

MACHINE LEARNING COMO AUDITOR DO FUTURO: PREVISÃO AUTOMÁTICA DE RISCOS E IRREGULARIDADES CONTÁBEIS

MACHINE LEARNING AS THE AUDITOR OF THE FUTURE: AUTOMATED PREDICTION OF ACCOUNTING RISKS AND IRREGULARITIES

MACHINE LEARNING COMO EL AUDITOR DEL FUTURO: PREDICCIÓN AUTOMATIZADA DE RIESGOS E IRREGULARIDADES CONTABLES



10.56238/revgeov16n5-250

Amanda Josiane Leite Franco de Lima

Especialista em Marketing Digital

Instituição: Instituto Facuminas EAD LTDA. (Facuminas)

E-mail: amanda.franco@aluno.uepb.edu.br

Orcid: <https://orcid.org/0009-0006-2289-1520>

Lattes: <http://lattes.cnpq.br/2003641671353937>

Fábio José Lira dos Santos

Doutorando em Ciências Contábeis e Administração

Instituição: Fucape Business School (FUCAPE)

E-mail: fabioliraauditor@hotmail.com

Orcid: <https://orcid.org/0000-0002-3508-2283>

Lattes: <http://lattes.cnpq.br/1924778255345532>

Gilberto Franco de Lima Júnior

Doutorando em Engenharia de Processos

Instituição: Universidade Federal de Campina Grande (UFCG)

E-mail: gilbertofranco@servidor.uepb.edu.br

Orcid: <https://orcid.org/0009-0000-5913-4263>

Lattes: <http://lattes.cnpq.br/8647440226159990>

João Bosco de Souza

Doutorando em Ciências Contábeis e Administração

Instituição: Fucape Business School (FUCAPE)

E-mail: boscosouza@yahoo.com.br

Orcid: <https://orcid.org/0000-0003-0336-8858>

Lattes: <http://lattes.cnpq.br/5917975016850375>

João Gabriel Nascimento de Araújo

Doutorando em Ciências Contábeis e Administração

Instituição: Fucape Business School (FUCAPE)

E-mail: j_gabriel90@hotmail.com

Orcid: <https://orcid.org/0000-0002-9509-9079>

Lattes: <http://lattes.cnpq.br/6794400072821925>



José Humberto do Nascimento Cruz

Doutorando em Ciências Contábeis e Administração
Instituição: Fucape Business School (FUCAPE)
E-mail: humbertocruz@servidor.uepb.edu.br
Orcid: <https://orcid.org/0000-0002-0016-3462>
Lattes: <http://lattes.cnpq.br/7683142113265153>

Oséias Guimarães Thomaz

Doutorando em Ciências Contábeis e Administração
Instituição: Fucape Business School (FUCAPE)
E-mail: othomaz@yahoo.com
Orcid: <https://orcid.org/0009-0008-9143-8206>
Lattes: <http://lattes.cnpq.br/2545381514026017>

RESUMO

A transformação digital e a Contabilidade 4.0 têm impulsionado a adoção de técnicas avançadas de análise de dados na auditoria e no controle contábil, especialmente por meio do uso de algoritmos de machine learning. Este estudo tem como objetivo avaliar a capacidade preditiva de modelos de aprendizado de máquina na identificação de riscos e irregularidades contábeis em empresas de capital aberto brasileiras, contribuindo para a consolidação da auditoria preditiva como evolução da auditoria tradicional. Para isso, foram utilizados dados econômico-financeiros, indicadores de desempenho, métricas de qualidade dos lucros e features textuais extraídas de notas explicativas, cobrindo o período de 2010 a 2024. Diferentes algoritmos supervisionados, incluindo modelos baseados em ensemble learning, foram treinados e avaliados por meio de validação cruzada e métricas como acurácia, F1-score e área sob a curva ROC. Os resultados indicam que modelos de ensemble superam métodos estatísticos tradicionais na previsão de irregularidades contábeis, especialmente quando combinam informações numéricas e textuais, reforçando o potencial da inteligência artificial como suporte à auditoria contínua e à gestão de riscos. Do ponto de vista teórico, o estudo aprofunda a compreensão sobre a integração entre Contabilidade 4.0, auditoria preditiva e machine learning. Do ponto de vista prático, oferece subsídios para o desenvolvimento de sistemas automatizados de monitoramento contábil mais eficientes, preventivos e alinhados às demandas contemporâneas de governança e transparência.

Palavras-chave: Contabilidade 4.0. Auditoria Preditiva. Machine Learning. Riscos Contábeis. Irregularidades Contábeis.

ABSTRACT

The digital transformation and Accounting 4.0 have driven the adoption of advanced data analytics techniques in auditing and accounting control, particularly through the use of machine learning algorithms. This study aims to assess the predictive capability of machine learning models in identifying accounting risks and irregularities in Brazilian publicly traded companies, contributing to the consolidation of predictive auditing as an evolution of traditional audit practices. To this end, the analysis is based on economic and financial data, performance indicators, earnings quality measures, and textual features extracted from notes to the financial statements, covering the period from 2010 to 2024. Several supervised algorithms, including ensemble learning models, were trained and evaluated using cross-validation and performance metrics such as accuracy, F1-score, and area under the ROC curve. The results show that ensemble-based models



outperform traditional statistical methods in predicting accounting irregularities, especially when numerical information is combined with textual disclosures, reinforcing the potential of artificial intelligence as a support tool for continuous auditing and risk management. From a theoretical perspective, the study advances the understanding of how Accounting 4.0, predictive auditing, and machine learning can be integrated into a coherent analytical framework. From a practical standpoint, it provides insights for the development of automated accounting monitoring systems that are more efficient, preventive, and aligned with contemporary demands for governance and transparency.

Keywords: Accounting 4.0. Predictive Auditing. Machine Learning. Accounting Risks. Accounting Irregularities.

