022, February). [Instance-dependent cost-sensitive learning for detecting transfer fraud](#). *European Journal of Operational Research*, 297(1), 291–300.
- (Citado 2 vezes nas páginas 3 e 31.)
- Ince, H., e Trafalis, T. B. (2006, November). [A hybrid model for exchange rate prediction](#). *Decision Support Systems*, 42(2), 1054–1062.
- (Citado 1 vez na página 69.)
- Jabbour, C. J. C. (2013, May). [Environmental training in organisations: From a literature review to a framework for future research](#). *Resources, Conservation and Recycling*, 74(1), 144–155.
- (Citado 8 vezes nas páginas xi, xiii, xv, 5, 30, 43, 46, e 73.)
- Jabbour, C. J. C., Sobreiro, V. A., de Sousa Jabbour, A. B. L., de Souza Campos, L. M., Mariano, E. B., e Renwick, D. W. S. (2017, June). [An analysis of the literature on humanitarian logistics and supply chain management: Paving the way for future studies](#). *Annals of Operations Research*, 283(1-2), 289–307.
- (Citado 1 vez na página 30.)
- James, N., Menzies, M., e Chan, J. (2021, March). [Changes to the extreme and erratic behaviour of cryptocurrencies during COVID-19](#). *Physica A: Statistical Mechanics and its Applications*, 565(1), 125581. (Online first.)
- (Citado 2 vezes nas páginas 37 e 54.)
- Johnson, P. C., Laurell, C., Ots, M., e Sandström, C. (2022, June). [Digital innovation and the effects of artificial intelligence on firms' research and development - Automation or augmentation, exploration or exploitation?](#) *Technological Forecasting and Social Change*, 179(1), 121636. (Online first.)
- (Citado 3 vezes nas páginas 9, 10, e 11.)
- Jones, B. H., e Chin, A. G. (2015, October). [On the efficacy of smartphone security: A critical analysis of modifications in business students' practices over time](#). *International Journal of Information Management*, 35(5), 561–571.
- (Citado 1 vez na página 3.)
- Jung, S. H., e Jeong, Y. J. (2020, November). [Twitter data analytical methodology development for prediction of start-up firms' social media marketing level](#). *Technology in Society*, 63(1), 101409. (Online first)
- (Citado 1 vez na página 31.)
- Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., e Caelen, O. (2018, June). [Sequence classification for credit-card fraud detection](#). *Expert Systems with Applications*, 100(1), 234–245.
- (Citado 6 vezes nas páginas 21, 37, 54, 58, 63, e 70.)
- Kankanhalli, A., Tan, B. C., Wei, K.-K., e Holmes, M. C. (2004, November). [Cross-cultural differences and information systems developer values](#). *Decision Support Systems*, 38(2), 183–195.
- (Citado 1 vez na página 58.)
- Karki, B. R., e Porras, J. (2021, October). [Digitalization for sustainable maintenance services: A systematic literature review](#). *Digital Business*, 1(2), 100011.
- (Citado 3 vezes nas páginas 1, 73, e 75.)
- Kaur, R., e Singh, S. (2016, July). [A survey of Data Mining and social network analysis based anomaly detection techniques](#). *Egyptian Informatics Journal*, 17(2), 199–216.
- (Citado 1 vez na página 4.)
- Kedar, Y., e Hon, G. (2017, February). [Natures and Laws: The making of the concept of law of nature - Robert Grosseteste \(c. 1168-1253\) and Roger Bacon \(1214/1220-1292\)](#). *Studies in History and Philosophy of Science Part A*, 61(1), 21–31.
- (Citado 1 vez na página 22.)
- Khan, A. I., e Al-Badi, A. (2020). [Open source machine learning frameworks for industrial Internet of Things](#). *Procedia Computer Science*, 170(1), 571–577.
- (Citado 1 vez na página 67.)
- Khan, M. A., Gu, L., Khan, M. A., e Meyer, N. (2021, March). [The effects of national culture on financial sector development: Evidence from emerging and developing economies](#). *Borsa Istanbul Review*, 22(1), 103–112.
- (Citado 1 vez na página 58.)
- Khudzari, J. M., Kurian, J., Tartakovsky, B., e Raghavan, G. (2018, August). [Bibliometric analysis of global research trends on microbial fuel cells using Scopus database](#). *Biochemical Engineering Journal*, 136(1), 51–60.
- (Citado 1 vez na página 30.)
- Kompas, T., e Ha, P. V. (2019, August). [The curse of dimensionality resolved: The effects of climate change and trade barriers in large dimensional modelling](#). *Economic Modelling*, 80(1), 103–110.
- (Citado 1 vez na página 26.)
- Konopik, J., Jahn, C., Schuster, T., Hoßbach, N., e Pflaum, A. (2022). [Mastering the digital transformation through organizational capabilities: A conceptual framework](#). *Digital Business*, 2(2), 100019. (Online first.)
- (Citado 3 vezes nas páginas 3, 5, e 74.)
- Koren, O., Koren, M., e Peretz, O. (2023, January). [A procedure for anomaly detection and analysis](#). *Engineering*

