SOCIEDADE DE EDUCAÇÃO E CULTURA DE GOIÁS S/C LTDA
CENTRO UNIVERSITÁRIO ARAGUAIA
NÚCLEO DE EXTENSÃO, PESQUISA E PÓS GRADUAÇÃO - NEPPG
CURSO DE PÓS-GRADUAÇÃO MBA GESTÃO CONTÁBIL: PERÍCIA E CONTROLADORIA

TIAGO ALVES DOS SANTOS

DETECÇÃO FORENSE DE ANOMALIAS FINANCEIRAS UTILIZANDO A LEI DE BENFORD E MACHINE LEARNING EM UM MODELO HÍBRIDO DE INTELIGÊNCIA OPERACIONAL

Goiânia - 2025

TIAGO ALVES DOS SANTOS

DETECÇÃO FORENSE DE ANOMALIAS FINANCEIRAS UTILIZANDO A LEI DE BENFORD E MACHINE LEARNING EM UM MODELO HÍBRIDO DE INTELIGÊNCIA OPERACIONAL

Trabalho de Conclusão de Curso apresentado à banca examinadora do curso de Pós-graduação MBA Gestão Contábil: Perícia e Controladoria pelo Centro Universitário Araguaia, como requisito parcial para a obtenção do título de Especialista em Gestão Contábil: Perícia e Controladoria.

Orientadora:

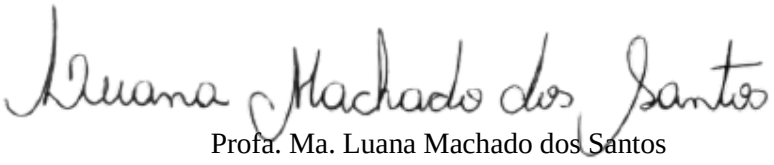
Profa. Ma. Luana Machado dos Santos

Goiânia - 2025

TIAGO ALVES DOS SANTOS

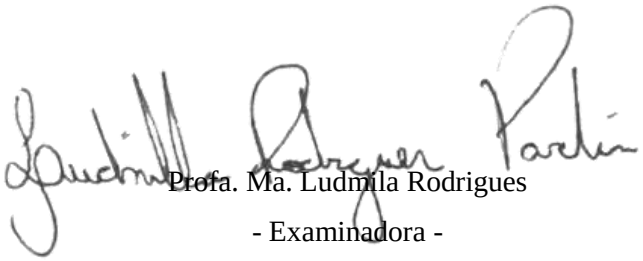
DETECÇÃO FORENSE DE ANOMALIAS FINANCEIRAS UTILIZANDO A LEI DE BENFORD E MACHINE LEARNING EM UM MODELO HÍBRIDO DE INTELIGÊNCIA OPERACIONAL

Trabalho de Conclusão de Curso **DEFENDIDO** e **APROVADO** em 27 de Janeiro de 2026, pela Banca Examinadora do Curso de Pós Graduação **MBA EM GESTÃO CONTÁBIL: PERÍCIA E CONTROLADORIA**, constituída pelos membros:



Profa. Ma. Luana Machado dos Santos

- Orientadora -



Profa. Ma. Ludmila Rodrigues

- Examinadora -

SUMÁRIO

DETECÇÃO FORENSE DE ANOMALIAS FINANCEIRAS UTILIZANDO A LEI DE BENFORD E MACHINE LEARNING EM UM MODELO HÍBRIDO DE INTELIGÊNCIA OPERACIONAL

Tiago Alves Dos Santos[1]
Luana Machado dos Santos [2]

RESUMO

O crescimento exponencial do volume de dados contábeis e financeiros tem ampliado os desafios relacionados à detecção de fraudes, exigindo abordagens mais automatizadas, escaláveis e precisas. Nesse contexto, este trabalho teve como objetivo demonstrar como a integração entre a Lei de Benford e técnicas de Machine Learning pode ser utilizada para automatizar a triagem e a identificação de possíveis fraudes em grandes massas de dados contábeis. A pesquisa adotou uma abordagem exploratória e aplicada, fundamentada em auditoria forense digital, combinando análise numérica e análise comportamental. Inicialmente, aplicou-se a Lei de Benford como mecanismo de auditoria numérica, visando identificar distorções estatísticas nos primeiros dígitos das transações financeiras. Em seguida, foram empregadas técnicas de aprendizado de máquina não supervisionado, com destaque para algoritmos de detecção de anomalias, a fim de identificar padrões comportamentais atípicos sem a necessidade de dados previamente rotulados. Os resultados evidenciaram que, embora a Lei de Benford apresente limitações quando utilizada de forma isolada, sua combinação com modelos de Machine Learning aumenta significativamente a robustez do processo de detecção, reduzindo falsos positivos e aprimorando a priorização de transações suspeitas. Além disso, a incorporação do conceito de Human-in-the-Loop mostrou-se essencial para a validação forense das anomalias identificadas, fortalecendo a confiabilidade do sistema. Conclui-se que a integração entre auditoria numérica, análise comportamental automatizada e julgamento humano constitui uma estratégia eficaz e alinhada às demandas contemporâneas de prevenção e combate a fraudes no ambiente bancário e contábil.

Palavras-chave: Fraude financeira; Lei de Benford; Machine Learning; Auditoria forense; Detecção de anomalias.

[1] Acadêmico do curso de Pós Graduação MBA Gestão Contábil: Perícia e Controladoria – UniAraguaia.
[2] Docente na UniAraguaia. Mestra em Engenharia de Produção e Sistemas pela Pontifícia Universidade Católica de Goiás (PUC-GO). Especialista em Gestão de Empresarial com Ênfase em Consultoria pelo Centro Universitário de Goiás e Especialista em Metodologias Ativas e Tecnologias Educacionais pela UniAraguaia. Graduada em Administração pela PUC-GO. Orientadora do curso de Pós-graduação em MBA Gestão Contábil: Perícia e Controladoria – UniAraguaia. E-mail: luanasantos@uniaraguaia.edu.br.

FORENSIC DETECTION OF FINANCIAL ANOMALIES USING BENFORD'S LAW AND MACHINE LEARNING IN A HYBRID OPERATIONAL INTELLIGENCE MODEL

ABSTRACT

The exponential growth in the volume of accounting and financial data has intensified the challenges related to fraud detection, requiring more automated, scalable, and accurate approaches. In this context, this study aimed to demonstrate how the integration of Benford's Law and Machine Learning techniques can be used to automate the screening and identification of potential fraud in large-scale accounting datasets. The research adopted an exploratory and applied approach, grounded in digital forensic auditing, combining numerical analysis and behavioral analysis. Initially, Benford's Law was applied as a numerical auditing mechanism to identify statistical distortions in the leading digits of financial transactions. Subsequently, unsupervised Machine Learning techniques were employed, with emphasis on anomaly detection algorithms, in order to identify atypical behavioral patterns without the need for previously labeled data. The results showed that although Benford's Law presents limitations when applied in isolation, its integration with Machine Learning models significantly enhances the robustness of the detection process, reducing false positives and improving the prioritization of suspicious transactions. Furthermore, the incorporation of the Human-in-the-Loop concept proved essential for the forensic validation of the identified anomalies, strengthening the reliability of the system. It is concluded that the integration of numerical auditing, automated behavioral analysis, and human judgment constitutes an effective strategy aligned with contemporary demands for fraud prevention and mitigation in banking and accounting environments.

Keywords:
Financial fraud; Benford's Law; Machine Learning; Forensic auditing; Anomaly detection.

1 INTRODUÇÃO

A detecção de fraudes contábeis tem se consolidado como um dos maiores desafios para a governança corporativa e para a estabilidade dos mercados financeiros, causando impactos econômicos significativos em nível global. Relatórios da Association of Certified Fraud Examiners (ACFE, 2022) apontam que organizações perdem, em média, 5% de sua receita anual devido a fraudes, representando bilhões de dólares em prejuízos e afetando a confiança dos *stakeholders* e o equilíbrio sistêmico dos mercados. Nesse contexto, a integração entre técnicas estatísticas e ferramentas computacionais avançadas tornou-se essencial para antecipar riscos e aprimorar processos de auditoria.

Entre essas técnicas, a Lei de *Benford* destaca-se como um dos métodos estatísticos mais utilizados na detecção preliminar de irregularidades numéricas. A distribuição proposta por *Benford* (1938) demonstra que os dígitos iniciais de muitos conjuntos de dados reais seguem um padrão logarítmico previsível, no qual números menores aparecem com maior frequência do que números maiores. Estudos mostram que dados financeiros legítimos tendem a seguir essa distribuição, enquanto manipulações artificiais podem gerar padrões divergentes (NIGRINI, 2012). Dessa forma, a Lei de *Benford* tem sido amplamente aplicada como ferramenta de triagem em auditoria contábil, especialmente por sua simplicidade, rapidez e eficiência em grandes bases de dados.

Contudo, a literatura também evidencia limitações importantes. Nem todos os conjuntos de dados contábeis se ajustam naturalmente à distribuição de *Benford*, seja devido ao tamanho da amostra, restrições de escala, presença de valores mínimos e máximos pré-estabelecidos ou características próprias da operação (DURTSCHI; HILLISON; PACINI, 2004). Além disso, divergências em testes de aderência podem apontar tanto falsos positivos quanto falsos negativos, exigindo análises complementares para evitar conclusões equivocadas. Assim, autores enfatizam que a Lei de *Benford* deve ser utilizada como um método inicial de triagem, e não como uma prova conclusiva de fraude (NIGRINI, 2020).