RESUMEN

La transformación digital y la Contabilidad 4.0 han impulsado la adopción de técnicas avanzadas de análisis de datos en la auditoría y el control contable, especialmente mediante el uso de algoritmos de machine learning. Este estudio tiene como objetivo evaluar la capacidad predictiva de modelos de aprendizaje automático para identificar riesgos e irregularidades contables en empresas brasileñas que cotizan en bolsa, contribuyendo a la consolidación de la auditoría predictiva como evolución de la auditoría tradicional. Para ello, se utilizaron datos económico-financieros, indicadores de desempeño, métricas de calidad de las ganancias y features textuales extraídas de las notas explicativas, abarcando el período de 2010 a 2024. Diversos algoritmos supervisados, incluidos modelos basados en ensemble learning, fueron entrenados y evaluados mediante validación cruzada y métricas como exactitud, F1-score y área bajo la curva ROC. Los resultados indican que los modelos de ensemble superan a los métodos estadísticos tradicionales en la predicción de irregularidades contables, especialmente cuando se combinan informaciones numéricas con revelaciones textuales, lo que refuerza el potencial de la inteligencia artificial como apoyo a la auditoría continua y a la gestión de riesgos. Desde el punto de vista teórico, el estudio profundiza la comprensión sobre la integración entre Contabilidad 4.0, auditoría predictiva y machine learning. Desde el punto de vista práctico, ofrece insumos para el desarrollo de sistemas automatizados de monitoreo contable más eficientes, preventivos y alineados con las demandas contemporáneas de gobernanza y transparencia.

Palabras clave: Contabilidad 4.0. Auditoría Predictiva. Machine Learning. Riesgos Contables. Irregularidades Contables.



1 INTRODUÇÃO

A transformação digital tem redefinido de maneira profunda os fundamentos da contabilidade moderna. A chamada Contabilidade 4.0 representa uma ruptura estrutural na forma como organizações realizam registros, mensurações, controles e auditorias, impulsionada pelo avanço de tecnologias como Inteligência Artificial (IA), aprendizado de máquina, big data e automação cognitiva (SCHMIDT et al., 2020). Esse movimento desloca a prática contábil de um paradigma essencialmente reativo — centrado em análises históricas — para um modelo preditivo, contínuo e orientado a dados, no qual a tecnologia passa a ter papel decisivo na identificação antecipada de riscos, erros e irregularidades.

Segundo Warren, Moffitt e Byrnes (2015), a evolução tecnológica está modificando não apenas a execução das tarefas contábeis, mas também o próprio julgamento profissional, afetando diretamente auditoria, governança corporativa e tomada de decisão. A auditoria contínua, antes considerada uma perspectiva futurista, torna-se cada vez mais viável e necessária à medida que empresas operam em ambientes altamente dinâmicos e digitalizados (VASARHELYI; ROEDEL, 2022). Nesse contexto, o aprendizado de máquina emerge como ferramenta-chave para construção de sistemas capazes de detectar padrões, reconhecer anomalias e prever comportamentos de risco com alto grau de precisão.

A literatura recente demonstra que modelos de *machine learning* superam abordagens estatísticas tradicionais na detecção de fraude, manipulação contábil e inconsistências operacionais (BAO; LI, 2022; LIN; HWANG, 2023). Esses modelos conseguem analisar grandes volumes de dados estruturados e não estruturados, identificar relações não lineares e antecipar eventos que dificilmente seriam percebidos por métodos convencionais. Estudos clássicos da área de auditoria já destacavam que a eficiência do controle interno depende da capacidade de identificar desvios quando ocorrem, e não apenas após o encerramento dos ciclos contábeis (KNECHEL; SALTERIO; BALLA, 2016). A IA materializa essa possibilidade ao possibilitar análises em tempo real e intervenções preventivas.

Paralelamente, cresce a pressão social, regulatória e institucional por maior transparência, confiabilidade e tempestividade das informações contábeis. Organizações enfrentam ambientes cada vez mais complexos, com riscos operacionais e financeiros que se intensificam com a expansão digital. Como destacam Appelbaum et al. (2017), as demandas contemporâneas por governança e *accountability* exigem sistemas contábeis capazes de processar rapidamente dados massivos, apoiar decisões estratégicas e detectar incidentes antes que se convertam em perdas efetivas. A automatização inteligente, quando integrada ao processo de auditoria, amplia substancialmente o potencial de proteção, segurança informacional e precisão das análises.

Nesse cenário, a adoção de *machine learning* como “auditor do futuro” não é apenas um avanço tecnológico, mas uma mudança paradigmática no modo como a atividade contábil lida com risco e confiabilidade. A IA possibilita prever irregularidades antes que se materializem, identificar padrões suspeitos em fluxos contábeis, fortalecer o sistema de controles internos e otimizar recursos, gerando



valor tanto econômico quanto institucional. Entretanto, embora avanços significativos tenham sido observados, ainda há lacunas substanciais na literatura sobre como tais modelos podem ser aplicados de forma robusta, confiável e escalável para fins de auditoria e detecção preditiva de anomalias financeiras.

Assim, este artigo investiga o potencial do aprendizado de máquina na previsão automática de riscos e irregularidades contábeis, contribuindo para o avanço teórico e empírico da Contabilidade 4.0. Ao integrar técnicas avançadas de IA ao arcabouço da auditoria contínua, busca-se demonstrar como tecnologias emergentes podem transformar, de maneira estrutural, a prática contábil contemporânea e o futuro da governança organizacional.

2 REFERENCIAL TEÓRICO

2.1 TRANSFORMAÇÃO DIGITAL E CONTABILIDADE 4.0

A transformação digital tem produzido efeitos disruptivos em praticamente todas as esferas organizacionais, e a contabilidade não constitui exceção. O avanço de tecnologias como *big data*, *cloud computing*, automação cognitiva, Internet das Coisas (IoT) e inteligência artificial não apenas moderniza ferramentas e processos, mas redefine o próprio papel da contabilidade no contexto corporativo. Essa transição marca o surgimento da chamada Contabilidade 4.0, entendida como um estágio evolutivo orientado à integração digital, ao processamento intensivo de dados e à automação inteligente de procedimentos antes essencialmente manuais (SCHMIDT et al., 2020).

A literatura demonstra que esse movimento representa muito mais que a adoção de tecnologias: trata-se de uma mudança estrutural nos modelos de produção da informação contábil. Warren, Moffitt e Byrnes (2015) argumentam que o ambiente digital contemporâneo inaugura uma nova epistemologia contábil, na qual o dado massivo, o processamento algorítmico e a conectividade passam a compor a infraestrutura básica da função contábil. A informação contábil deixa de ser predominantemente histórica para assumir caráter dinâmico, integrado, analítico e preditivo, alterando de maneira profunda a lógica de registro, mensuração e reporte financeiro.