- Applications of Artificial Intelligence*, 117(1), 105503. (Online first.)
(Citado 1 vez na página 20.)
- Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K., e Kou, S. (2021, January). [Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection](#). *Decision Support Systems*, 140(1), 113429. (Online first.)
(Citado 2 vezes nas páginas 38 e 54.)
- Koyuncugil, A. S., e OZgulbas, N. (2012, May). [Financial early warning system model and data mining application for risk detection](#). *Expert Systems with Applications*, 39(6), 6238–6253.
(Citado 1 vez na página 4.)
- Kristóf, T., e Virág, M. (2022, October). [EU-27 bank failure prediction with C5.0 decision trees and deep learning neural networks](#). *Research in International Business and Finance*, 61(1), 101644. (Online first.)
(Citado 2 vezes nas páginas 38 e 54.)
- Kundu, D., Pal, R. K., e Mandal, D. P. (2020, February). [Preference enhanced hybrid expertise retrieval system in community question answering services](#). *Decision Support Systems*, 129(1), 113164. (Online first.)
(Citado 1 vez na página 69.)
- Kuzey, C., Uyar, A., e Delen, D. (2014, March). [The impact of multinationality on firm value: A comparative analysis of machine learning techniques](#). *Decision Support Systems*, 59(1), 127–142.
(Citado 1 vez na página 67.)
- Königstorfer, F., e Thalmann, S. (2020, September). [Applications of Artificial Intelligence in commercial banks - A research agenda for behavioral finance](#). *Journal of Behavioral and Experimental Finance*, 27(1), 100352. (Online first.)
(Citado 2 vezes nas páginas 9 e 10.)
- Lage Junior, M., e Godinho Filho, M. (2010, May). [Variations of the Kanban System: Literature review and classification](#). *International Journal of Production Economics*, 125(1), 13–21.
(Citado 4 vezes nas páginas 5, 30, 43, e 73.)
- Lavergne, P., e Patilea, V. (2008, March). [Breaking the curse of dimensionality in nonparametric testing](#). *Journal of Econometrics*, 143(1), 103–122.
(Citado 2 vezes nas páginas 26 e 27.)
- Learning, E. (2020). [How does a transaction get into the blockchain?](#) (Publicado online.)
(Citado 1 vez na página 15.)
- Lee, J. K. Y., Gholami, H., Saman, M. Z. M., Ngadiman, N. H. A. B., Zakuan, N., Mahmood, S., e Omain, S. Z. (2021). [Sustainability-Oriented Application of Value Stream Mapping: A review and classification](#). *IEEE Access*, 9(1), 68414–68434.
(Citado 1 vez na página 30.)
- Lee, K. C., Han, I., e Kwon, Y. (1996, September). [Hybrid neural network models for bankruptcy predictions](#). *Decision Support Systems*, 18(1), 63–72.
(Citado 1 vez na página 69.)
- Lenine, F. (2017). [Como selecionar atributos para resolver a maldição da dimensionalidade](#). Disponível em <https://medium.com/@fabiolenine/como-selecionar-atributos-para-resolver-a-maldição-da-dimensionalidade-5c810bc8449f> (Artigo online.)
(Citado 1 vez na página 26.)
- León, C. (2020). [Detecting anomalous payments networks: A dimensionality-reduction approach](#). *Latin American Journal of Central Banking*, 1(1–4), 100001. (Online first.)
(Citado 3 vezes nas páginas 3, 38, e 54.)
- León, C., Barucca, P., Acero, O., Gage, G., e Ortega, F. (2020). [Pattern recognition of financial institutions' payment behavior](#). *Latin American Journal of Central Banking*, 1(1–4), 100011. (Online first.)
(Citado 5 vezes nas páginas 22, 38, 54, 67, e 68.)
- Li, D., Wang, Q., Wang, J., e Yao, Y. (2008). [Mitigation of curse of dimensionality in dynamic programming](#). *IFAC Proceedings Volumes*, 41(2), 7778–7783.
(Citado 1 vez na página 75.)
- Li, J., Herdem, M. S., Nathwani, J., e Wen, J. Z. (2023, January). [Methods and applications for Artificial Intelligence, Big Data, Internet of Things, and Blockchain in smart energy management](#). *Energy and AI*, 11(1), 100208. (Online first.)
(Citado 3 vezes nas páginas 11, 14, e 16.)
- Li, K., Gao, X., Fu, S., Diao, X., Ye, P., Xue, B., ... Huang, Z. (2022, November). [Robust outlier detection based on the changing rate of directed density ratio](#). *Expert Systems with Applications*, 207(1), 117988. (Online first.)
(Citado 1 vez na página 20.)
- Liang, T.-P., Chandler, J. S., e Han, I. (1990, January). [Integrating statistical and inductive learning methods for knowledge acquisition](#). *Expert Systems with Applications*, 1(4), 391–401.
(Citado 1 vez na página 69.)

- Liao, C.-H., Guan, X.-Q., Cheng, J.-H., e Yuan, S.-M. (2022, October). [Blockchain-based identity management and access control framework for open banking ecosystem](#). *Future Generation Computer Systems*, 135(1), 450–466.
(Citado 1 vez na página 17.)
- Lien, D. (2017, January). [Business Finance and Enterprise Management in the Era of Big Data: An introduction](#). *The North American Journal of Economics and Finance*, 39(1), 143–144.
(Citado 1 vez na página 12.)
- Lima, M. A. F. (2020, October). [O PIX diante dos crimes patrimoniais](#). (Publicado online.)
(Citado 2 vezes nas páginas 19 e 76.)
- Lin, F., Lin, L.-J., Yeh, C.-C., e Wang, T.-S. (2018, May). [Does the board of directors as fat cats exert more earnings management? Evidence from Benford's law](#). *The Quarterly Review of Economics and Finance*, 68(1), 158–170.
(Citado 5 vezes nas páginas 23, 24, 38, 54, e 59.)
- Liu, H., Lin, F., He, J., e Cai, Y. (2010, December). [New approach for the sequential pattern mining of high-dimensional sequence databases](#). *Decision Support Systems*, 50(1), 270–280.
(Citado 2 vezes nas páginas 65 e 74.)
- Liu, Q., Wang, C., Zhang, P., e Zheng, K. (2021, November). [Detecting stock market manipulation via machine learning: Evidence from China Securities Regulatory Commission punishment cases](#). *International Review of Financial Analysis*, 78(1), 101887. (Online first.)
(Citado 2 vezes nas páginas 38 e 54.)
- Liu, S., Koster, S., e Chen, X. (2022, December). [Digital divide or dividend? The impact of digital finance on the migrants' entrepreneurship in less developed regions of China](#). *Cities*, 131(1), 103896. (Online first.)
(Citado 2 vezes nas páginas 9 e 10.)
- Lo, F.-Y., Rey-Martí, A., e Botella-Carrubi, D. (2020, July). [Research methods in business: Quantitative and qualitative comparative analysis](#). *Journal of Business Research*, 115(1), 221–224.
(Citado 1 vez na página 60.)
- López-Robles, J. R., Rodríguez-Salvador, M., Gamboa-Rosales, N. K., Ramirez-Rosales, S., e Cobo, M. J. (2019). [The last five years of Big Data Research in Economics, Econometrics and Finance: Identification and conceptual analysis](#). *Procedia Computer Science*, 162(1), 729–736.
(Citado 1 vez na página 13.)
- Lozano, M. G., Schreiber, J., e Brynielsson, J. (2017, July). [Tracking geographical locations using a geo-aware topic model for analyzing social media data](#). *Decision Support Systems*, 99(1), 18–29.
(Citado 1 vez na página 57.)
- Lucas, Y., Portier, P.-E., Laporte, L., He-Guelton, L., Caelen, O., Granitzer, M., e Calabretto, S. (2020, January). [Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs](#). *Future Generation Computer Systems*, 102(1), 393–402.
(Citado 3 vezes nas páginas 38, 54, e 64.)
- Ma, C., Zhang, R., Lin, X., e Chen, G. (2012, June). [DuoWave: Mitigating the curse of dimensionality for uncertain data](#). *Data & Knowledge Engineering*, 76-78(1), 16–38.
(Citado 1 vez na página 26.)
- Maass, W., e Storey, V. C. (2021, July). [Pairing conceptual modeling with machine learning](#). *Data & Knowledge Engineering*, 134(1), 101909. (Online first.)
(Citado 1 vez na página 60.)
- Mallinguh, E. B., e Zéman, Z. (2020, August). [Financial distress, prediction, and strategies by firms: A systematic review of literature](#). *Periodica Polytechnica Social and Management Sciences*, 28(2), 162–176.
(Citado 2 vezes nas páginas 5 e 74.)
- Mamede, S. d. P. N., e Malaquias, R. F. (2017, January). [Monday effect in Brazilian hedge funds with immediate redemption](#). *Research in International Business and Finance*, 39(1), 47–53.
(Citado 4 vezes nas páginas 38, 54, 59, e 71.)
- Mansfield-Devine, S. (2016, October). [Open banking: Opportunity and danger](#). *Computer Fraud & Security*, 2016(10), 8–13.
(Citado 2 vezes nas páginas 17 e 18.)
- Mariano, E. B., Sobreiro, V. A., e Rebelatto, D. A. N. (2015, July). [Human development and data envelopment analysis: A structured literature review](#). *Omega*, 54(1), 33–49.
(Citado 1 vez na página 30.)
- Marikyan, D., Papagiannidis, S., Rana, O. F., e Ranjan, R. (2022). [Blockchain: A business model innovation analysis](#). *Digital Business*, 2(2), 100033.
(Citado 4 vezes nas páginas 1, 73, 75, e 76.)
- Marthews, A., e Tucker, C. (2022, October). [What blockchain can and can't do: Applications to marketing and privacy](#). *International Journal of Research in Marketing*, 1(1), 1–5. (Online first.)
(Citado 1 vez na página 15.)