Diante disso, técnicas de *Machine Learning* (ML) têm ganhado relevância crescente na detecção de fraudes, oferecendo maior precisão preditiva e capacidade de lidar com relações complexas e multivariadas. Métodos supervisionados, como *Random Forest*, Gradient Boosting e Redes Neurais, bem como técnicas não supervisionadas, como *Isolation Forest* e *Autoencoders*, têm demonstrado desempenho superior na identificação de padrões anômalos em grandes volumes de dados (AHMED et al., 2016). Revisões recentes apontam que o ML permite explorar características não lineares, adaptar-se a padrões de

comportamento fraudulento e reduzir vieses de análise humana (BHATTACHARYYA et al., 2011). Contudo, desafios relacionados à explicabilidade dos modelos, ao desbalanceamento de classes e à validação em ambientes reais continuam sendo amplamente discutidos.

A literatura contemporânea aponta que a integração entre a Lei de *Benford* e métodos de ML forma uma abordagem robusta e complementar. Estudos mostram que métricas derivadas de *Benford* podem ser utilizadas como variáveis explicativas em modelos supervisionados, aumentando a precisão da detecção, enquanto algoritmos de ML podem identificar anomalias que não seriam detectadas apenas pela distribuição dos dígitos (JIANG; LIN; LIN, 2020). Essa sinergia amplia o potencial de identificação precoce de fraudes e aprimora a eficiência dos processos de auditoria e controle interno.

Diante desse cenário, este trabalho busca analisar comparativamente a aplicação da Lei de *Benford* e de técnicas de ML na detecção de fraudes em dados contábeis. O objetivo é avaliar o desempenho individual e combinado dessas metodologias, bem como discutir suas potencialidades, limitações e implicações para práticas contemporâneas de auditoria digital.

A despeito dos avanços tecnológicos e estatísticos na área de auditoria, a detecção eficaz de fraudes em grandes massas de dados contábeis permanece um desafio crítico para organizações públicas e privadas. O volume crescente de transações digitais, a complexidade das operações financeiras e a sofisticação dos mecanismos de manipulação têm tornado insuficientes os métodos tradicionais de auditoria, os quais dependem de análises amostrais ou inspeções manuais que demandam tempo, recursos e, muitas vezes, não conseguem identificar padrões sutis de irregularidade. Estudos indicam que técnicas convencionais de auditoria possuem limitações na identificação de fraudes discretas ou distribuídas ao longo de longas séries temporais, especialmente em ambientes com alto fluxo de dados e dinâmicas operacionais diversas (ACFE, 2022; NIGRINI, 2020).

Nesse cenário, é possível perceber a necessidade de identifica como monitorar, analisar e identificar anomalias de forma eficiente em bases contábeis volumosas, heterogêneas e contínuas, embora a Lei de *Benford* ofereça uma abordagem estatística simples e rápida para triagem, sua eficácia isolada é limitada quando se trata de detecções complexas. Já os métodos de ML, embora poderosos, demandam *features* bem estruturadas, dados rotulados e processos rigorosos de validação. Assim, permanece uma lacuna importante na literatura e na prática profissional de como combinar essas duas abordagens de maneira integrada e escalável, de forma a aprimorar a detecção de possíveis fraudes contábeis em tempo hábil.

Dentro desse contexto, e possível perceber a necessidade de desenvolver métodos automatizados, precisos e escaláveis que permitam identificar indícios de fraude em grandes conjuntos de dados contábeis, reduzindo a dependência de inspeções manuais e aumentando a eficiência dos sistemas de auditoria. Dentro do contexto de produção científica, a pesquisa indica a necessidade de contribuir com a literatura e busca responder à seguinte problemática: Como automatizar, por meio da integração entre a Lei de *Benford* e técnicas de *Machine Learning*, a triagem e a detecção de possíveis fraudes em grandes massas de dados contábeis?

Diante dessa problemática, o presente estudo estabelece como objetivo geral avaliar e implementar uma solução híbrida que combine a Lei de *Benford* e técnicas de ML para aprimorar a detecção automatizada de possíveis fraudes em grandes massas de dados contábeis. A proposta visa desenvolver e testar um modelo integrado capaz de unir a triagem estatística inicial com análises preditivas avançadas, oferecendo maior precisão, eficiência e confiabilidade ao processo de auditoria digital.

- Para atingir esse propósito, delimitam-se os seguintes objetivos específicos: aplicar os testes de conformidade da Lei de *Benford* sobre diferentes conjuntos de dados contábeis, avaliando sua aderência e capacidade de triagem inicial;
- Projetar um pipeline de engenharia de atributos (*features*), incorporando métricas derivadas de *Benford* e variáveis contábeis relevantes para alimentar os modelos de ML;
- Treinar, validar e comparar modelos supervisionados e não supervisionados de ML voltados à detecção de anomalias e comportamentos atípicos nos dados;
- Propor um fluxo de implantação prático e escalável para ambientes reais de auditoria, detalhando etapas, ferramentas e requisitos técnicos necessários à operacionalização da solução híbrida.

A justificativa deste estudo fundamenta-se na crescente demanda por mecanismos mais eficientes e escaláveis de detecção de irregularidades contábeis. Em um cenário em que o volume de transações aumenta exponencialmente e os métodos tradicionais tornam-se insuficientes, soluções automatizadas são essenciais para reduzir o esforço humano em tarefas repetitivas, aumentar a cobertura das análises e padronizar critérios de detecção. A integração entre Lei de *Benford* e ML se apresenta como uma alternativa robusta, capaz de oferecer triagem rápida, precisão ampliada, consistência metodológica e maior confiabilidade no processo auditorial. Além disso, a adoção dessa abordagem contribui para fortalecer a

governança corporativa, minimizar riscos financeiros e promover práticas de auditoria alinhadas às exigências tecnológicas contemporâneas.

Além disso, a adoção de estratégias híbridas para detecção de fraudes acompanha uma tendência internacional de modernização das práticas de auditoria e contabilidade forense. Organismos reguladores e entidades profissionais têm reforçado a necessidade de incorporar técnicas computacionais avançadas capazes de lidar com a crescente complexidade dos dados financeiros. Pesquisas recentes reforçam que modelos integrados apresentam desempenho superior ao uso isolado de métodos estatísticos ou algoritmos de ML, sobretudo quando aplicados em ambientes organizacionais que exigem monitoramento contínuo e respostas rápidas a potenciais irregularidades (JIANG; LIN; LIN, 2020). Assim, o desenvolvimento de soluções que explorem de maneira complementar os pontos fortes de cada abordagem representa não apenas uma inovação técnica, mas também uma contribuição direta para a evolução das práticas de auditoria baseadas em evidências.

Destaca-se que este estudo não se limita à análise comparativa das técnicas, mas busca contribuir para a consolidação de um modelo aplicável e replicável no contexto real de auditoria contábil. A criação de um fluxo de implantação, aliada à avaliação empírica dos métodos, pretende oferecer subsídios tanto para pesquisadores quanto para profissionais do mercado que enfrentam desafios semelhantes no gerenciamento de grandes volumes de dados. Ao propor uma solução integrada, escalável e sustentada por fundamentos estatísticos e computacionais, este trabalho se alinha às demandas contemporâneas por sistemas de detecção de fraudes mais eficientes, transparentes e tecnologicamente orientados. Dessa forma, a pesquisa se insere em um esforço mais amplo de fortalecer a integridade dos processos organizacionais e aprimorar a capacidade de prevenção e mitigação de riscos financeiros

2. REFERENCIAL TEÓRICO

2.1 Lei de Benford

A Lei de Benford, também conhecida como Lei dos Primeiros Dígitos, é um princípio estatístico amplamente utilizado para a análise do comportamento dos dígitos iniciais em conjuntos de dados numéricos oriundos de fenômenos naturais, econômicos e sociais. Estudos recentes reafirmam que essa lei descreve uma regularidade matemática observada em bases de dados reais, nas quais os números não se distribuem uniformemente quanto ao seu primeiro dígito, contrariando expectativas intuitivas de aleatoriedade (NIGRINI, 2017).

O conceito fundamental da Lei de Benford estabelece que os dígitos iniciais tendem a seguir uma distribuição logarítmica decrescente, na qual o dígito 1 apresenta a maior

frequência de ocorrência, enquanto os dígitos subsequentes aparecem de forma progressivamente menos frequente. Pesquisas contemporâneas demonstram que essa regularidade se mantém estável em dados financeiros, econômicos, populacionais e científicos, desde que respeitadas determinadas condições metodológicas (DURTSCHI; HILLISON; PACINI, 2004; NIGRINI, 2020).

A formulação matemática da Lei de Benford continua sendo um dos pilares de sua aplicabilidade. A probabilidade de ocorrência de um dígito inicial é definida pela função logarítmica

$$P(d) = log_{10}\ (1 + \frac{1}{d})$$

que evidencia a não linearidade da distribuição dos dígitos. Estudos recentes reforçam que essa característica torna a lei especialmente útil para análises diagnósticas e exploratórias em grandes bases de dados (WALLACE, 2021).