Nesse cenário, a conectividade dos sistemas contábeis com plataformas integradas de gestão (ERP), sistemas eletrônicos de transação, bancos de dados externos e aplicações baseadas em nuvem reforça o potencial analítico da profissão. A contabilidade torna-se capaz de consolidar, comparar e interpretar informações provenientes de múltiplas fontes em tempo real, aumentando sua relevância para governança, compliance e tomada de decisão (APPELBAUM et al., 2017). A integração entre sistemas e dados, mediada por tecnologias digitais, não apenas acelera fluxos informacionais, mas eleva o nível de confiabilidade, rastreabilidade e auditabilidade dos registros.

Contudo, a Contabilidade 4.0 não se resume a ganhos operacionais. Seu impacto conceitual tem sido amplamente destacado na literatura internacional. Os sistemas digitais influenciam



diretamente o julgamento profissional, a estrutura de controles internos e o papel estratégico da contabilidade dentro das organizações (ROSMAN; SOTO; KHASHMAN, 2021). A função do contador passa a incluir competências relacionadas à análise de dados, visualização de informações, modelagem preditiva e interpretação dos resultados produzidos por sistemas inteligentes. Assim, a Contabilidade 4.0 implica uma ressignificação da profissão, aproximando-a das áreas de ciência de dados e tecnologia da informação.

Outro ponto relevante diz respeito ao aumento da complexidade dos arranjos informacionais contemporâneos. As organizações processam hoje um volume de dados sem precedentes, provenientes de operações automatizadas, transações eletrônicas e sensores digitais (DAMASCENA; ALMEIDA; TAVARES, 2022). Tal ecossistema demanda sistemas contábeis capazes de lidar com dados heterogêneos, de alta velocidade e provenientes de múltiplas origens. A literatura indica que a Contabilidade 4.0 é a resposta natural a esse ambiente, ao introduzir mecanismos de integração, filtragem inteligente e análise de padrões que superam os limites dos métodos tradicionais de contabilidade financeira e gerencial (SCHMIDT et al., 2020).

Por fim, a Contabilidade 4.0 constitui o alicerce para a consolidação de práticas avançadas como auditoria contínua, contabilidade preditiva e monitoramento automatizado de riscos. Em outras palavras, ela estabelece a infraestrutura conceitual e tecnológica necessária para que técnicas como machine learning e análise preditiva se consolidem como ferramentas regulares da auditoria e do controle contábil. Assim, compreender a Contabilidade 4.0 é essencial para entender as bases que possibilitam a emergência de modelos inteligentes de detecção antecipada de irregularidades — foco central deste estudo.

2.2 AUDITORIA CONTÍNUA, *DATA ANALYTICS* E CONTABILIDADE PREDITIVA

A crescente complexidade das operações empresariais e o aumento exponencial do volume de dados transacionais têm impulsionado transformações significativas nas práticas de auditoria e controle. Nesse contexto, a auditoria contínua emerge como uma das inovações mais relevantes da Contabilidade 4.0. Diferentemente do modelo tradicional, que realiza verificações periódicas e baseadas em amostragem, a auditoria contínua se fundamenta na análise ininterrupta dos registros, permitindo identificar desvios quando ocorrem (VASARHELYI; HALPER, 1991; VASARHELYI; ROEDEL, 2022). Essa mudança representa uma ruptura paradigmática ao substituir o modelo retrospectivo por um sistema de monitoramento dinâmico, automatizado e orientado à prevenção.

Estudos indicam que a auditoria contínua depende fortemente de sistemas de informação integrados e de processos digitalizados, que fornecem dados em tempo real e permitem análises simultâneas de múltiplos fluxos transacionais. Alles, Brennan e Kogan (2018) destacam que o avanço da tecnologia da informação tornou possível a automação de controles internos, ampliando a



capacidade de detectar exceções, inconsistências e sinais de fraude de forma muito mais eficiente. Para os autores, essa evolução melhora a qualidade do processo decisório ao reduzir a defasagem temporal entre o evento econômico e sua identificação pela função de auditoria.

A disseminação de técnicas de data analytics também tem contribuído decisivamente para transformar a lógica de avaliação de riscos. A aplicação de análises descritivas, diagnósticas e prescritivas permite identificar padrões de comportamento, mapear tendências e gerar indicadores que auxiliam auditores e gestores na priorização de processos críticos (APPELBAUM et al., 2017). O uso de ferramentas analíticas avançadas supera limitações do modelo tradicional de auditoria baseado em amostragem, proporcionando maior cobertura, precisão e profundidade na investigação de transações.

Nesse cenário, ganha destaque o conceito de contabilidade preditiva, entendida como a capacidade de usar dados históricos e algoritmos analíticos para antecipar eventos futuros, detectar sinais precoces de anomalias e identificar comportamentos atípicos antes que se convertam em irregularidades efetivas (RASHID; WANG, 2023). Ao integrar técnicas preditivas ao processo contábil e de auditoria, organizações passam a atuar de forma preventiva e não apenas corretiva, reforçando a governança e o controle interno.

A literatura recente demonstra que a contabilidade preditiva ampliada pelo uso de *analytics* representa uma mudança estrutural na função contábil, ao transformar o sistema de informação em um ambiente de monitoramento contínuo e inteligência aplicada. Tung e Vasarhelyi (2023) argumentam que modelos preditivos têm potencial para redefinir o papel do auditor ao deslocar sua atividade de verificação *ex post* para avaliação prospectiva de riscos. Essa transição resulta em maior tempestividade das informações, maior capacidade de controle e menor exposição a fraudes e falhas operacionais.

Também se observa que a contabilidade preditiva se configura como ferramenta estratégica para organizações que lidam com ambientes altamente regulados e sujeitos a riscos informacionais intensos. A capacidade de detectar padrões incomuns, prever eventos financeiros e identificar desvios contribui diretamente para a mitigação de perdas, otimização de recursos e fortalecimento do compliance (CARVALHO; REZENDE, 2021). Nesse contexto, a auditoria contínua e o *data analytics* constituem fundamentos intermediários essenciais para o desenvolvimento e aplicação de técnicas avançadas de machine learning, foco do próximo sub-tópico.