- Masudin, I., e Fernanda, F. W. (2019, November). [A review of literature on types, stages of recovery and humanitarian logistics operations in the tsunami and earthquake disaster in Indonesia](#). *IOP Conference Series: Materials Science and Engineering*, 674(1), 012043. (Online first.)
(Citado 1 vez na página 30.)
- McCarthy, J., Minsky, M. L., Rochester, N., e Shannon, C. E. (2006). [A proposal for the Dartmouth summer research project on Artificial Intelligence, August 31, 1955](#). *AI Magazine*, 27(4), 12–14.
(Citado 1 vez na página 10.)
- Metawa, N., Dogan, E., e Taskin, D. (2022, December). [Analyzing the nexus of green economy, clean and financial technology](#). *Economic Analysis and Policy*, 76(1), 385–396.
(Citado 1 vez na página 14.)
- Miller, G. J. (2022, December). [Stakeholder roles in artificial intelligence projects](#). *Project Leadership and Society*, 3(1), 100068. (Online first.)
(Citado 3 vezes nas páginas 10, 11, e 12.)
- Misra, S., Thakur, S., Ghosh, M., e Saha, S. K. (2020). [An autoencoder based model for detecting fraudulent credit card transaction](#). *Procedia Computer Science*, 167(1), 254–262.
(Citado 5 vezes nas páginas 3, 39, 54, 60, e 63.)
- Mittal, A. (2017, May). [Machine Learning Process And Scenarios](#). (Published online)
(Citado 1 vez na página 22.)
- Mukherjee, R., Gamble, R. F., e Parkinson, J. A. (1997, December). [Classifying and detecting anomalies in hybrid knowledge-based systems](#). *Decision Support Systems*, 21(4), 231–251.
(Citado 1 vez na página 73.)
- Muniyandi, A. P., Rajeswari, R., e Rajaram, R. (2012). [Network anomaly detection by cascading k-means clustering and C4.5 decision tree algorithm](#). *Procedia Engineering*, 30(1), 174–182.
(Citado 1 vez na página 4.)
- Mustafa, A. A., Lin, C.-Y., e Kakinaka, M. (2022, June). [Detecting market pattern changes: A machine learning approach](#). *Finance Research Letters*, 47(1), 102621. (Online first.)
(Citado 2 vezes nas páginas 39 e 54.)
- Nagler, T., e Czado, C. (2016, October). [Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas](#). *Journal of Multivariate Analysis*, 151(1), 69–89.
(Citado 1 vez na página 26.)
- Najem, R., Amr, M. F., Bahnasse, A., e Talea, M. (2022). [Artificial Intelligence for digital finance, axes and techniques](#). *Procedia Computer Science*, 203(1), 633–638.
(Citado 2 vezes nas páginas 10 e 12.)
- Nazário, R. T. F., Silva, J. L. e., Sobreiro, V. A., e Kimura, H. (2017, November). [A literature review of technical analysis on stock markets](#). *The Quarterly Review of Economics and Finance*, 66(1), 115–126.
(Citado 2 vezes nas páginas 30 e 47.)
- Newman, J., Mintrom, M., e O'Neill, D. (2022, February). [Digital technologies, artificial intelligence, and bureaucratic transformation](#). *Futures*, 136(1), 102886. (Online first.)
(Citado 1 vez na página 8.)
- Nigrini, M. J. (2012). *Benford's law*. Wiley.
(Citado 3 vezes nas páginas 22, 24, e 25.)
- Nikkel, B. (2020, June). [Fintech forensics: Criminal investigation and digital evidence in financial technologies](#). *Forensic Science International: Digital Investigation*, 33(1), 200908. (Online first.)
(Citado 5 vezes nas páginas 20, 39, 48, 54, e 61.)
- Nissim, D. (2022, November). [Big data, accounting information, and valuation](#). *The Journal of Finance and Data Science*, 8(1), 69–85.
(Citado 2 vezes nas páginas 39 e 54.)
- Niu, Y., Wen, W., Wang, S., e Li, S. (2023, January). [Breaking barriers to innovation: The power of digital transformation](#). *Finance Research Letters*, 51(1), 103457. (Online first.)
(Citado 2 vezes nas páginas 8 e 9.)
- Ojha, P. (2021, March). [The Upsurge of Graphical Abstracts: A Neuroscience Perspective](#). *Neuroscience*, 457(1), 218–219.
(Citado 1 vez na página 31.)
- Okoli, C. (2015). [A Guide to Conducting a Standalone Systematic Literature Review](#). *Communications of the Association for Information Systems*, 37(1), 879–910.
(Citado 1 vez na página 30.)
- Ouyang, L., Zhang, W., e Wang, F.-Y. (2022, December). [Intelligent contracts: Making smart contracts smart for blockchain intelligence](#). *Computers and Electrical Engineering*, 104(1), 108421. (Online first.)
(Citado 2 vezes nas páginas 14 e 16.)
- Pallathadka, H., Ramirez-Asis, E. H., Loli-Poma, T. P., Kaliyaperumal, K., Ventayen, R. J. M., e Naved, M. (2021,