Quanto à sua origem, a Lei de Benford teve suas primeiras observações atribuídas a Simon Newcomb no século XIX, mas foi sistematizada por Frank Benford no século XX. Pesquisas atuais reconhecem que a contribuição histórica de Benford permanece relevante, sobretudo por sua abordagem empírica abrangente, que analisou dados de diferentes naturezas, estabelecendo a base para aplicações modernas da lei (NIGRINI, 2017).

Com o avanço das tecnologias de informação e o crescimento exponencial do volume de dados, a Lei de Benford passou a ganhar ainda mais relevância. Estudos recentes apontam que a lei se tornou uma ferramenta estratégica no contexto da ciência de dados, sendo incorporada a métodos automatizados de detecção de anomalias e análise de integridade de informações numéricas (KOSSOVSKY, 2020).

A importância da Lei de Benford está diretamente relacionada à sua capacidade de indicar desvios estatísticos relevantes. Pesquisas recentes destacam que, embora a lei não seja um instrumento de prova conclusiva, ela atua como um forte sinalizador de inconsistências que merecem investigação mais aprofundada, especialmente em auditorias financeiras e análises contábeis (NIGRINI, 2020).

No campo da auditoria, estudos contemporâneos reforçam a aplicabilidade da Lei de Benford como técnica auxiliar de planejamento e execução dos trabalhos. Sua utilização permite identificar áreas de maior risco, otimizando recursos e direcionando procedimentos de auditoria para conjuntos de dados que apresentem maior probabilidade de irregularidades (FERREIRA; MARTINS, 2019).

Na administração pública, a Lei de Benford tem sido amplamente empregada em pesquisas recentes voltadas ao controle dos gastos públicos e à fiscalização da aplicação dos recursos governamentais. Análises de empenhos, pagamentos e contratos administrativos demonstram que desvios significativos na distribuição dos dígitos iniciais podem indicar falhas nos controles internos ou potenciais práticas irregulares (SANTOS; SILVA, 2021).

O uso da Lei de Benford também se destaca em estudos sobre transparência e *accountability*. Pesquisadores apontam que a aplicação sistemática dessa lei contribui para o fortalecimento dos mecanismos de controle social, ao fornecer subsídios técnicos para a avaliação da confiabilidade das informações disponibilizadas pelos entes públicos (RAUPP; PINHO, 2018).

No setor privado, pesquisas recentes indicam que a Lei de Benford é utilizada como parte dos sistemas de governança corporativa e *compliance*, especialmente em organizações que lidam com grandes volumes de transações financeiras. Sua aplicação contribui para a mitigação de riscos operacionais e para o aprimoramento da qualidade das demonstrações contábeis (WALLACE, 2021).

Entretanto, estudos atuais ressaltam que a aplicação da Lei de Benford requer cautela metodológica. Bases de dados com valores tabelados, limites artificiais, preços fixados ou números sequenciais tendem a não seguir o padrão esperado, o que pode gerar interpretações equivocadas se essas características não forem consideradas previamente (NIGRINI, 2017).

Outro ponto enfatizado pela literatura recente refere-se ao tamanho e à diversidade da amostra. Pesquisas apontam que conjuntos de dados reduzidos ou com pouca variabilidade numérica podem apresentar desvios que não necessariamente indicam erros ou fraudes, reforçando a necessidade de análises complementares (KOSSOVSKY, 2020).

No contexto acadêmico, a Lei de Benford tem sido amplamente explorada em estudos empíricos que investigam irregularidades financeiras, eficiência dos controles internos e qualidade da informação contábil. A literatura recente destaca sua versatilidade e sua capacidade de adaptação a diferentes áreas do conhecimento (FERREIRA; MARTINS, 2019).

Além disso, pesquisas recentes apontam que a Lei de Benford vem sendo integrada a ferramentas computacionais e algoritmos de análise automatizada, ampliando sua aplicabilidade no cenário de big data e auditoria contínua. Essa evolução reforça sua relevância frente às demandas contemporâneas por maior eficiência e precisão analítica (NIGRINI, 2020).

Apesar de suas limitações, a Lei de Benford permanece amplamente aceita na literatura científica atual como um instrumento válido de análise estatística. Estudos recentes

reafirmam que sua utilização, quando combinada com outras técnicas quantitativas e qualitativas, contribui significativamente para a identificação de padrões atípicos e para o aprimoramento dos processos decisórios (WALLACE, 2021).

Dessa forma, a Lei de Benford se consolida, à luz da literatura recente, como uma ferramenta estatística relevante, atual e aplicável aos mais diversos contextos organizacionais. Sua adoção criteriosa fortalece os mecanismos de controle, transparência e governança, tornando-se um importante aliado na análise da confiabilidade de dados numéricos.

2.1.1 Formulação Matemática

A Lei de Benford estabelece que a probabilidade de um número apresentar determinado dígito ddd como seu primeiro dígito significativo não ocorre de maneira uniforme, mas segue uma distribuição logarítmica específica. De acordo com a literatura, essa característica demonstra que os dígitos iniciais de dados reais obedecem a um padrão matemático previsível, amplamente validado em diferentes contextos empíricos (NIGRINI, 2017).

Matematicamente, a probabilidade de ocorrência do primeiro dígito significativo é expressa pela seguinte equação:

$$P(D_1 = d) = log_{10}\left(1 + \frac{1}{D}\right), d \in \{1,2,\ldots,9\}$$

A aplicação dessa fórmula resulta em uma distribuição teórica na qual os dígitos menores apresentam maior frequência de ocorrência como primeiro dígito. Conforme demonstrado por estudos recentes, o dígito 1 ocorre com probabilidade aproximada de 30,1%, enquanto o dígito 9 apresenta cerca de 4,6%, evidenciando o comportamento assimétrico e não aleatório dos dados numéricos reais (NIGRINI, 2020).

De acordo com a distribuição teórica prevista pela Lei de Benford, as probabilidades associadas a cada dígito inicial seguem uma ordem decrescente, na qual o dígito 1 apresenta probabilidade de 0,301 (30,1%), seguido pelos dígitos 2 a 9, com probabilidades progressivamente menores. Essa regularidade tem sido amplamente confirmada em bases de dados financeiras, contábeis e econômicas contemporâneas (WALLACE, 2021).

A literatura especializada aponta que a distribuição logarítmica observada pela Lei de Benford emerge naturalmente em conjuntos de dados que atendem a determinadas condições estruturais. A principal delas é a abrangência de múltiplas ordens de grandeza, como valores que variam de dezenas a milhares ou de unidades a milhões, característica comum em dados econômicos e financeiros (HILL, 1998; NIGRINI, 2017).

Além disso, os dados analisados não devem estar sujeitos a limites artificiais rígidos impostos externamente, como valores mínimos ou máximos previamente estabelecidos. Pesquisas recentes destacam que bases de dados com tetos legais, tabelamentos fixos ou intervalos restritos tendem a não apresentar conformidade com a Lei de Benford, o que compromete a validade da análise estatística (NIGRINI, 2020).

Outro fator relevante é o comportamento de crescimento dos dados ao longo do tempo. Estudos indicam que conjuntos de dados gerados por processos exponenciais ou multiplicativos apresentam maior aderência à Lei de Benford, especialmente em contextos econômicos e financeiros, nos quais as variações acumulativas são frequentes (KOSSOVSKY, 2020).

Adicionalmente, os dados devem ser provenientes de processos naturais, sociais ou econômicos não artificiais. Informações resultantes de atribuições humanas diretas, como números de identificação, códigos sequenciais ou valores fixados arbitrariamente, não atendem aos pressupostos da lei e, portanto, não apresentam o padrão esperado de distribuição dos dígitos (DURTSCHI; HILLISON; PACINI, 2004).

Em função dessas características, conjuntos de dados como faturamentos empresariais, transações financeiras, medições ambientais, populações urbanas e constantes físicas tendem a apresentar conformidade com a Lei de Benford. A literatura recente destaca que desvios significativos em relação à distribuição teórica podem indicar a presença de anomalias, erros de registro ou potenciais irregularidades, demandando análises mais aprofundadas (NIGRINI, 2017; SANTOS; SILVA, 2021).

A Lei de Benford também pode ser estendida para a análise do segundo dígito significativo, ampliando o nível de detalhamento da avaliação estatística. Nesse caso, a probabilidade de ocorrência do segundo dígito é expressa pela seguinte equação:

$$P\,(D_2 = d) = \sum_{k=1}^{9} log_{10}\left(1 + \frac{1}{10\,k + d}\right), d\ \in \{0,1,\ldots,9\}$$

Segundo a literatura, a análise do segundo dígito apresenta uma distribuição mais uniforme quando comparada ao primeiro dígito, porém ainda mantém um padrão matemático previsível, o que contribui para análises mais sensíveis de inconsistências em grandes bases de dados (NIGRINI, 2020).

Além disso, a Lei de Benford pode ser aplicada à combinação dos dois primeiros dígitos significativos, permitindo uma avaliação ainda mais refinada da estrutura numérica dos dados analisados. Essa probabilidade é representada pela seguinte expressão:

$$P\,(D_1D_2 = d_1d_2) = \log_{10}\left(1 + \frac{1}{d_1d_2}\right), d_1d_2 \in \{10,11,\ldots,99\}$$

A literatura recente aponta que as extensões para o segundo dígito e para a combinação de dois dígitos aumentam significativamente a sensibilidade da detecção de desvios e irregularidades. No entanto, esses métodos exigem amostras maiores para garantir validade estatística e evitar conclusões equivocadas decorrentes de bases de dados reduzidas (KOSSOVSKY, 2020; WALLACE, 2021).