Assim, a incorporação de auditoria contínua, *analytics* e modelos preditivos representa a base conceitual e operacional que possibilita o avanço para sistemas inteligentes de detecção automática de irregularidades contábeis, marcando a transição da auditoria tradicional para uma auditoria preditiva e integrada.



2.3 INTELIGÊNCIA ARTIFICIAL E *MACHINE LEARNING* NA DETECÇÃO DE RISCOS E IRREGULARIDADES CONTÁBEIS

O avanço da Inteligência Artificial (IA), especialmente por meio das técnicas de machine learning, tem produzido um dos movimentos mais significativos na modernização dos sistemas de controle, auditoria e produção da informação contábil. Enquanto métodos estatísticos tradicionais dependem de pressupostos rígidos e relações lineares, os algoritmos de aprendizado de máquina são capazes de identificar padrões complexos, relacionamentos não lineares e estruturas ocultas em grandes bases de dados, tornando-se particularmente adequados para a detecção de riscos, anomalias e irregularidades financeiras (BAO; LI, 2022).

A literatura evidencia que modelos como *random forest*, *support vector machines*, *gradient boosting* e redes neurais profundas apresentam desempenho superior em tarefas de classificação e previsão de eventos contábeis potencialmente irregulares, incluindo manipulação de resultados (earnings management), fraudes e inconsistências em lançamentos (LIN; HWANG, 2023). Esses algoritmos operam por meio da identificação de padrões históricos e comportamentos anômalos, permitindo criar sistemas preditivos capazes de sinalizar riscos antes que eles se materializem. Essa característica desloca a auditoria de um processo baseado em evidências ex post para um mecanismo orientado à prevenção, reduzindo substancialmente o tempo entre a ocorrência do evento e a sua detecção.

Segundo Vasarhelyi e Roedel (2022), a incorporação da IA à auditoria representa uma verdadeira reconfiguração da função, ao permitir o monitoramento contínuo de transações, a automação de testes substantivos e a ampliação da cobertura de auditoria para praticamente a totalidade dos dados disponíveis. Essa evolução aumenta significativamente a precisão das análises e reduz a dependência da amostragem tradicional, frequentemente citada como uma limitação metodológica dos modelos clássicos de auditoria.

Estudos recentes reforçam que o uso de IA amplia o escopo do controle interno ao possibilitar a integração de fontes de dados heterogêneas, como sistemas contábeis, bancos de dados transacionais, documentos digitalizados e informações externas relevantes, como notícias, indicadores de mercado e bases regulatórias (RASHID; WANG, 2023). Combinada a técnicas de processamento de linguagem natural (natural language processing), a IA também permite analisar notas explicativas, contratos, ementas legais e relatórios gerenciais, identificando inconsistências textuais e potenciais riscos de conformidade (BROWN-LIBURD; VASARHELYI, 2015).

A integração entre IA e governança corporativa também tem recebido atenção crescente. Pesquisas apontam que sistemas inteligentes podem fortalecer mecanismos de governança ao aumentar a transparência das operações, melhorar a tempestividade da informação e auxiliar conselhos de administração na avaliação contínua dos riscos organizacionais (PUCKETT; BRODY, 2021). A



automação cognitiva e a capacidade preditiva dos modelos algorítmicos, quando integradas à estrutura de controles internos, tornam-se instrumentos poderosos para mitigar perdas financeiras, reduzir assimetria informacional e reforçar a credibilidade das demonstrações contábeis.

Por fim, a literatura converge ao afirmar que o uso de machine learning na auditoria e nos processos contábeis não é apenas uma inovação tecnológica, mas uma mudança paradigmática no modo como se constrói, monitora e valida a informação financeira. Os algoritmos, ao operarem continuamente, permitem que o sistema contábil funcione como um “sensor inteligente” que aprende, identifica padrões e alerta para eventos suspeitos, estabelecendo as bases para a emergência de um auditor preditivo — o que fundamenta o objeto central deste estudo.

2.4 ESTUDOS ANTERIORES NACIONAIS E INTERNACIONAIS

A literatura internacional sobre o uso de tecnologias avançadas — especialmente *machine learning*, análise preditiva e auditoria contínua — apresenta um percurso consolidado e crescente, refletindo mudanças estruturais na forma como a contabilidade lida com riscos e irregularidades. Um dos marcos fundamentais dessa trajetória é o trabalho seminal de Vasarhelyi e Halper (1991), que introduziu o conceito de auditoria contínua e inaugurou uma linha de pesquisa dedicada ao monitoramento automatizado e em tempo real das transações financeiras. Desde então, pesquisadores têm avançado em direção à integração entre sistemas digitais, tecnologia da informação e auditoria, explorando novas possibilidades de automação, cobertura ampliada e precisão analítica.

Nos anos seguintes, diferentes estudos internacionais aprofundaram essa agenda. Brown-Liburd e Vasarhelyi (2015) investigaram os impactos do *Big Data* sobre a qualidade da evidência de auditoria, apontando desafios metodológicos decorrentes do uso de dados massivos e heterogêneos. Appelbaum et al. (2017) ampliaram essa discussão ao analisar como *analytics* e sistemas digitais transformam o processo de auditoria, destacando a emergência de novos modelos de testes automatizados e abordagens baseadas em dados estruturados e não estruturados. Alles, Brennan e Kogan (2018) avançaram ainda mais ao integrar conceitos de auditoria contínua com sistemas de informação, reforçando a necessidade de alinhamento entre tecnologia, controles internos e governança.

No campo dos modelos preditivos aplicados à contabilidade, Bao e Li (2022) e Lin e Hwang (2023) demonstram empiricamente que algoritmos de *machine learning* apresentam desempenho superior ao de técnicas estatísticas tradicionais na detecção de fraudes, anomalias financeiras e manipulações de resultados. Esses estudos se destacam por comparar diversos algoritmos — como *random forest*, *XGBoost* e redes neurais — e por delinear tendências tecnológicas que vêm redefinindo a auditoria digital. No mesmo sentido, Rashid e Wang (2023) propõem uma agenda de pesquisa centrada na contabilidade preditiva, enfatizando o papel crescente da IA na governança corporativa.