- July). [Applications of Artificial Intelligence in business management, e-commerce and finance](#). *Materials Today: Proceedings*, 1(1), 1–4. (Online first.)
(Citado 1 vez na página 22.)
- Palmié, M., Wincent, J., Parida, V., e Caglar, U. (2020, February). [The evolution of the financial technology ecosystem: An introduction and agenda for future research on disruptive innovations in ecosystems](#). *Technological Forecasting and Social Change*, 151(1), 119779. (Online first.)
(Citado 1 vez na página 14.)
- Papík, M., e Papíková, L. (2022, June). [Detecting accounting fraud in companies reporting under US GAAP through data mining](#). *International Journal of Accounting Information Systems*, 45(1), 100559. (Online first.)
(Citado 2 vezes nas páginas 39 e 54.)
- Parnes, D. (2020, May). [Exploring economic anomalies in the S&P 500 index](#). *The Quarterly Review of Economics and Finance*, 76(1), 292–309.
(Citado 5 vezes nas páginas 3, 21, 39, 54, e 59.)
- Patel, R., Migliavacca, M., e Oriani, M. E. (2022, December). [Blockchain in banking and finance: A bibliometric review](#). *Research in International Business and Finance*, 62(1), 101718. (Online first.)
(Citado 2 vezes nas páginas 14 e 15.)
- Patil, S., Nemade, V., e Soni, P. K. (2018). [Predictive modelling for credit card fraud detection using data analytics](#). *Procedia Computer Science*, 132(1), 385–395.
(Citado 5 vezes nas páginas 39, 54, 61, 65, e 71.)
- Perdisci, R., Ariu, D., Fogla, P., Giacinto, G., e Lee, W. (2009, April). [McPAD: A multiple classifier system for accurate payload-based anomaly detection](#). *Computer Networks*, 53(6), 864–881.
(Citado 1 vez na página 4.)
- Pflüger, D., Peherstorfer, B., e Bungartz, H.-J. (2010, October). [Spatially adaptive sparse grids for high-dimensional data-driven problems](#). *Journal of Complexity*, 26(5), 508–522.
(Citado 1 vez na página 26.)
- Pham, N. T., Hoang, H. T., e Phan, Q. P. T. (2019, October). [Green human resource management: A comprehensive review and future research agenda](#). *International Journal of Manpower*, 41(7), 845–878.
(Citado 1 vez na página 30.)
- Pick, J. B., Turetken, O., Deokar, A. V., e Sarkar, A. (2017, July). [Location analytics and decision support: Reflections on recent advancements, a research framework, and the path ahead](#). *Decision Support Systems*, 99(1), 1–8.
(Citado 1 vez na página 57.)
- Picoto, W. N., e Pinto, I. (2021, January). [Cultural impact on mobile banking use – A multi-method approach](#). *Journal of Business Research*, 124(1), 620–628.
(Citado 1 vez na página 62.)
- Pietronudo, M. C., Croidieu, G., e Schiavone, F. (2022, September). [A solution looking for problems? A systematic literature review of the rationalizing influence of Artificial Intelligence on decision-making in innovation management](#). *Technological Forecasting and Social Change*, 182(1), 121828. (Online first.)
(Citado 2 vezes nas páginas 11 e 12.)
- Pinheiro, M. A. P., Seles, B. M. R. P., Fiorini, P. D. C., Jugend, D., Jabbour, A. B. L. S., Silva, H. M. R., e Latan, H. (2019, April). [The role of new product development in underpinning the circular economy](#). *Management Decision*, 57(4), 840–862.
(Citado 1 vez na página 30.)
- Pinto, S. O., e Sobreiro, V. A. (2022). [Literature review: Anomaly detection approaches on digital business financial systems](#). *Digital Business*, 2(2), 100038. (Online first.)
(Citado 4 vezes nas páginas 31, 32, 44, e 75.)
- Plc, R. (2021, February). [Annual report and financial statements 2020](#). (Report published by Elsevier®.)
(Citado 1 vez na página 31.)
- Pourhabibi, T., Ong, K.-L., Kam, B. H., e Boo, Y. L. (2020, June). [Fraud detection: A systematic literature review of graph-based anomaly detection approaches](#). *Decision Support Systems*, 133(1), 113303. (Online first.)
(Citado 15 vezes nas páginas 3, 4, 20, 21, 29, 39, 43, 44, 47, 48, 49, 54, 60, 64, e 66.)
- Prakash, R., Anoop, V., e Asharaf, S. (2022, November). [Blockchain technology for cybersecurity: A text mining literature analysis](#). *International Journal of Information Management Data Insights*, 2(2), 100112. (Online first.)
(Citado 2 vezes nas páginas 14 e 16.)
- Raisin, U. (2022). [Open banking explained](#). (Publicado online.)
(Citado 1 vez na página 17.)
- Raj, A., Agrahari, A., e Srivastava, S. K. (2020, September). [Do pressures foster sustainable public procurement? An empirical investigation comparing developed and developing economies](#). *Journal of Cleaner Production*, 266(1), 122055. (Online first.)