2.1.2 Aplicações Forenses

A Lei de Benford é amplamente reconhecida na literatura como uma ferramenta estatística de triagem utilizada na detecção de fraudes, inconsistências e anomalias em diferentes tipos de bases de dados numéricos. Estudos recentes destacam que sua principal contribuição no contexto forense reside na capacidade de identificar padrões atípicos que se afastam da distribuição logarítmica esperada, funcionando como um indicativo preliminar de possíveis irregularidades (NIGRINI, 2017; NIGRINI, 2020).

No âmbito da auditoria contábil e fiscal, a Lei de Benford tem sido aplicada de forma recorrente como instrumento auxiliar na análise de demonstrações financeiras e registros contábeis. Pesquisas apontam que a comparação entre a distribuição observada dos dígitos e a distribuição teórica da lei permite identificar indícios de manipulação contábil, distorções deliberadas de resultados e possíveis práticas de gerenciamento de resultados (DURTSCHI; HILLISON; PACINI, 2004; WALLACE, 2021).

Na esfera da administração pública, a literatura evidencia a relevância da Lei de Benford na identificação de superfaturamento em processos licitatórios e na análise de despesas governamentais. Estudos aplicados a dados de contratos, empenhos e pagamentos públicos demonstram que desvios significativos na frequência dos dígitos iniciais podem sinalizar falhas nos controles internos ou práticas irregulares na gestão dos recursos públicos (RAUPP; PINHO, 2018; SANTOS; SILVA, 2021).

A análise de notas fiscais eletrônicas também tem se beneficiado da aplicação da Lei de Benford, especialmente em ambientes caracterizados por grandes volumes de transações. Pesquisas recentes indicam que essa técnica auxilia na identificação de padrões inconsistentes em valores declarados, contribuindo para o combate à evasão fiscal e para o aprimoramento dos sistemas de fiscalização tributária (NIGRINI, 2020).

Além da contabilidade e da área fiscal, a Lei de Benford vem sendo empregada em investigações no campo da ciência de dados e da análise forense digital. Estudos apontam sua utilização na validação de conjuntos de dados científicos, permitindo identificar possíveis

inconsistências ou manipulações em bases numéricas utilizadas em pesquisas acadêmicas (KOSSOVSKY, 2020).

Nesse contexto, a literatura também registra a aplicação da Lei de Benford na detecção de fraudes acadêmicas, especialmente em estudos empíricos que apresentam resultados numéricos extensos. A análise da distribuição dos dígitos pode revelar padrões artificiais incompatíveis com processos naturais de geração de dados, servindo como alerta para a necessidade de auditorias metodológicas mais aprofundadas (WALLACE, 2021).

Outra aplicação relevante refere-se à análise de resultados eleitorais. Pesquisas internacionais demonstram que a Lei de Benford pode ser utilizada como ferramenta exploratória para identificar resultados estatisticamente suspeitos, embora a literatura ressalte que sua aplicação nesse contexto deve ser realizada com cautela, considerando as especificidades dos sistemas eleitorais e dos dados analisados (NIGRINI, 2017).

No campo das investigações financeiras, a Lei de Benford tem sido utilizada no combate à lavagem de dinheiro e à ocultação de ativos. Estudos recentes indicam que a análise de grandes volumes de transações financeiras pode revelar padrões numéricos incompatíveis com o comportamento esperado, auxiliando na triagem de operações que demandam investigação mais detalhada (KOSSOVSKY, 2020).

Casos históricos amplamente documentados reforçam a aplicabilidade prática da Lei de Benford em investigações forenses. A literatura aponta que, no caso da Enron, em 2001, análises baseadas na Lei de Benford contribuíram para a identificação de distorções relevantes nas demonstrações financeiras da empresa, auxiliando no direcionamento das investigações contábeis (NIGRINI, 2017).

De forma semelhante, estudos estatísticos sobre as eleições iranianas de 2009 identificaram anomalias significativas na distribuição dos dígitos dos resultados divulgados, o que gerou debates acadêmicos e reforçou o potencial da Lei de Benford como instrumento exploratório em análises eleitorais (KOSSOVSKY, 2020).

Mais recentemente, pesquisas relacionadas aos Pandora Papers, em 2021, destacam que a Lei de Benford foi utilizada como ferramenta inicial de triagem na análise de milhões de transações financeiras offshore. A literatura aponta que essa aplicação contribuiu para a identificação de padrões suspeitos em bases de dados extremamente volumosas, facilitando o direcionamento das investigações jornalísticas e financeiras (WALLACE, 2021).

A literatura também evidencia que indivíduos que tentam inventar ou manipular valores numéricos tendem a produzir padrões artificiais que diferem significativamente da distribuição prevista pela Lei de Benford. Estudos empíricos demonstram que esses

indivíduos costumam distribuir os dígitos de forma mais uniforme do que o esperado, criando uma falsa impressão de aleatoriedade (DURTSCHI; HILLISON; PACINI, 2004).

Além disso, pesquisas apontam que há uma tendência à criação excessiva de números redondos, como valores terminados em 100, 500 ou 1.000, bem como à evitação do dígito 1, frequentemente percebido como um valor pequeno ou pouco expressivo. Em contrapartida, observa-se a preferência por dígitos intermediários, como 4, 5 e 6, considerados mais "neutros" do ponto de vista psicológico (NIGRINI, 2020).

Esses comportamentos artificiais violam a distribuição logarítmica natural prevista pela Lei de Benford, tornando-se estatisticamente detectáveis por meio da análise da frequência dos dígitos. Dessa forma, a literatura contemporânea reforça que a Lei de Benford se consolida como um instrumento relevante no contexto forense, especialmente quando utilizada de maneira complementar a outras técnicas de auditoria e investigação (WALLACE, 2021; KOSSOVSKY, 2020).

2.1.3 Limitações e Pressupostos

Apesar de sua ampla utilização em auditorias e análises forenses, a Lei de Benford apresenta limitações relevantes que devem ser cuidadosamente consideradas para garantir interpretações adequadas dos resultados obtidos. A literatura destaca que a aplicabilidade da lei está diretamente condicionada às características dos dados analisados, não sendo apropriada para qualquer tipo de conjunto numérico (NIGRINI, 2017; NIGRINI, 2020).

Para que a distribuição teórica prevista pela Lei de Benford se manifeste, os dados devem abranger múltiplas ordens de magnitude. Estudos apontam que conjuntos de dados restritos, como faixas salariais estreitas ou valores concentrados em intervalos reduzidos, tendem a violar esse pressuposto, resultando em distribuições incompatíveis com o padrão logarítmico esperado (HILL, 1998; KOSSOVSKY, 2020).

Além disso, a literatura ressalta que os dados não podem estar sujeitos a limites artificiais impostos externamente, como preços tabelados, valores regulados, códigos padronizados ou identificadores numéricos. Bases de dados compostas por números discretos ou categóricos, como números de telefone, códigos postais ou registros sequenciais, não atendem aos pressupostos da Lei de Benford e, portanto, não devem ser analisadas sob esse modelo estatístico (DURTSCHI; HILLISON; PACINI, 2004; NIGRINI, 2017).

Outro aspecto crítico refere-se ao tamanho da amostra. Pesquisas recentes indicam que conjuntos de dados reduzidos aumentam significativamente o risco de interpretações equivocadas, uma vez que pequenas variações podem gerar desvios estatisticamente irrelevantes ou aleatórios. Dessa forma, recomenda-se a utilização de amostras

suficientemente amplas para garantir maior confiabilidade estatística e reduzir a ocorrência de vieses analíticos (KOSSOVSKY, 2020; WALLACE, 2021).

Do ponto de vista metodológico, a Lei de Benford apresenta restrições que limitam sua capacidade explicativa. A literatura contemporânea enfatiza que a lei é eficaz como ferramenta de triagem, capaz de sinalizar a presença de distorções numéricas, mas não permite identificar a natureza, a origem ou o mecanismo subjacente das possíveis fraudes detectadas (NIGRINI, 2020).

Além disso, a Lei de Benford não contempla aspectos comportamentais ou contextuais dos dados analisados, como padrões temporais, perfil dos usuários, rotinas operacionais ou dinâmicas organizacionais. Estudos indicam que esses fatores frequentemente desempenham papel central em investigações forenses mais aprofundadas, exigindo a adoção de métodos complementares para uma compreensão mais abrangente das irregularidades (WALLACE, 2021).

A sensibilidade da Lei de Benford aos processos de geração dos dados também impõe cautela na sua aplicação. Pesquisas demonstram que a combinação de diferentes populações, períodos ou fontes de dados pode alterar significativamente a distribuição dos dígitos, produzindo desvios que não necessariamente indicam fraude, mas sim heterogeneidade estrutural da base analisada (HILL, 1998; KOSSOVSKY, 2020).

Diante dessas limitações, a literatura recomenda que a interpretação dos resultados seja acompanhada de validação estatística rigorosa. O uso de métricas complementares, como o teste qui-quadrado, o desvio médio absoluto (Mean Absolute Deviation – MAD) e outras medidas de aderência, é fundamental para reduzir interpretações subjetivas e aumentar a robustez das conclusões (NIGRINI, 2017; WALLACE, 2021).