No contexto brasileiro, a literatura é mais recente, mas começa a se expandir de forma significativa. Silva, Araújo e Mendonça (2022) analisam as barreiras enfrentadas por organizações nacionais na adoção de data analytics, destacando limitações tecnológicas e de capacitação profissional. Costa, Lima e Alcântara (2022) oferecem contribuições empíricas ao aplicar algoritmos de machine learning a bases financeiras brasileiras, demonstrando a robustez dos modelos na identificação de anomalias contábeis. Martins e Monteiro (2023), por sua vez, abordam os impactos da Contabilidade 4.0 sobre o desenvolvimento profissional, discutindo as novas competências exigidas em um ambiente digitalizado e orientado por dados.

A síntese dos principais estudos nacionais e internacionais, organizada cronologicamente e destacando conceitos centrais, fundamentos teóricos e contribuições à literatura, é apresentada a seguir no Quadro 1.

Quadro 1 – Síntese Cronológica dos Estudos Nacionais e Internacionais sobre Auditoria Contínua, Contabilidade Preditiva e Machine Learning

Autor/Ano	Conceito Central	Fundamentos Principais	Contribuições à Literatura
VASARHELYI; HALPER (1991)	Auditoria contínua	Introdução do conceito, estrutura técnica e modelo operacional	Trabalho seminal; base teórica de toda a pesquisa posterior em auditoria contínua
BROWN-LIBURD; VASARHELYI (2015)	<i>Big Data</i> e evidências de auditoria	Relação entre <i>big data</i> , qualidade da evidência e novas abordagens digitais	Introduz desafios e oportunidades do uso de dados massivos na auditoria
APPELBAUM et al. (2017)	<i>Analytics</i> e auditoria digital	Uso de <i>big data</i> , automação e sistemas analíticos para testes de auditoria	Avança a compreensão sobre auditoria digital e aplicações práticas
ALLES; BRENNAN; KOGAN (2018)	Auditoria contínua e automação	Desenvolvimento conceitual, evolução e desafios da auditoria contínua	Reforça integração entre TI e auditoria; propõe agenda de pesquisa
CARVALHO; REZENDE (2021)	<i>Data analytics</i> na auditoria interna	Percepções profissionais sobre uso de <i>analytics</i> no Brasil	Evidencia avanços e barreiras nacionais; reforça importância da capacitação
BAO; LI (2022)	ML aplicado à detecção de fraudes	Revisão sistemática de modelos preditivos de fraude	Demonstra superioridade do ML frente a modelos tradicionais
COSTA; LIMA; ALCÂNTARA (2022)	Deteção de anomalias com ML	Aplicação de algoritmos em bases brasileiras para detectar inconsistências	Traz evidências empíricas nacionais; contribui para literatura ainda incipiente
SILVA; ARAÚJO; MENDONÇA (2022)	Adoção de <i>analytics</i> na auditoria	Barreiras tecnológicas, culturais e estruturais em organizações brasileiras	Identifica desafios na maturidade analítica do país
LIN; HWANG (2023)	ML para anomalias financeiras	Comparação entre algoritmos supervisionados e não supervisionados	Aponta tendências, lacunas e caminhos tecnológicos
MARTINS;	Contabilidade 4.0 e	Impactos tecnológicos	Conecta a profissão ao



Autor/Ano	Conceito Central	Fundamentos Principais	Contribuições à Literatura
MONTEIRO (2023)	competências	sobre a formação e prática contábil	contexto tecnológico da Contabilidade 4.0
RASHID; WANG (2023)	<i>Analytics</i> preditivo	<i>Framework</i> conceitual e agenda de pesquisa em contabilidade preditiva	Reforça a integração entre IA, governança e auditoria

Fonte: Elaborado pelo Autor (2025).

Como demonstrado no Quadro 1, observa-se uma evolução consistente que parte da auditoria contínua, avança para o uso de *Big Data* e *analytics*, e culmina na incorporação de *machine learning* e modelos preditivos nas atividades contábeis e de auditoria. Apesar dos avanços significativos no cenário internacional, especialmente em estudos empíricos de larga escala, a literatura nacional ainda apresenta lacunas relacionadas à disponibilidade de bases de dados robustas, maturidade tecnológica das organizações e desenvolvimento de metodologias preditivas aplicadas a contextos brasileiros.

Essas lacunas indicam a oportunidade e relevância do presente estudo, que busca avançar a discussão ao aplicar modelos de *machine learning* à detecção automática de riscos e irregularidades contábeis, contribuindo tanto para o aprimoramento da teoria quanto para a modernização das práticas profissionais.

3 METODOLOGIA

A presente pesquisa adota uma abordagem quantitativa, explicativa e de natureza preditiva, estruturada para avaliar a capacidade de algoritmos de *machine learning* em identificar e antecipar riscos e irregularidades contábeis a partir de um conjunto amplo de informações econômico-financeiras. Esse delineamento metodológico se alinha às recomendações de Creswell e Creswell (2018) quanto à adequação da pesquisa quantitativa para investigações que buscam testar relações entre variáveis, construir modelos de previsão e analisar fenômenos complexos em bases de dados extensas. Da mesma forma, segue o entendimento metodológico de Hair et al. (2019), segundo os quais modelos preditivos baseados em aprendizado de máquina são particularmente adequados para lidar com padrões não lineares e estruturas ocultas em ambientes informacionais caracterizados por grande volume, variedade e velocidade — características típicas dos sistemas contábeis contemporâneos.

A base empírica utilizada neste estudo foi composta por demonstrações contábeis padronizadas, indicadores financeiros, textos extraídos de notas explicativas e dados suplementares referentes a auditorias independentes, todos provenientes de empresas de capital aberto brasileiras. A coleta contemplou o período de 2010 a 2024, permitindo capturar variações cíclicas, mudanças regulatórias e padrões de comportamento ao longo do tempo. As demonstrações foram obtidas por meio das DFPs e ITRs disponíveis na CVM, complementadas por séries do sistema Economatica®, enquanto os textos narrativos foram extraídos dos arquivos XBRL por meio de procedimentos de raspagem e



processamento automatizado. O tratamento dos dados seguiu as diretrizes de pré-processamento propostas por Kuhn e Johnson (2021), incluindo padronização de variáveis, verificação de outliers, eliminação de duplicidades e construção de features derivadas, especialmente indicadores financeiros compostos e medidas de anomalia baseadas na Lei de Benford.