- (Citado 1 vez na página 55.)
Rajaeian, M. M., Cater-Steel, A., e Lane, M. (2017, October). [A systematic literature review and critical assessment of model-driven decision support for IT outsourcing](#). *Decision Support Systems*, 102(1), 42–56.
- (Citado 1 vez na página 29.)
R.B., A., e K.R., S. K. (2021, June). [Credit card fraud detection using Artificial Neural Network](#). *Global Transitions Proceedings*, 2(1), 35–41.
- (Citado 3 vezes nas páginas 40, 54, e 71.)
Rishabh. (2021, March). [Essentials of Digital Transformation in Financial Services](#). (Published online.)
- (Citado 1 vez na página 9.)
Rocha-Salazar, J.-d.-J., Segovia-Vargas, M.-J., e del Mar Camacho-Miñano, M. (2021, May). [Money laundering and terrorism financing detection using neural networks and an abnormality indicator](#). *Expert Systems with Applications*, 169(1), 114470. (Online first.)
- (Citado 2 vezes nas páginas 40 e 54.)
Rodrigues, A. R. D., Ferreira, F. A., Teixeira, F. J., e Zopounidis, C. (2022, April). [Artificial Intelligence, digital transformation and cybersecurity in the banking sector: A multi-stakeholder cognition-driven framework](#). *Research in International Business and Finance*, 60(1), 101616. (Online first.)
- (Citado 1 vez na página 10.)
Routledge, B. R. (2018, August). [Comments on: Big data in finance and the growth of large firms, by Juliane Begenau & Maryam Farboodi & Laura Veldkamp](#). *Journal of Monetary Economics*, 97(1), 88–90.
- (Citado 1 vez na página 13.)
Rtayli, N., e Enneya, N. (2020, December). [Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization](#). *Journal of Information Security and Applications*, 55(1), 102596. (Online first.)
- (Citado 5 vezes nas páginas 4, 40, 56, 60, e 70.)
Rubio, J., Barucca, P., Gage, G., Arroyo, J., e Morales-Resendiz, R. (2020). [Classifying payment patterns with Artificial Neural Networks: An autoencoder approach](#). *Latin American Journal of Central Banking*, 1(1–4), 100013. (Online first.)
- (Citado 7 vezes nas páginas 3, 20, 21, 40, 56, 63, e 65.)
Sá, A. G. d., Pereira, A. C., e Pappa, G. L. (2018, June). [A customized classification algorithm for credit card fraud detection](#). *Engineering Applications of Artificial Intelligence*, 72(1), 21–29.
- (Citado 5 vezes nas páginas 5, 40, 48, 56, e 64.)
Sabeti, L., e Heijmans, R. (2021, June). [Shallow or deep? Training an autoencoder to detect anomalous flows in a retail payment system](#). *Latin American Journal of Central Banking*, 2(2), 100031. (Online first.)
- (Citado 2 vezes nas páginas 40 e 56.)
Saeed, A., Riaz, H., e Baloch, M. S. (2022, September). [Does big data utilization improve firm legitimacy?](#) *Technological Forecasting and Social Change*, 182(1), 121847. (Online first.)
- (Citado 1 vez na página 13.)
Saleh, K., Boujarwah, A. A., e Al-Dallal, J. (2001, December). [Anomaly detection in concurrent Java programs using dynamic data flow analysis](#). *Information and Software Technology*, 43(15), 973–981.
- (Citado 1 vez na página 4.)
Salim, N., Rahman, M. N. A., e Wahab, D. A. (2019, February). [A systematic literature review of internal capabilities for enhancing eco-innovation performance of manufacturing firms](#). *Journal of Cleaner Production*, 209(1), 1445–1460.
- (Citado 1 vez na página 30.)
Salimi, A., Ziaii, M., Amiri, A., Zadeh, M. H., Karimpouli, S., e Moradkhani, M. (2018, April). [Using a feature subset selection method and support vector machine to address curse of dimensionality and redundancy in hyperion hyperspectral data classification](#). *The Egyptian Journal of Remote Sensing and Space Science*, 21(1), 27–36.
- (Citado 1 vez na página 25.)
Sami, F. (2022, December). [Optimize electric automation control using artificial intelligence \(AI\)](#). *Optik*, 271(1), 170085. (Online first.)
- (Citado 2 vezes nas páginas 11 e 12.)
Sánchez, M. A. (2022, August). [A multi-level perspective on financial technology transitions](#). *Technological Forecasting and Social Change*, 181(1), 121766. (Online first.)
- (Citado 1 vez na página 14.)
Sankarrao, L., Ghose, D. K., e Rathinsamy, M. (2021, November). [Predicting land-use change: Intercomparison of different hybrid machine learning models](#). *Environmental Modelling & Software*, 145(1), 105207. (Online first.)
- (Citado 1 vez na página 49.)
Sarlin, P., e Björk, K.-M. (2017, November). [Machine learning in finance - Guest editorial](#). *Neurocomputing*,

264(1), 1. (Online first.)

(Citado 1 vez na página 21.)

Savage, D., Zhang, X., Yu, X., Chou, P., e Wang, Q. (2014, October). [Anomaly detection in online social networks](#). *Social Networks*, 39(1), 62–70.

(Citado 1 vez na página 4.)

Serneels, S. (2022, September). [Detecting wash trading for nonfungible tokens](#). *Finance Research Letters*, 51(1), 103374. (Online first.)

(Citado 2 vezes nas páginas 40 e 56.)

Serrano-Silva, Y. O., Villuendas-Rey, Y., e Yáñez-Márquez, C. (2018, March). [Automatic feature weighting for improving financial Decision Support Systems](#). *Decision Support Systems*, 107(1), 78–87.

(Citado 1 vez na página 3.)

Services, A. W. (2022). [O que é Inteligência Artificial?](#) (Publicação online AWS.)

(Citado 1 vez na página 52.)

Setzer, T., e Bichler, M. (2013, May). [Using matrix approximation for high-dimensional discrete optimization problems: Server consolidation based on cyclic time-series data](#). *European Journal of Operational Research*, 227(1), 62–75.

(Citado 2 vezes nas páginas 25 e 26.)

Seuring, S. (2013, March). [A review of modeling approaches for sustainable supply chain management](#). *Decision Support Systems*, 54(4), 1513–1520.

(Citado 1 vez na página 29.)

Severino, M. K., e Peng, Y. (2021, September). [Machine learning algorithms for fraud prediction in property insurance: Empirical evidence using real-world microdata](#). *Machine Learning with Applications*, 5(1), 100074. (Online first.)

(Citado 5 vezes nas páginas 5, 40, 56, 64, e 71.)

Sgueglia, A., Sorbo, A. D., Visaggio, C. A., e Canfora, G. (2022, September). [A systematic literature review of IoT time series anomaly detection solutions](#). *Future Generation Computer Systems*, 134(1), 170–186.

(Citado 1 vez na página 21.)

Shaikh, A. A., Sharma, R., e Karjaluo, H. (2020, September). [Digital innovation & enterprise in the sharing economy: An action research agenda](#). *Digital Business*, 1(1), 100002. (Online first.)

(Citado 3 vezes nas páginas 1, 2, e 73.)

Shi, J., Ausloos, M., e Zhu, T. (2018, February). [Benford's law first significant digit and distribution distances for testing the reliability of financial reports in developing countries](#). *Physica A: Statistical Mechanics and its Applications*, 492(1), 878–888.

(Citado 9 vezes nas páginas 23, 24, 25, 41, 48, 56, 58, 68, e 76.)

Shtanko, A., e Kulik, S. (2021). [Increasing the effectiveness of intelligent module by enlarging training dataset from real data](#). *Procedia Computer Science*, 190(1), 712–716.

(Citado 1 vez na página 49.)

Silva, A. d. B., Gomes, M. M., Costa, C. A., Righi, R. d. R., Barbosa, J. L. V., Pessin, G., ... Federizzi, G. (2020, June). [Intelligent personal assistants: A systematic literature review](#). *Expert Systems with Applications*, 147(1), 113193. (Online first.)