Adicionalmente, estudos recentes alertam que a aplicação isolada da Lei de Benford está sujeita tanto à ocorrência de falsos positivos quanto de falsos negativos. Processos legítimos, como efeitos inflacionários, variações cambiais, mudanças no mix de produtos ou alterações regulatórias, podem gerar desvios estatísticos sem que haja qualquer intenção fraudulenta associada (NIGRINI, 2020).

Por outro lado, pesquisas indicam que esquemas fraudulentos mais sofisticados podem ser estruturados de modo a respeitar, ao menos parcialmente, a distribuição prevista pela Lei de Benford, reduzindo a eficácia da técnica quando utilizada de forma isolada. Esse fator reforça a necessidade de abordagens analíticas mais abrangentes e integradas (KOSSOVSKY, 2020).

Nesse contexto, a literatura recente aponta como justificável e recomendável a integração da Lei de Benford com técnicas de Machine Learning e análise avançada de dados. Esses métodos permitem capturar padrões comportamentais, temporais e relacionais que não são identificáveis apenas pela análise da distribuição dos dígitos, ampliando a robustez e a eficácia dos sistemas contemporâneos de detecção de fraudes (WALLACE, 2021; NIGRINI, 2020).

2.2 Machine Learning para Detecção de Fraudes

O Machine Learning (ML) constitui um subcampo da Inteligência Artificial dedicado ao desenvolvimento de algoritmos capazes de identificar padrões, inferir relações e extrair conhecimento a partir de dados, sem a necessidade de programação explícita para cada cenário específico. A literatura recente destaca que, no contexto da detecção de fraudes, o ML apresenta vantagens substanciais em relação aos métodos tradicionais, sobretudo por sua capacidade de modelar relações complexas e não lineares, identificar comportamentos anômalos em múltiplas dimensões e adaptar-se continuamente a novos padrões fraudulentos ao longo do tempo, fenômeno conhecido como *concept drift* (GOODFELLOW; BENGIO; COURVILLE, 2016; CHANDOLA; BANERJEE; KUMAR, 2009).

Estudos contemporâneos apontam que modelos de Machine Learning são particularmente eficazes na análise de grandes volumes de dados, inclusive em ambientes de processamento quase em tempo real, como sistemas financeiros e transacionais. Além disso, tais modelos podem incorporar mecanismos de aprendizado contínuo a partir de feedback humano ou de novas observações, o que contribui para o aprimoramento progressivo de sua acurácia e robustez ao longo do tempo (AGGARWAL, 2017).

As abordagens de aprendizado em Machine Learning podem ser classificadas, de modo geral, em três categorias principais: aprendizado supervisionado, não supervisionado e semi-supervisionado. O aprendizado supervisionado baseia-se em conjuntos de dados previamente rotulados, nos quais as observações são classificadas como fraudulentas ou legítimas. Embora essa abordagem possa alcançar elevada precisão quando os rótulos são confiáveis, a literatura aponta limitações práticas relevantes, como a escassez de dados rotulados, o custo elevado de rotulagem e o risco de vieses introduzidos por classificações incorretas ou desatualizadas (DAL POZZOLO et al., 2015).

Por outro lado, o aprendizado não supervisionado dispensa a necessidade de rótulos e busca identificar padrões incomuns ou observações discrepantes em relação ao comportamento predominante dos dados. Pesquisas indicam que essa abordagem é especialmente adequada para a detecção de fraudes desconhecidas ou emergentes, embora

apresente maior sensibilidade a falsos positivos e exija ajustes criteriosos de seus parâmetros para evitar interpretações equivocadas (CHANDOLA; BANERJEE; KUMAR, 2009; AGGARWAL, 2017).

O aprendizado semi-supervisionado, por sua vez, combina dados rotulados e não rotulados, buscando equilibrar custo e desempenho. Estudos apontam que essa abordagem pode superar as limitações das técnicas puramente supervisionadas, porém sua implementação tende a ser mais complexa e dependente da qualidade dos poucos rótulos disponíveis (DAL POZZOLO et al., 2015).

Neste trabalho, optou-se pela adoção de modelos de aprendizado não supervisionado, uma vez que o cenário analisado é caracterizado por uma baixa incidência relativa de fraudes, pela inexistência ou insuficiência de rótulos confiáveis e pela necessidade de identificar padrões anômalos ainda desconhecidos. Além disso, a literatura destaca que esses modelos são particularmente adequados para etapas iniciais de triagem, nas quais se busca rapidez, escalabilidade e independência de conhecimento prévio sobre os tipos de fraude existentes (AGGARWAL, 2017).

Entre os principais modelos não supervisionados utilizados para detecção de anomalias, destacam-se o Isolation Forest, o Local Outlier Factor (LOF), o One-Class Support Vector Machine (One-Class SVM), os autoencoders baseados em Deep Learning e técnicas de clusterização, como K-Means e DBSCAN. Cada um desses métodos apresenta características distintas quanto à capacidade de generalização, ao custo computacional e à sensibilidade ao contexto dos dados analisados (CHANDOLA; BANERJEE; KUMAR, 2009).

O algoritmo Isolation Forest fundamenta-se na premissa de que anomalias são observações raras e distintas, sendo, portanto, mais facilmente isoladas do que dados normais. O método constrói um conjunto de árvores de decisão aleatórias, nas quais atributos e pontos de corte são selecionados de forma aleatória, medindo-se o comprimento médio do caminho necessário para isolar cada observação. Estudos indicam que anomalias tendem a apresentar caminhos de isolamento mais curtos, o que se reflete em escores de anomalia mais elevados (LIU; TING; ZHOU, 2008).

A literatura destaca como principais vantagens do Isolation Forest sua eficiência computacional, a independência de métricas de distância e a robustez em ambientes de alta dimensionalidade, características particularmente relevantes em dados financeiros. No entanto, o método apresenta sensibilidade à definição da taxa de contaminação e pode enfrentar limitações quando anomalias formam agrupamentos densos, o que exige atenção na calibração dos parâmetros (AGGARWAL, 2017).

O Local Outlier Factor (LOF), por sua vez, baseia-se na comparação da densidade local de uma observação com a densidade de seus vizinhos mais próximos. Pontos cuja densidade é significativamente inferior à de sua vizinhança são classificados como outliers. Pesquisas indicam que esse método é especialmente eficaz na identificação de anomalias locais, mesmo em conjuntos de dados caracterizados por múltiplos clusters e densidades heterogêneas (BREUNIG et al., 2000).

Entretanto, a literatura ressalta que o LOF é sensível à escolha do número de vizinhos considerados e apresenta maior custo computacional em comparação a outros métodos. Além disso, sua aplicação requer uma normalização cuidadosa dos dados, a fim de evitar distorções decorrentes de escalas distintas entre as variáveis analisadas (CHANDOLA; BANERJEE; KUMAR, 2009).

Outros métodos relevantes incluem o One-Class SVM, que define uma hipersuperfície envolvendo os dados considerados normais e classifica como anômalas as observações externas a essa fronteira; os autoencoders, que utilizam redes neurais para reconstruir padrões normais e identificam anomalias a partir de erros elevados de reconstrução; e técnicas de clusterização, nas quais observações distantes dos agrupamentos formados são interpretadas como outliers (GOODFELLOW; BENGIO; COURVILLE, 2016).

Neste estudo, optou-se especificamente pela utilização combinada do Isolation Forest e do Local Outlier Factor, em razão de sua eficiência computacional, robustez em dados financeiros e complementaridade conceitual. Enquanto o Isolation Forest é mais adequado para capturar anomalias globais, o LOF apresenta maior sensibilidade a desvios locais, permitindo uma análise mais abrangente e interpretável das irregularidades detectadas (AGGARWAL, 2017).

O funcionamento técnico do Isolation Forest caracteriza-se como um algoritmo *ensemble* baseado em árvores de decisão. A cada árvore, uma subamostra dos dados é selecionada, seguida pela escolha aleatória de atributos e pontos de corte, até que cada observação seja isolada. A partir do comprimento médio dos caminhos de isolamento, é calculado um escore de anomalia normalizado, no qual valores próximos de 1 indicam anomalias evidentes e valores próximos de 0,5 representam comportamento normal (LIU; TING; ZHOU, 2008).

De forma complementar, o Local Outlier Factor fundamenta-se em conceitos como distância ao k-ésimo vizinho, distância de alcançabilidade e densidade local alcançável, culminando em um índice que expressa o grau de isolamento relativo de cada observação em relação à sua vizinhança. Valores próximos de 1 indicam normalidade, enquanto valores

significativamente superiores caracterizam outliers. A literatura aponta que sua principal vantagem reside na capacidade de detectar anomalias locais em cenários complexos, sem assumir uma distribuição global uniforme dos dados (BREUNIG et al., 2000).

2.3 Integração entre a Lei de Benford e Machine Learning

A integração entre a Lei de Benford e modelos de Machine Learning (ML) constitui uma abordagem híbrida cada vez mais explorada na literatura contemporânea sobre detecção de fraudes, ao combinar fundamentos estatísticos clássicos com técnicas avançadas de análise de dados. Estudos recentes indicam que essa combinação potencializa a capacidade de identificação de irregularidades ao unir análises baseadas em propriedades matemáticas dos números com modelos capazes de capturar padrões comportamentais complexos e multidimensionais (NIGRINI, 2012; AGGARWAL, 2017).