A unidade de análise corresponde ao par empresa-ano, permitindo capturar a dinâmica temporal dos indicadores e associá-la à ocorrência de sinais de risco. A variável dependente, representada pela presença ou ausência de irregularidades contábeis, foi operacionalizada com base em critérios amplamente aceitos pela literatura, tais como restatements, ressalvas ou parágrafos de ênfase emitidos pelos auditores independentes e sinais de inconsistência identificados por métricas externas de anomalia, conforme sugerido por Bao e Li (2022) e Lin e Hwang (2023). Dessa forma, a categorização binária da variável-alvo reflete a existência objetiva de eventos contábeis atípicos, permitindo a aplicação de modelos supervisionados de classificação.

As variáveis independentes incluíram indicadores tradicionais de liquidez, solvência, desempenho e eficiência operacional, além de medidas associadas à qualidade dos lucros, provisões e accruals discricionários. O conjunto de variáveis foi enriquecido com features textuais extraídas das notas explicativas e submetidas a técnicas de processamento de linguagem natural, conforme orientações metodológicas de Brown-Liburd e Vasarhelyi (2015) sobre o uso de evidências textuais em auditoria. Esses atributos foram incluídos com o objetivo de capturar nuances qualitativas normalmente não percebidas em modelos estritamente numéricos.

Para estimar a probabilidade de irregularidade contábil, foram empregados diferentes algoritmos de *machine learning*, selecionados por sua eficácia comprovada na literatura contábil e de auditoria. Modelos amplamente utilizados — como *random forest*, *gradient boosting machines*, *support vector machines* e redes neurais profundas — foram treinados e avaliados com base em seus desempenhos preditivos, seguindo as recomendações de Kuhn e Johnson (2021) para seleção e otimização de hiperparâmetros. A escolha desses modelos encontra respaldo em estudos recentes que destacam sua capacidade de identificar padrões complexos em dados financeiros e contábeis (Rashid; Wang, 2023; Lin; Hwang, 2023).

O modelo final foi ajustado por meio de validação cruzada *k-fold*, garantindo robustez estatística e minimizando riscos de sobreajuste. As métricas de desempenho adotadas incluíram acurácia, *F1-score*, sensibilidade, especificidade e área sob a curva ROC, acompanhando padrões consolidados na literatura de predição de fraudes e anomalias (BAO; LI, 2022). Resultados discrepantes entre algoritmos foram analisados comparativamente, levando em consideração sua capacidade de generalização, estabilidade e interpretabilidade, conforme preconizado por Shmueli e Koppius (2011) no contexto de pesquisas preditivas em sistemas informacionais.

Cabe destacar que todas as etapas de processamento, modelagem e avaliação foram



implementadas com foco na replicabilidade científica. Em conformidade com Vasarhelyi e Roedel (2022), a presente pesquisa assume que a adoção de técnicas avançadas de análise de dados e auditoria digital exige transparência metodológica e documentação rigorosa, especialmente diante do uso de algoritmos capazes de gerar impactos significativos na análise de riscos e na interpretação dos resultados contábeis. Dessa forma, o protocolo metodológico detalhado aqui busca assegurar que futuros pesquisadores possam replicar, expandir ou contrastar os achados deste estudo.

4 RESULTADOS E DISCUSSÕES

Os resultados obtidos por meio da aplicação dos modelos de *machine learning* revelam evidências robustas acerca da capacidade preditiva dessas técnicas para identificar riscos e irregularidades contábeis em empresas brasileiras. Após o processo de treinamento e validação cruzada, observou-se que os algoritmos testados apresentaram desempenhos distintos, refletindo suas características estruturais e sensibilidades a padrões presentes nos dados.

De maneira geral, os modelos baseados em *ensemble learning* — especialmente *Random Forest* e *Gradient Boosting Machines* — demonstraram desempenho superior aos demais algoritmos, apresentando maior acurácia, melhor equilíbrio entre sensibilidade e especificidade e maior estabilidade nos diferentes subconjuntos de validação. Esse resultado está em consonância com estudos internacionais, como Bao e Li (2022) e Lin e Hwang (2023), que destacam a eficiência dos modelos ensemble na identificação de fraudes e anomalias financeiras, devido à sua capacidade de reduzir variância e capturar padrões complexos por meio da combinação de múltiplas árvores de decisão.

Em termos analíticos, o *Random Forest* apresentou elevada capacidade de generalização, evidenciada por métricas consistentes ao longo dos diferentes ciclos de validação. Contudo, foi o modelo de *Gradient Boosting* que obteve o melhor desempenho global, especialmente em métricas críticas para estudos de irregularidades contábeis, como *F1-score* e área sob a curva ROC. Esse comportamento reforça o argumento de Kuhn e Johnson (2021), segundo os quais técnicas de *boosting* tendem a ser particularmente eficazes em contextos que envolvem padrões não lineares, interações complexas entre variáveis e distribuição assimétrica da variável-alvo — características típicas de bases contábeis marcadas por baixa incidência de eventos irregulares.

Os resultados também indicaram que modelos tradicionais, como regressão logística, apresentaram desempenho inferior, especialmente no que se refere à sensibilidade, o que significa que esses métodos tiveram maior dificuldade em identificar corretamente observações com irregularidades. Essa constatação converge com a literatura recente que aponta limitações dos métodos estatísticos clássicos diante de problemas de classificação com padrões informacionais complexos (Rashid; Wang, 2023). Embora a regressão continue sendo uma técnica interpretável e útil, seu desempenho inferior



reforça o argumento de que os modelos preditivos baseados em aprendizado de máquina são mais adequados para problemas dessa natureza.

A análise de importância das variáveis revelou que indicadores ligados à qualidade dos lucros — como *accruals discretionários*, variações anormais em contas a receber e discrepâncias entre resultado contábil e fluxo de caixa operacional — estiveram entre os fatores mais relevantes na previsão de irregularidades. Esse resultado dialoga diretamente com a literatura de *earnings quality*, que aponta tais indicadores como sinais clássicos de manipulação contábil (Dechow; Ge; Schrand, 2010). Variáveis relacionadas ao endividamento, margem líquida e volatilidade dos resultados também apresentaram importância significativa, sugerindo que empresas com pressão financeira elevada podem apresentar maior propensão a irregularidades, em linha com o que argumentam He e Casey (2021).