(Citado 1 vez na página 29.)

Silva, W., Kimura, H., e Sobreiro, V. A. (2017, February). [An analysis of the literature on systemic financial risk: A survey](#). *Journal of Financial Stability*, 28(1), 91–114.

(Citado 1 vez na página 30.)

Singh, K., e Best, P. (2019, September). [Anti-money laundering: Using data visualization to identify suspicious activity](#). *International Journal of Accounting Information Systems*, 34(1), 100418. (Online first.)

(Citado 7 vezes nas páginas 4, 21, 41, 48, 56, 66, e 71.)

Singh, S., Bhardwaj, N., Sharma, G. D., Kaya, T., Mahendru, M., e Erkut, B. (2019, July). [Research in market-calibrated option pricing analysis](#). *Qualitative Research in Financial Markets*, 12(2), 159–176.

(Citado 1 vez na página 5.)

Skiles, N. (2022). [Diferença entre mineração de dados supervisionada e não supervisionada](#). (Artigo online.)

(Citado 1 vez na página 23.)

Smith, S. D., Dixon, M. J., Tays, W. J., e Bulman-Fleming, M. (2004, August). [Anomaly detection in the right hemisphere: The influence of visuospatial factors](#). *Brain and Cognition*, 55(3), 458–462.

(Citado 1 vez na página 4.)

Sokolova, K., Perez, C., e Lemercier, M. (2017, January). [Android application classification and anomaly detection with graph-based permission patterns](#). *Decision Support Systems*, 93(1), 62–76.

(Citado 1 vez na página 2.)

Soulé, E., Michonneau, P., Michel, N., e Bockstaller, C. (2021, November). [Environmental sustainability assessment in agricultural systems: A conceptual and methodological review](#). *Journal of Cleaner Production*, 325(1),

129291. (Online first.)
(Citado 1 vez na página 48.)
- Stefanelli, N. O., Teixeira, A. A., Oliveira, J. H. C. D., Ferreira, M. A., e Sehnem, S. (2019, May). [Environmental training: A systematic review of the state of the art of the theme](#). *Benchmarking: An International Journal*, 27(7), 2048–2076.
(Citado 1 vez na página 5.)
- Stripling, E., Baesens, B., Chizi, B., e vanden Broucke, S. (2018, July). [Isolation-based conditional anomaly detection on mixed-attribute data to uncover workers' compensation fraud](#). *Decision Support Systems*, 111(1), 13–26.
(Citado 4 vezes nas páginas 3, 41, 56, e 64.)
- Subudhi, S., e Panigrahi, S. (2020, June). [Use of optimized Fuzzy C-Means clustering and supervised classifiers for automobile insurance fraud detection](#). *Journal of King Saud University - Computer and Information Sciences*, 32(5), 568–575.
(Citado 2 vezes nas páginas 41 e 56.)
- Taillard, E. (1993, January). [Benchmarks for basic scheduling problems](#). *European Journal of Operational Research*, 64(2), 278–285.
(Citado 1 vez na página 75.)
- Talan, G., e Sharma, G. (2019, January). [Doing well by doing good: A systematic review and research agenda for sustainable investment](#). *Sustainability*, 11(2), 353.
(Citado 1 vez na página 5.)
- Tang, L., Li, J., Du, H., Li, L., Wu, J., e Wang, S. (2022, February). [Big Data in forecasting research: A literature review](#). *Big Data Research*, 27(1), 100289. (Online first.)
(Citado 1 vez na página 13.)
- Tardelli, S., Avvenuti, M., Tesconi, M., e Cresci, S. (2021, March). [Detecting inorganic financial campaigns on Twitter](#). *Information Systems*, 103(1), 101769. (Online first.)
(Citado 2 vezes nas páginas 41 e 56.)
- Thompson, B. S. (2017, sep). [Can Financial Technology Innovate Benefit Distribution in Payments for Ecosystem Services and REDD? Ecological Economics](#), 139(1), 150–157.
(Citado 1 vez na página 2.)
- Tian, G., Li, B., e Cheng, Y. (2022, October). [Does digital transformation matter for corporate risk-taking? Finance Research Letters](#), 49(1), 103107. (Online first.)
(Citado 2 vezes nas páginas 9 e 10.)
- Tian, X., Han, R., Wang, L., Lu, G., e Zhan, J. (2015, December). [Latency critical big data computing in finance](#). *The Journal of Finance and Data Science*, 1(1), 33–41.
(Citado 2 vezes nas páginas 13 e 14.)
- Timmons, S., McGowan, F. P., e Lunn, P. D. (2019, September). [Setting defaults for online banking transactions: Experimental evidence from personal loan repayment terms](#). *Journal of Behavioral and Experimental Finance*, 23(1), 161–165.
(Citado 1 vez na página 2.)
- Tiwari, R., Srivastava, S., e Gera, R. (2020). [Investigation of Artificial Intelligence Techniques in Finance and Marketing](#). *Procedia Computer Science*, 173(1), 149–157.
(Citado 1 vez na página 10.)
- Tra, V., Amayri, M., e Bouguila, N. (2022, November). [Unsupervised outlier detection using neural network-based mixtures of probabilistic principal component analyzers for building chiller fault diagnosis](#). *Building and Environment*, 225(1), 109620. (Online first.)
(Citado 1 vez na página 20.)
- Trujillo, J., Davis, K. C., Du, X., Damiani, E., e Storey, V. C. (2021, September). [Conceptual modeling in the era of Big Data and Artificial Intelligence: Research topics and introduction to the special issue](#). *Data & Knowledge Engineering*, 135(1), 101911. (Online first.)
(Citado 3 vezes nas páginas 25, 61, e 65.)
- Tsai, C.-F., e Chen, Z.-Y. (2014, May). [Towards high dimensional instance selection: An evolutionary approach](#). *Decision Support Systems*, 61(1), 79–92.
(Citado 1 vez na página 65.)
- Tung, L. L., e Quaddus, M. (2002, June). [Cultural differences explaining the differences in results in GSS: Implications for the next decade](#). *Decision Support Systems*, 33(2), 177–199.
(Citado 2 vezes nas páginas 58 e 62.)
- Turner, A. (2021, June). [How many smartphones are in the world?](#) (Report published by BankMyCell.)
(Citado 1 vez na página 2.)
- Vandervorst, F., Verbeke, W., e Verdonck, T. (2022, August). [Data misrepresentation detection for insurance underwriting fraud prevention](#). *Decision Support Systems*, 159(1), 113798. (Online first.)