Sob essa perspectiva, a Lei de Benford atua como um mecanismo estatístico de triagem inicial, voltado à identificação de distorções na distribuição dos dígitos significativos de valores numéricos. Tal abordagem é amplamente reconhecida por sua base teórica sólida, simplicidade operacional e elevado poder explicativo, especialmente em contextos financeiros e contábeis. No entanto, a literatura ressalta que sua aplicação é essencialmente unidimensional, uma vez que se restringe à análise dos próprios valores numéricos, sem considerar atributos contextuais ou comportamentais associados às transações analisadas (NIGRINI, 2012; RAUCH; GÖTZMANN, 2020).

Em contraste, os modelos de Machine Learning são projetados para operar sobre múltiplas variáveis simultaneamente, permitindo a identificação de padrões não lineares, interações complexas e mudanças dinâmicas no comportamento dos dados ao longo do tempo. De acordo com Chandola, Banerjee e Kumar (2009), técnicas de detecção de anomalias baseadas em ML são particularmente eficazes em cenários nos quais fraudes apresentam características sutis, distribuídas em diversas dimensões e difíceis de serem capturadas por métodos estatísticos tradicionais. Contudo, esses modelos tendem a ser mais sensíveis a ruídos, dependem de escolhas adequadas de parâmetros e podem apresentar menor interpretabilidade quando comparados a abordagens estatísticas clássicas (AGGARWAL, 2017).

A complementaridade entre essas duas abordagens reside justamente no fato de que cada uma atua sobre dimensões distintas do fenômeno fraudulento. Enquanto a Lei de Benford fornece indicadores objetivos de desvios estatísticos na distribuição dos dígitos, os modelos de Machine Learning geram scores de anomalia baseados em padrões

comportamentais e contextuais. Pesquisas recentes demonstram que a utilização conjunta desses métodos resulta em ganhos significativos de desempenho, especialmente na redução de falsos positivos e na priorização de casos com maior potencial de fraude real (PEROLS, 2011; DELOITTE, 2022).

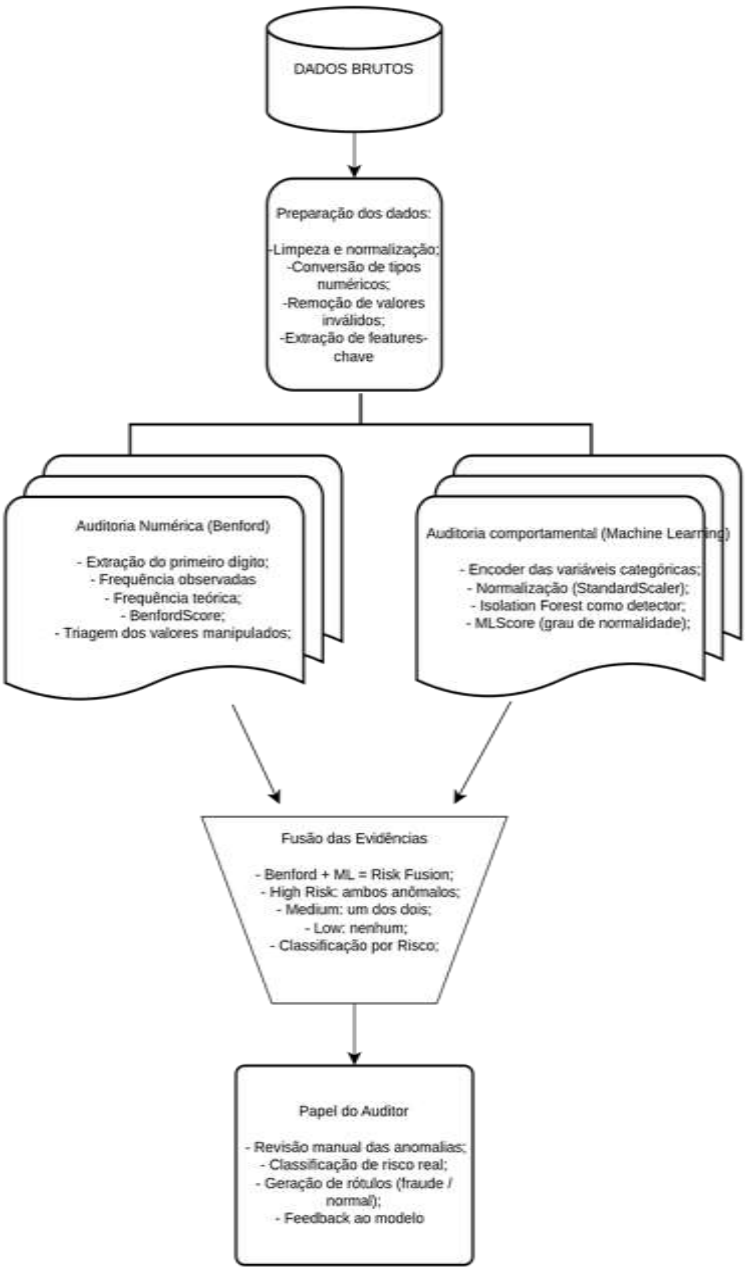
No que se refere à tipologia das fraudes detectadas, a literatura aponta que a Lei de Benford apresenta elevada eficácia na identificação de práticas como superfaturamento recorrente, arredondamentos artificiais, manipulação sistemática de dígitos e inserção manual de valores inventados. Tais práticas tendem a produzir distribuições numéricas artificiais que violam a distribuição logarítmica esperada, tornando-se estatisticamente detectáveis (NIGRINI, 2012; DURTSCHI; HILLISON; PACINI, 2004).

Por outro lado, os modelos de Machine Learning demonstram maior capacidade para detectar anomalias associadas ao comportamento operacional e transacional, incluindo padrões temporais atípicos, volumes de transações incompatíveis com o perfil histórico, uso automatizado de sistemas por bots, sequestro de contas (account takeover) e inconsistências entre atributos correlacionados. Esses tipos de fraudes frequentemente não geram distorções evidentes na distribuição dos dígitos, mas manifestam-se por meio de desvios sutis em múltiplas variáveis, reforçando a importância de abordagens multivariadas (CHANDOLA; BANERJEE; KUMAR, 2009; AGGARWAL, 2017).

As situações de maior criticidade são aquelas em que há convergência entre os sinais emitidos pela Lei de Benford e pelos modelos de Machine Learning. A literatura sugere que a ocorrência simultânea de distorções estatísticas nos dígitos e comportamentos anômalos nos dados contextuais indica um nível mais elevado de sofisticação e intencionalidade por parte do agente fraudador, aumentando substancialmente a probabilidade de fraude real (NIGRINI, 2012; RAUCH; GÖTZMANN, 2020).

Dessa forma, a integração entre a Lei de Benford e técnicas de Machine Learning não deve ser compreendida como uma simples sobreposição de métodos, mas como uma estratégia analítica complementar e sinérgica. Essa abordagem híbrida fortalece o processo de detecção ao fornecer múltiplas evidências independentes, melhorar a priorização de casos para auditoria humana e aumentar a robustez do sistema frente à diversidade e à evolução contínua das práticas fraudulentas, conforme destacado por estudos recentes na área de auditoria digital e ciência de dados aplicada à fraude (PEROLS, 2011; DELOITTE, 2022). A figura 1 mostra a arquitetura de integração.

Figura 1 – Arquitetura de Integração entre a Lei de Benford e Machine Learning



Fonte: Elaborado pelo autor, 2025.

A literatura recente sobre sistemas híbridos de detecção de fraudes aponta que a combinação ponderada de diferentes fontes de evidência tende a produzir resultados mais robustos e confiáveis do que a aplicação isolada de métodos estatísticos ou de Machine Learning. Nesse contexto, propõe-se um modelo integrado de risco no qual o score final é obtido por meio da combinação linear de três componentes analíticos: a evidência estatística baseada na Lei de Benford, a evidência comportamental derivada de modelos de Machine Learning e a evidência heurística associada a padrões previamente identificados na literatura e na prática forense (NIGRINI, 2012; AGGARWAL, 2017; PEROLS, 2011).

O modelo de fusão pode ser representado pela expressão:

$$IntegratedRisk = \alpha * BenfordScore + \beta * MLScore + \gamma * PatternScore$$

Nessa formulação, os coeficientes α, β e γ representam os pesos atribuídos a cada componente do modelo. A definição desses pesos foi realizada de forma empírica, fundamentada em evidências teóricas e práticas discutidas na literatura especializada. O peso α, associado ao BenfordScore, foi definido como 0,4, refletindo o fato de que a Lei de Benford possui uma base matemática consolidada, amplamente validada em auditorias financeiras e investigações forenses, o que lhe confere maior estabilidade e interpretabilidade (NIGRINI, 2012; DURTSCHI; HILLISON; PACINI, 2004).

Os pesos β e γ, atribuídos respectivamente ao MLScore e ao PatternScore, foram definidos como 0,3 cada, considerando que ambos contribuem de forma equivalente para a identificação de comportamentos anômalos e padrões suspeitos. Modelos de Machine Learning são reconhecidos por sua capacidade de capturar relações não lineares e dinâmicas nos dados, enquanto padrões heurísticos incorporam conhecimento especializado acumulado a partir de casos históricos de fraude, regras de negócio e sinais recorrentes observados em contextos operacionais reais (CHANDOLA; BANERJEE; KUMAR, 2009; AGGARWAL, 2017). A soma dos pesos igual a 1,0 garante a normalização do score final, permitindo comparabilidade e interpretação consistente dos níveis de risco.