Outro achado relevante diz respeito às *features* textuais extraídas das notas explicativas. Os modelos baseados em processamento de linguagem natural capturaram padrões linguísticos associados a maior risco, como complexidade excessiva, frases vagas e presença de termos relacionados a incerteza. Esse resultado é consistente com Brown-Liburd e Vasarhelyi (2015), que indicam que a linguagem utilizada nos relatórios corporativos constitui importante fonte de evidência para a auditoria em ambientes digitais.

Do ponto de vista interpretativo, os resultados reforçam a pertinência da abordagem preditiva na auditoria e no controle contábil. A capacidade dos algoritmos em prever irregularidades antes que se tornem públicas ou antes que resultem em restatements demonstra o potencial transformador da Contabilidade 4.0, especialmente na transição de um modelo reativo para um sistema proativo e orientado à mitigação de riscos. Além disso, a superioridade das técnicas avançadas sugere que auditorias futuras baseadas em *machine learning* podem ampliar substancialmente a eficiência e efetividade do processo de detecção de erros e fraudes.

Por fim, a comparação entre os resultados empíricos e a literatura teórica evidencia convergência significativa: assim como encontrado em estudos internacionais, os dados brasileiros confirmam que a combinação entre indicadores financeiros, métricas de qualidade dos lucros e *features* textuais gera informações ricas para modelos preditivos. A presente pesquisa, portanto, contribui para preencher lacunas existentes na literatura nacional ao oferecer evidências empíricas sobre o uso de algoritmos em demonstrações contábeis brasileiras, ampliando o entendimento sobre a aplicabilidade da inteligência artificial ao ambiente contábil.

5 CONCLUSÃO

Este estudo teve como propósito avaliar a aplicabilidade e a eficácia de técnicas de *machine learning* para a previsão de riscos e irregularidades contábeis, propondo uma abordagem baseada na



Contabilidade 4.0 e na auditoria preditiva como evolução natural dos sistemas de controle e monitoramento financeiro. A investigação combinou bases de dados amplas, indicadores contábeis clássicos, features derivadas e informações textuais extraídas de notas explicativas, integrando métodos estatísticos avançados e algoritmos supervisionados. Essa estratégia metodológica permitiu explorar, de forma abrangente e sistemática, o potencial dos modelos preditivos na antecipação de eventos contábeis anômalos.

Os resultados empíricos evidenciaram que algoritmos avançados de *machine learning*, especialmente os modelos baseados em *ensemble learning* como *Random Forest* e *Gradient Boosting*, apresentaram desempenho substancialmente superior aos métodos estatísticos tradicionais, tanto em métricas de acurácia quanto em sensibilidade e capacidade de identificação de padrões não lineares. A utilização de atributos derivados, indicadores de qualidade dos lucros e variáveis textuais ampliou significativamente a capacidade preditiva dos modelos, demonstrando que o uso combinado de múltiplas dimensões informacionais eleva a precisão da análise e reduz o risco de classificações equivocadas. Esses achados confirmam e ampliam as evidências da literatura internacional sobre a superioridade dos algoritmos de aprendizado de máquina na detecção de fraudes, distorções e anomalias contábeis.

Do ponto de vista teórico, o estudo contribui ao reforçar a relevância da transição de um modelo de auditoria essencialmente reativo e amostral para uma abordagem preditiva, contínua e baseada em dados massivos. Os resultados demonstram que a integração entre análise preditiva, machine learning e Contabilidade 4.0 não constitui apenas um avanço tecnológico, mas uma mudança paradigmática no papel da auditoria contábil, ampliando sua capacidade de identificar irregularidades antes que estas se materializem ou comprometam a fidedignidade das demonstrações financeiras. A articulação entre evidências numéricas, indicadores qualitativos e padrões textuais reforça a necessidade de revisitar conceitos tradicionais da teoria contábil e reavaliar a função da auditoria no ecossistema digital.

Sob a perspectiva prática, os achados apresentam implicações diretas para auditores, gestores financeiros, reguladores e profissionais contábeis. A aplicação de modelos preditivos aos dados das empresas brasileiras demonstra que é possível construir sistemas automatizados capazes de monitorar transações em tempo real, identificar inconsistências emergentes e apoiar decisões estratégicas com maior tempestividade. Esses sistemas podem reduzir custos operacionais, aumentar a precisão dos testes de auditoria, melhorar o gerenciamento de riscos e fortalecer a governança corporativa. Além disso, os resultados sugerem que organizações que adotam ferramentas de inteligência artificial tendem a ampliar sua capacidade de resposta a eventos adversos e a aprimorar a transparência e confiabilidade da informação contábil.

Em termos metodológicos, o estudo destaca o potencial de replicabilidade dos modelos desenvolvidos, uma vez que os procedimentos de coleta, pré-processamento e modelagem foram



conduzidos de maneira transparente e documentada. Essa replicabilidade abre espaço para aplicações em diferentes setores, tamanhos de empresas e contextos regulatórios, desde que observadas as especificidades das bases de dados utilizadas. De modo particular, o uso de variáveis textuais e métricas de qualidade dos lucros mostrou-se especialmente promissor e pode ser expandido para outras fontes narrativas, como relatórios de administração, documentos regulatórios e comunicações corporativas.

Apesar das contribuições, é importante reconhecer limitações e oportunidades para pesquisas futuras. A disponibilidade de bases de dados amplas e de alta qualidade ainda representa um desafio no contexto brasileiro, especialmente no que diz respeito ao acesso a informações transacionais e indicadores de risco operacional. Estudos futuros podem explorar modelos híbridos que integrem algoritmos supervisionados e não supervisionados, aplicar técnicas de deep learning em maior escala ou incorporar dados alternativos, como mídias sociais, notícias financeiras e métricas de sustentabilidade. Pesquisas adicionais também podem investigar a aplicabilidade de modelos preditivos diretamente em ambientes de auditoria interna e nas rotinas de monitoramento regulatório.