- (Citado 2 vezes nas páginas 41 e 56.)
- Venancio, V. (2021, November). *Sua empresa está realmente na Transformação Digital?* (Publicado online.)
- (Citado 1 vez na página 8.)
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J., e Baesens, B. (2012, April). *New insights into churn prediction in the telecommunication sector: A profit driven data mining approach.* *European Journal of Operational Research*, 218(1), 211–229.
- (Citado 1 vez na página 2.)
- Vial, G. (2019, June). *Understanding digital transformation: A review and a research agenda.* *The Journal of Strategic Information Systems*, 28(2), 118–144.
- (Citado 1 vez na página 8.)
- Waite, N. (1999, April). *A real-time system-adapted anomaly detector.* *Information Sciences*, 115(1–4), 221–259.
- (Citado 1 vez na página 4.)
- Wang, H., Ma, S., Dai, H. N., Imran, M., e Wang, T. (2020, September). *Blockchain-based data privacy management with Nudge theory in open banking.* *Future Generation Computer Systems*, 110(1), 812–823.
- (Citado 2 vezes nas páginas 17 e 18.)
- Wang, H., e Weigend, A. S. (2004, September). *Data mining for financial decision making.* *Decision Support Systems*, 37(4), 457–460.
- (Citado 1 vez na página 49.)
- Wang, Y., e Xu, W. (2018, January). *Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud.* *Decision Support Systems*, 105(1), 87–95.
- (Citado 4 vezes nas páginas 41, 56, 64, e 68.)
- Wei, L., Peng, M., e Wu, W. (2021, November). *Financial literacy and fraud detection - Evidence from China.* *International Review of Economics & Finance*, 76(1), 478–494.
- (Citado 2 vezes nas páginas 41 e 56.)
- Wessels, M., van den Brink, P., Verburch, T., Cadet, B., e van Ruijven, T. (2021, October). *Understanding incentives for cybersecurity investments: Development and application of a typology.* *Digital Business*, 1(2), 100014.
- (Citado 4 vezes nas páginas 2, 5, 73, e 76.)
- Wever, M., Shah, M., e O’Leary, N. (2022, February). *Designing early warning systems for detecting systemic risk: A case study and discussion.* *Futures*, 136(1), 102882. (Online first.)
- (Citado 2 vezes nas páginas 42 e 56.)
- Wimmer, H., Yoon, V. Y., e Sugumaran, V. (2016, August). *A multi-agent system to support evidence based medicine and clinical decision making via data sharing and data privacy.* *Decision Support Systems*, 88(1), 51–66.
- (Citado 1 vez na página 66.)
- Wu, N., e Zhang, J. (2006, October). *Factor-analysis based anomaly detection and clustering.* *Decision Support Systems*, 42(1), 375–389.
- (Citado 2 vezes nas páginas 3 e 4.)
- Wu, Y., e Huang, S. (2022, August). *The effects of digital finance and financial constraint on financial performance: Firm-level evidence from China’s new energy enterprises.* *Energy Economics*, 112(1), 106158. (Online first.)
- (Citado 1 vez na página 10.)
- Wu, Y., Ngai, E. W., Wu, P., e Wu, C. (2020, May). *Fake online reviews: Literature review, synthesis, and directions for future research.* *Decision Support Systems*, 132(1), 113280. (Online first.)
- (Citado 1 vez na página 29.)
- Xia, X., Pan, X., Li, N., He, X., Ma, L., Zhang, X., e Ding, N. (2022, July). *GAN-based anomaly detection: A review.* *Neurocomputing*, 493(1), 497–535.
- (Citado 1 vez na página 20.)
- Yan, C., Li, Y., Liu, W., Li, M., Chen, J., e Wang, L. (2020, June). *An artificial bee colony-based kernel ridge regression for automobile insurance fraud identification.* *Neurocomputing*, 393(1), 115–125.
- (Citado 6 vezes nas páginas 42, 48, 56, 60, 68, e 71.)
- Yankol-Schalck, M. (2022, December). *The value of cross-data set analysis for automobile insurance fraud detection.* *Research in International Business and Finance*, 63(1), 101769. (Online first.)
- (Citado 2 vezes nas páginas 42 e 56.)
- Yasami, Y., e Safaei, F. (2017, March). *A statistical infinite feature cascade-based approach to anomaly detection for dynamic social networks.* *Computer Communications*, 100(1), 52–64.
- (Citado 1 vez na página 4.)
- Yoon, K., Liu, Y., Chiu, T., e Vasarhelyi, M. A. (2021, September). *Design and evaluation of an advanced continuous data level auditing system: A three-layer structure.* *International Journal of Accounting Information Systems*, 42(1), 100524. (Online first.)
- (Citado 2 vezes nas páginas 42 e 56.)
- Yu, W., Wong, C. Y., Chavez, R., e Jacobs, M. A. (2021, June). *Integrating big data analytics into supply chain finance: The roles of information processing and data-driven culture.* *International Journal of Production*