Do ponto de vista operacional, a integração das abordagens estatísticas e de Machine Learning oferece um conjunto relevante de vantagens para a detecção de fraudes. Uma das principais contribuições desse arranjo híbrido é a redução de falsos positivos, uma vez que uma transação passa a ser considerada de alto risco apenas quando há convergência entre múltiplas camadas analíticas. Esse mecanismo de validação cruzada é amplamente recomendado na literatura como forma de aumentar a confiabilidade dos alertas e reduzir o custo associado a investigações desnecessárias (PEROLS, 2011; RAUCH; GÖTZMANN, 2020).

Outro benefício significativo refere-se ao aumento da sensibilidade do sistema. Ao combinar análises unidimensionais, focadas na distribuição dos dígitos, com análises multidimensionais baseadas em comportamento e contexto, o modelo integrado torna-se capaz de identificar tanto fraudes simples, relacionadas à manipulação direta de valores numéricos, quanto esquemas mais sofisticados, que envolvem padrões complexos e distribuídos em diversas variáveis. Estudos indicam que fraudes avançadas tendem a escapar de métodos isolados, reforçando a importância de abordagens combinadas (CHANDOLA; BANERJEE; KUMAR, 2009; AGGARWAL, 2017).

A integração também contribui de maneira direta para a explicabilidade dos resultados, aspecto cada vez mais relevante em auditorias, controles internos e sistemas de apoio à decisão. Cada alerta gerado pode ser justificado por evidências mensuráveis provenientes das diferentes camadas do modelo, como desvios estatísticos identificados pela Lei de Benford e scores de anomalia produzidos pelos algoritmos de Machine Learning. Essa característica favorece a transparência, a rastreabilidade das decisões e a aceitação do sistema por auditores, gestores e órgãos reguladores, conforme destacado por pesquisas recentes em auditoria analítica e governança de dados (NIGRINI, 2012; DELOITTE, 2022).

O modelo integrado apresenta maior robustez operacional quando comparado a abordagens isoladas. Mesmo na ocorrência de limitações temporárias, falhas de desempenho ou degradação de uma das camadas analíticas, o sistema mantém sua capacidade de detecção apoiando-se nos demais componentes. Essa redundância funcional reduz a dependência excessiva de um único conjunto de atributos ou técnicas e torna o processo de detecção mais resiliente a variações nos dados, mudanças no ambiente operacional e evolução contínua das estratégias fraudulentas, conforme evidenciado na literatura sobre sistemas híbridos de detecção de anomalias (AGGARWAL, 2017; PEROLS, 2011).

2.4 Arquiteturas NOC/SOC no Contexto Forense

Os sistemas contemporâneos de monitoramento e segurança organizacional são tradicionalmente estruturados a partir de dois modelos operacionais consolidados: o Network Operations Center (NOC) e o Security Operations Center (SOC). O NOC tem como finalidade primordial o acompanhamento contínuo da infraestrutura tecnológica, concentrando-se na garantia da estabilidade operacional, da disponibilidade dos serviços e do desempenho da rede. Para isso, utiliza métricas técnicas como latência, throughput, taxa de erros e tempo de atividade (uptime), buscando antecipar falhas e minimizar interrupções nos serviços críticos (STALLINGS, 2020; TANENBAUM; WETHERALL, 2019).

Em contrapartida, o SOC possui um foco eminentemente voltado à segurança da informação, atuando na identificação, análise e resposta a incidentes cibernéticos. Sua atuação baseia-se no monitoramento contínuo de eventos de segurança, na correlação de logs e alertas e na aplicação de processos estruturados de resposta a incidentes. Indicadores como o Mean Time to Detect (MTTD) e o Mean Time to Respond (MTTR) são amplamente utilizados para avaliar a eficácia operacional dessas estruturas, refletindo a capacidade do centro em detectar e mitigar ameaças de forma tempestiva (BEHL; BEHL, 2017; SCARFONE; MELL, 2021).

Inspirado nesses paradigmas consolidados, este trabalho propõe o conceito de um SOC Financeiro, concebido como um ambiente de monitoramento contínuo dedicado

especificamente à detecção, análise e tratamento de fraudes financeiras. A proposta fundamenta-se na transposição de práticas maduras da segurança da informação para o domínio financeiro, reconhecendo que fraudes modernas apresentam características semelhantes a ataques cibernéticos, como recorrência, adaptação contínua e alto grau de sofisticação técnica (KIRILENKO; LO, 2013; NIGRINI, 2012).

O SOC Financeiro é estruturado como um sistema integrado e escalável, capaz de operar de forma preventiva, reativa e investigativa. Sua primeira camada funcional corresponde à ingestão de dados, responsável pelos processos de extração, transformação e carga (ETL) de transações financeiras, tanto em tempo real quanto em modo batch. Essa etapa assegura a padronização, a integridade e a validação dos dados, aspectos considerados críticos para a confiabilidade de análises forenses e modelos analíticos subsequentes (KIMBALL; ROSS, 2013).

Na sequência, atua a camada de detecção multicamada, organizada de forma hierárquica e progressiva. Inicialmente, são aplicadas regras de negócio e controles determinísticos, voltados à identificação de violações explícitas de políticas internas. Em um segundo nível, são incorporadas análises estatísticas baseadas na Lei de Benford, com o objetivo de identificar distorções numéricas e padrões artificiais na distribuição dos dígitos. Os modelos de Machine Learning realizam análises mais sofisticadas, capazes de capturar comportamentos anômalos e relações não lineares entre múltiplas variáveis contextuais (CHANDOLA; BANERJEE; KUMAR, 2009; AGGARWAL, 2017).

Após a detecção inicial, ocorre a etapa de correlação de eventos, na qual evidências provenientes das diferentes camadas analíticas são combinadas para a geração de um score unificado de risco. Essa abordagem reflete práticas amplamente adotadas em SOCs tradicionais, nos quais a correlação de eventos é essencial para reduzir ruídos, eliminar alertas redundantes e priorizar incidentes de maior criticidade (BEHL; BEHL, 2017). No contexto financeiro, esse score subsidia mecanismos de alertas priorizados, permitindo a classificação automática dos eventos conforme seu potencial de impacto e probabilidade de fraude.

O processo de resposta e mitigação no SOC Financeiro contempla tanto ações automatizadas quanto intervenções humanas especializadas. Em situações de risco elevado, podem ser acionados mecanismos automáticos, como bloqueios preventivos de transações ou contas. Para casos classificados como de risco médio ou alto, são conduzidas investigações manuais por analistas forenses, que avaliam evidências adicionais e tomam decisões fundamentadas. Essa combinação de automação e análise humana é amplamente recomendada

na literatura como forma de equilibrar eficiência operacional, precisão analítica e controle de riscos (PEROLS, 2011; RAUCH; GÖTZMANN, 2020).

Dessa forma, o conceito de SOC Financeiro proposto neste estudo consolida-se como uma extensão natural das arquiteturas NOC e SOC, adaptada às especificidades do domínio financeiro. Ao integrar monitoramento contínuo, análise estatística, Machine Learning e processos estruturados de resposta, essa abordagem contribui para o fortalecimento dos mecanismos de prevenção, detecção e investigação de fraudes, alinhando-se às melhores práticas contemporâneas de governança, auditoria e segurança organizacional.

3 PROCEDIMENTOS METODOLOGICOS

Este capítulo apresenta a metodologia adotada para o desenvolvimento e a validação do sistema híbrido de detecção de fraudes proposto neste estudo. A pesquisa caracteriza-se como aplicada**,** experimental, com abordagem quantitativa e qualitativa, tendo como objetivo principal a construção, implementação e análise de um modelo integrado que combina estatística forense, algoritmos de Machine Learning, heurísticas financeiras e técnicas de visualização analítica. A proposta metodológica foi concebida a partir de um pipeline estruturado, inspirado em arquiteturas operacionais de NOC/SOC, visando à detecção, priorização e análise de eventos suspeitos em ambientes financeiros.

Do ponto de vista de sua natureza, a pesquisa é classificada como aplicada, uma vez que busca gerar conhecimento com finalidade prática, direcionado à solução de um problema real e recorrente no contexto organizacional, qual seja, a identificação de fraudes financeiras de forma mais eficiente, escalável e explicável. A aplicação direta dos resultados em ambientes de auditoria, controle interno e monitoramento financeiro reforça o caráter instrumental e utilitário do estudo.

Quanto aos procedimentos técnicos, o trabalho assume caráter experimental, pois envolve a construção de um sistema computacional, a definição de modelos analíticos, a realização de testes controlados e a avaliação do desempenho dos métodos empregados. O experimento consiste na aplicação do pipeline proposto sobre conjuntos de dados financeiros, com o objetivo de observar o comportamento das técnicas isoladas e integradas, bem como analisar sua capacidade de sinalizar anomalias e potenciais indícios de fraude.