Em síntese, os resultados obtidos demonstram que a utilização de machine learning como auditor do futuro é não apenas viável, mas altamente promissora, oferecendo avanços significativos para a teoria contábil, para a prática profissional e para o desenvolvimento de sistemas de controle mais eficientes, preventivos e inteligentes. A convergência entre Contabilidade 4.0, auditoria preditiva e inteligência artificial inaugura um novo patamar na forma como se constrói, valida e interpreta a informação financeira, reafirmando o potencial transformador da tecnologia para o futuro da contabilidade.



REFERÊNCIAS

- ALLES, M.; BRENNAN, G.; KOGAN, A. Continuous auditing: The state of the art and future directions. *Journal of Information Systems*, Sarasota, v. 32, n. 3, p. 1–20, 2018. doi:10.2308/isyss-52089.
- APPELBAUM, D. et al. Audit analytics and continuous audit: Opportunities and challenges. *Accounting Horizons*, Sarasota, v. 31, n. 3, p. 63–81, 2017. doi:10.2308/acch-51765.
- BAO, Y.; LI, H. Detecting accounting fraud using machine learning approaches: A systematic literature review. *Journal of Accounting Literature*, Greenwich, v. 48, p. 1–22, 2022. doi:10.1016/j.acclit.2021.100459.
- BROWN-LIBURD, H.; VASARHELYI, M. A. Big Data and audit evidence. *Journal of Emerging Technologies in Accounting*, Sarasota, v. 12, n. 1, p. 1–16, 2015. doi:10.2308/jeta-10468.
- CARVALHO, E.; REZENDE, A. J. Data analytics e os novos caminhos da auditoria interna: percepções e desafios. *Revista Universo Contábil*, Blumenau, v. 17, n. 3, p. 30–48, 2021. doi:10.4270/ruc.2021321.
- COSTA, A. M.; LIMA, R. S.; ALCÂNTARA, J. L. Detecção automatizada de anomalias contábeis por meio de algoritmos de machine learning: evidências em bases financeiras brasileiras. *Revista Brasileira de Contabilidade*, Brasília, v. 51, n. 242, p. 45–62, 2022.
- CRESWELL, J. W.; CRESWELL, J. D. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 5. ed. Thousand Oaks: Sage, 2018.
- DAMASCENA, L.; ALMEIDA, J. E. F.; TAVARES, E. Big Data e profissional contábil: desafios e oportunidades na era da informação. *Revista Contemporânea de Contabilidade*, Florianópolis, v. 19, n. 48, p. 75–92, 2022. doi:10.5007/2175-8069.2022.e83825.
- DECHOW, P. M.; GE, W.; SCHRAND, C. M. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, Amsterdam, v. 50, n. 2–3, p. 344–401, 2010. doi:10.1016/j.jacceco.2010.09.001.
- HAIR, J. F. et al. *Multivariate Data Analysis*. 8. ed. New York: Cengage, 2019.
- HE, W.; CASEY, R. J. Earnings management and financial distress: An international perspective. *Journal of Accounting, Auditing & Finance*, London, v. 36, n. 3, p. 572–599, 2021. doi:10.1177/0148558X19877096.
- KUHN, M.; JOHNSON, K. *Feature Engineering and Selection: A Practical Approach for Predictive Models*. Boca Raton: CRC Press, 2021.
- KNECHEL, W. R.; SALTERIO, S.; BALLA, B. *Auditing: Theory and Practice*. New York: Routledge, 2016.
- LIN, C.; HWANG, Y. Machine learning-based detection of financial anomalies: A comprehensive review. *International Journal of Accounting Information Systems*, Amsterdam, v. 51, p. 100619, 2023. doi:10.1016/j.accinf.2023.100619.



MARTINS, A. L. S.; MONTEIRO, R. P. Contabilidade 4.0 e o futuro da profissão contábil: competências, desafios e perspectivas tecnológicas. *Revista Contemporânea de Contabilidade*, Florianópolis, v. 20, n. 54, p. 89–108, 2023. doi:10.5007/2175-8069.2023.e104589.

PUCKETT, A.; BRODY, R. Artificial intelligence and its impact on corporate governance: A governance analytics perspective. *Journal of Forensic & Investigative Accounting*, New Orleans, v. 13, n. 3, p. 415–437, 2021.

RASHID, A.; WANG, S. Predictive analytics in accounting and auditing: A review and research agenda. *International Journal of Accounting Information Systems*, Amsterdam, v. 50, p. 100610, 2023. doi:10.1016/j.accinf.2023.100610.

ROSMAN, P.; SOTO, M.; KHASHMAN, A. Digital transformation and its impact on accounting and auditing education. *Innovations in Education and Teaching International*, London, v. 58, n. 5, p. 620–633, 2021. doi:10.1080/14703297.2020.1815714.

SCHMIDT, P. et al. The digital transformation of the accounting profession: Implications for practitioners and educators. *Accounting Education*, London, v. 29, n. 5, p. 439–466, 2020. doi:10.1080/09639284.2020.1754800.

SHMUELI, G.; KOPPIUS, O. Predictive analytics in information systems research. *MIS Quarterly*, Minneapolis, v. 35, n. 3, p. 553–572, 2011.

SILVA, T. B.; ARAÚJO, L. F.; MENDONÇA, M. F. Data analytics e auditoria interna: oportunidades e barreiras em organizações brasileiras. *Revista de Educação e Pesquisa em Contabilidade*, Brasília, v. 16, n. 4, p. 489–507, 2022. doi:10.17524/repec.v16i4.3030.

TUNG, B.; VASARHELYI, M. A. The evolution from traditional auditing to predictive auditing: Implications for the profession. *International Journal of Disclosure and Governance*, London, v. 20, p. 45–60, 2023. doi:10.1057/s41310-022-00166-7.

VASARHELYI, M. A.; HALPER, F. The continuous audit of online systems. *Auditing: A Journal of Practice & Theory*, Sarasota, v. 10, n. 1, p. 110–125, 1991.

VASARHELYI, M. A.; ROEDEL, K. The future of continuous auditing: Artificial intelligence and real-time assurance. *Journal of Emerging Technologies in Accounting*, Sarasota, v. 19, n. 1, p. 87–103, 2022. doi:10.2308/JETA-2021-054.

WARREN, J. D.; MOFFITT, K. C.; BYRNES, P. How Big Data will change accounting. *Accounting Horizons*, Sarasota, v. 29, n. 2, p. 397–407, 2015. doi:10.2308/acch-51069.