- Economics*, 236(1), 108135. (Online first.)
(Citado 1 vez na página 13.)
- Yuan, K., Chi, G., Zhou, Y., e Yin, H. (2022, January). [A novel two-stage hybrid default prediction model with k-means clustering and support vector domain description](#). *Research in International Business and Finance*, 59(1), 101536. (Online first.)
(Citado 2 vezes nas páginas 42 e 56.)
- Zaffar, M. A., Kumar, R. L., e Zhao, K. (2019, February). [Using agent-based modelling to investigate diffusion of mobile-based branchless banking services in a developing country](#). *Decision Support Systems*, 117(1), 62–74.
(Citado 2 vezes nas páginas 2 e 59.)
- Zhai, J., Cao, Y., Yao, Y., Ding, X., e Li, Y. (2017, January). [Computational intelligent hybrid model for detecting disruptive trading activity](#). *Decision Support Systems*, 93(1), 26–41.
(Citado 3 vezes nas páginas 42, 56, e 70.)
- Zhang, G., Atasoy, H., e Vasarhelyi, M. A. (2022, September). [Continuous monitoring with machine learning and interactive data visualization: An application to a healthcare payroll process](#). *International Journal of Accounting Information Systems*, 46(1), 100570. (Online first.)
(Citado 3 vezes nas páginas 42, 56, e 59.)
- Zhang, J.-T., Zhou, B., Guo, J., e Zhu, T. (2021, July). [Two-sample Behrens-Fisher problems for high-dimensional data: A normal reference approach](#). *Journal of Statistical Planning and Inference*, 213(1), 142–161.
(Citado 2 vezes nas páginas 25 e 74.)
- Zhang, K. Z., e Benyoucef, M. (2016, June). [Consumer behavior in social commerce: A literature review](#). *Decision Support Systems*, 86(1), 95–108.
(Citado 1 vez na página 29.)
- Zhang, W., Xie, R., Wang, Q., Yang, Y., e Li, J. (2022, June). [A novel approach for fraudulent reviewer detection based on weighted topic modelling and nearest neighbors with asymmetric Kullback–Leibler divergence](#). *Decision Support Systems*, 157(1), 113765. (Online first.)
(Citado 2 vezes nas páginas 42 e 56.)
- Zhang, Y., Hu, A., Wang, J., e Zhang, Y. (2022, May). [Detection of fraud statement based on word vector: Evidence from financial companies in China](#). *Finance Research Letters*, 46(1), 102477. (Online first.)
(Citado 2 vezes nas páginas 42 e 56.)
- Zhang, Z., Jin, J., Li, S., e Zhang, Y. (2023, February). [Digital transformation of incumbent firms from the perspective of portfolios of innovation](#). *Technology in Society*, 72(1), 102149. (Online first.)
(Citado 1 vez na página 9.)
- Zhou, Y., Lu, Z., Hu, J., e Hu, Y. (2020, June). [Surrogate modeling of high-dimensional problems via data-driven polynomial chaos expansions and sparse partial least square](#). *Computer Methods in Applied Mechanics and Engineering*, 364(1), 112906. (Online first.)
(Citado 1 vez na página 26.)
- Zimek, A., Schubert, E., e Kriegel, H.-P. (2012, August). [A survey on unsupervised outlier detection in high-dimensional numerical data](#). *Statistical Analysis and Data Mining*, 5(5), 363–387.
(Citado 1 vez na página 74.)
- Zoričák, M., Gnip, P., Drotár, P., e Gazda, V. (2020, January). [Bankruptcy prediction for small- and medium-sized companies using severely imbalanced datasets](#). *Economic Modelling*, 84(1), 165–176.
(Citado 2 vezes nas páginas 42 e 56.)

APÊNDICE A

O cálculo que estabelece a distribuição da Lei de Benford indica a probabilidade de aparição de dígitos significativos dentro de um conjunto de dados. O primeiro, e mais comum, é o cálculo da $\text{Prob}(D_1)$, em que D_1 representa o primeiro dígito significativo de um número:

$$\text{Prob}(D_1 = d_1) = \log \left(1 + \frac{1}{d_1} \right); d_1 \in \{1, 2, 3, \dots, 9\}$$

Quando d_1 é substituído pelos números 1 a 9, um a um, como nos exemplos a seguir, encontra-se exatamente os valores relacionados às porcentagens de distribuição mais conhecidas da Lei de Benford:

$$\text{Prob}(D_1 = 1) = \log \left(1 + \frac{1}{1} \right) = \log 2 = 0,30103$$

$$\text{Prob}(D_1 = 2) = \log \left(\frac{3}{2} \right) = 0,17609$$

⋮

$$\text{Prob}(D_1 = 9) = \log \left(\frac{10}{9} \right) = 0,0457$$

No entanto, o cálculo de distribuição das frequências pode ser feito também a partir do cálculo da $\text{Prob}(D_2)$, em que D_2 representa o segundo dígito significativo de um número:

$$\text{Prob}(D_2 = d_2) = \sum_{d_1=1}^9 \log \left(1 + \frac{1}{d_1 d_2} \right); d_2 \in \{0, 1, 2, \dots, 9\}$$

Nesse caso, o denominador $d_1 d_2$ representa os valores do primeiro e segundo dígitos significativos concatenados, criando uma outra distribuição conforme os exemplos a seguir:

$$\begin{aligned} \text{Prob}(D_2 = 0) &= \sum_{d_1=1}^9 \log \left(1 + \frac{1}{d_1 d_2} \right) \\ &= \log \left(1 + \frac{1}{10} \right) + \log \left(1 + \frac{1}{20} \right) + \log \left(1 + \frac{1}{30} \right) \\ &+ \log \left(1 + \frac{1}{40} \right) + \log \left(1 + \frac{1}{50} \right) + \log \left(1 + \frac{1}{60} \right) \\ &+ \log \left(1 + \frac{1}{70} \right) + \log \left(1 + \frac{1}{80} \right) + \log \left(1 + \frac{1}{90} \right) \\ &= 0,11968 \end{aligned}$$

$$\begin{aligned}
\text{Prob}(D_2 = 1) &= \sum_{d_1=1}^9 \log\left(1 + \frac{1}{d_1 d_2}\right) \\
&= \log\left(1 + \frac{1}{11}\right) + \log\left(1 + \frac{1}{21}\right) + \log\left(1 + \frac{1}{31}\right) \\
&\quad + \log\left(1 + \frac{1}{41}\right) + \log\left(1 + \frac{1}{51}\right) + \log\left(1 + \frac{1}{61}\right) \\
&\quad + \log\left(1 + \frac{1}{71}\right) + \log\left(1 + \frac{1}{81}\right) + \log\left(1 + \frac{1}{91}\right) \\
&= 0,11389
\end{aligned}$$

⋮

$$\begin{aligned}
\text{Prob}(D_2 = 9) &= \sum_{d_1=1}^9 \log\left(1 + \frac{1}{d_1 d_2}\right) \\
&= \log\left(1 + \frac{1}{19}\right) + \log\left(1 + \frac{1}{29}\right) + \log\left(1 + \frac{1}{39}\right) \\
&\quad + \log\left(1 + \frac{1}{49}\right) + \log\left(1 + \frac{1}{59}\right) + \log\left(1 + \frac{1}{69}\right) \\
&\quad + \log\left(1 + \frac{1}{79}\right) + \log\left(1 + \frac{1}{89}\right) + \log\left(1 + \frac{1}{99}\right) \\
&= 0,08500
\end{aligned}$$

Todos os resultados e proporções podem ser vistos na Tabela 5.1:

Dígito	Prob(D₁)	Prob(D₂)
0		0,11968
1	0,30103	0,11389
2	0,17609	0,10882
3	0,12494	0,10433
4	0,09691	0,10031
5	0,07918	0,09668
6	0,06695	0,09337
7	0,05799	0,09035
8	0,05115	0,08757
9	0,04576	0,08500

Tabela 5.1: Proporções do primeiro e segundo dígito da Lei de Benford.

Ambas as formas de calcular as frequências resultam em distribuições que podem ser utilizadas para a análise da conformidade de dados com a Lei de Benford. O cálculo das probabilidades pode ser estendido para o restante dos dígitos significativos. No entanto, o mais habitual dentre os artigos presentes na base de literatura deste trabalho é o cálculo da probabilidade $\text{Prob}(D_1)$, uma vez que indica um padrão de distribuição mais aparente.