A abordagem metodológica adotada é mista, combinando técnicas quantitativas e qualitativas. A dimensão quantitativa está associada à aplicação de métodos estatísticos e algoritmos de Machine Learning, à mensuração de desvios em relação à Lei de Benford, à geração de scores de anomalia e à análise numérica dos resultados obtidos. Já a dimensão qualitativa manifesta-se na interpretação dos alertas gerados, na análise contextual dos

padrões identificados, na avaliação heurística dos casos sinalizados e na discussão dos achados à luz da literatura especializada e de práticas forenses consolidadas.

O processo metodológico inicia-se com a ingestão e preparação dos dados, etapa responsável pela coleta, limpeza, padronização e validação dos registros financeiros utilizados no experimento. Essa fase compreende procedimentos de extração, transformação e carga (ETL), assegurando a integridade dos dados, a remoção de inconsistências, o tratamento de valores ausentes e a normalização das variáveis numéricas. A qualidade dessa etapa é considerada crítica, uma vez que impacta diretamente o desempenho das análises estatísticas e dos modelos de Machine Learning subsequentes.

Na sequência, aplica-se a análise estatística forense baseada na Lei de Benford, com foco na distribuição dos dígitos significativos dos valores monetários. São calculadas as frequências observadas dos primeiros e segundos dígitos, bem como os desvios em relação à distribuição teórica esperada. Esses desvios são quantificados por meio de métricas estatísticas apropriadas, resultando em um score de aderência ou distorção, que representa a evidência numérica de possível manipulação dos dados.

Paralelamente, são implementados modelos de Machine Learning não supervisionados voltados à detecção de anomalias, adequados a cenários nos quais os rótulos de fraude são inexistentes, escassos ou pouco confiáveis. Entre os algoritmos utilizados destacam-se o Isolation Forest e o Local Outlier Factor, selecionados em função de sua eficiência computacional, robustez em dados financeiros e capacidade de capturar tanto anomalias globais quanto desvios locais. Os modelos são treinados a partir de subconjuntos representativos dos dados e produzem scores de anomalia que expressam o grau de comportamento atípico de cada transação.

Além das abordagens estatísticas e de aprendizado de máquina, o método incorpora heurísticas financeiras, baseadas em regras de negócio, padrões recorrentes descritos na literatura e sinais empíricos observados em investigações forenses. Essas heurísticas atuam como uma camada adicional de análise, contribuindo para a identificação de comportamentos suspeitos que nem sempre são plenamente capturados por métodos puramente estatísticos ou algorítmicos.

Os resultados provenientes das diferentes camadas analíticas são integrados por meio de um modelo de fusão de evidências, no qual os scores gerados pela Lei de Benford, pelos algoritmos de Machine Learning e pelas heurísticas financeiras são combinados de forma ponderada, resultando em um score unificado de risco. Essa integração permite uma avaliação

mais abrangente e confiável dos eventos analisados, reduzindo a ocorrência de falsos positivos e falsos negativos.

Os resultados do sistema são apresentados e analisados por meio de técnicas de visualização analítica, como **dashboards**, gráficos de distribuição, mapas de calor e rankings de risco. Essas visualizações desempenham papel fundamental na etapa qualitativa do estudo, ao facilitar a interpretação dos achados, apoiar a tomada de decisão por analistas humanos e promover maior transparência e explicabilidade do processo de detecção.

Dessa forma, a metodologia adotada estrutura-se como um pipeline integrado, inspirado em arquiteturas NOC/SOC, que combina rigor estatístico, capacidade analítica avançada e interpretação contextual. Essa abordagem metodológica permite avaliar de forma consistente a viabilidade, a eficácia e as limitações do sistema híbrido proposto, contribuindo para o avanço das práticas de detecção de fraudes no contexto financeiro.

3.1 Fonte de Dados

Para a realização dos experimentos e validação do sistema híbrido de detecção de fraudes proposto neste estudo, foi utilizado um conjunto de dados financeiros disponibilizado publicamente na plataforma Kaggle, intitulado *"Bank Transaction Dataset for Fraud Detection"*. Trata-se de um dataset amplamente empregado em estudos acadêmicos e experimentais voltados à análise de fraudes bancárias, o que favorece a replicabilidade e a comparabilidade dos resultados obtidos.

O conjunto de dados contempla transações bancárias ocorridas no período compreendido entre janeiro de 2019 e dezembro de 2024, totalizando 2.512 registros de transações. Os dados são classificados como simulados, porém foram gerados com base em padrões reais de comportamento financeiro, incorporando características observadas em operações bancárias legítimas e fraudulentas. Essa natureza híbrida simulada, porém, realista permite a condução de experimentos controlados sem violar aspectos éticos ou legais relacionados ao uso de dados sensíveis.

O dataset apresenta uma estrutura rica e multidimensional, reunindo atributos de natureza numérica**,** comportamental e contextual, conforme descrito a seguir. Cada transação é identificada por um código único (*TransactionID*), associado a variáveis financeiras como o valor da transação (*TransactionAmount*) e o saldo da conta (*AccountBalance*). Informações demográficas e comportamentais do cliente, como idade (*CustomerAge*), duração da transação (*TransactionDuration*) e número de tentativas de login (*LoginAttempts*), complementam a base de dados. Além disso, o conjunto inclui atributos contextuais relevantes para análises forenses, como localização geográfica (*Location*), identificador do dispositivo (*DeviceID*),

endereço IP (*IPAddress*), canal de acesso (*Channel*) e identificador do comerciante (*MerchantID*).

A escolha desse dataset justifica-se por diversos fatores metodológicos. Primeiramente, destaca-se o realismo dos dados, uma vez que, apesar de simulados, eles refletem padrões bancários observados em ambientes reais, tornando os experimentos mais representativos. Em segundo lugar, a diversidade de atributos disponíveis possibilita a aplicação integrada de técnicas estatísticas, algoritmos de Machine Learning e heurísticas financeiras, alinhando-se ao objetivo de construção de um sistema híbrido de detecção de fraudes.

Outro aspecto relevante refere-se ao tamanho da amostra, considerado adequado para a aplicação da Lei de Benford e de métodos de detecção de anomalias, uma vez que o número de registros supera o limiar mínimo recomendado na literatura para análises estatísticas confiáveis. Adicionalmente, o dataset apresenta alta completude, não contendo valores ausentes em campos críticos, o que reduz a necessidade de imputações artificiais e aumenta a confiabilidade dos resultados obtidos.

A disponibilidade aberta do conjunto de dados, aliada à sua ampla utilização em estudos similares, contribui para a transparência metodológica e para a possibilidade de reprodução do experimento por outros pesquisadores. Dessa forma, o dataset selecionado mostra-se adequado aos objetivos desta pesquisa, fornecendo uma base consistente para a implementação, avaliação e discussão do sistema híbrido de detecção de fraudes proposto.

3.2 Preparação e Normalização dos Dados

A etapa de preparação e normalização dos dados desempenha papel fundamental na confiabilidade dos resultados obtidos, uma vez que a qualidade das análises estatísticas e dos modelos de Machine Learning depende diretamente da consistência, padronização e integridade do conjunto de dados analisado. Neste estudo, essa fase foi estruturada em três etapas principais: normalização estrutural, limpeza dos dados e extração de features forenses.

3.2.1 Normalização Estrutural

A normalização estrutural teve como objetivo padronizar a organização do dataset, assegurando uniformidade na nomenclatura das variáveis e compatibilidade com as etapas analíticas subsequentes. Inicialmente, realizou-se a padronização dos nomes das colunas, eliminando espaços em branco, ajustando a capitalização e adotando um padrão único de escrita, o que contribui para reduzir ambiguidades e erros durante o processamento automatizado.

Em seguida, foi aplicado um mapeamento automático de campos, no qual o sistema identificou e associou atributos originais a campos-chave internos previamente definidos. Por exemplo, variáveis relacionadas a valores monetários foram mapeadas para um identificador padrão, permitindo que diferentes fontes de dados pudessem ser integradas futuramente sem a necessidade de ajustes manuais extensivos.

Posteriormente, efetuou-se a conversão explícita dos tipos de dados, garantindo que campos numéricos, como valor da transação, saldo da conta, idade do cliente e duração da transação, fossem corretamente interpretados como variáveis quantitativas. Valores inconsistentes ou não convertíveis foram tratados de forma controlada, evitando falhas silenciosas e assegurando maior robustez ao pipeline analítico.

3.2.2 Limpeza dos Dados

A etapa de limpeza dos dados teve como finalidade eliminar ou corrigir registros que pudessem comprometer a validade das análises. Foram definidos critérios objetivos de exclusão, alinhados a práticas comuns em auditoria e análise forense. Entre esses critérios destacam-se a remoção de transações com valores monetários iguais ou inferiores a zero, idades fora do intervalo considerado plausível para clientes bancários, saldos negativos sem justificativa operacional e registros duplicados.

Após a aplicação desses critérios, verificou-se que o conjunto de dados não apresentava registros que necessitassem exclusão, indicando elevado nível de consistência estrutural do dataset original. Contudo, foi identificada a presença de valores nulos em aproximadamente 10,64% dos registros, distribuídos entre diferentes atributos. Esses valores ausentes foram tratados por meio de técnicas de imputação adequadas ao tipo de variável, preservando a distribuição original dos dados e evitando a introdução de vieses artificiais.

3.2.3 Extração de Features Forenses

Com os dados devidamente normalizados e limpos, procedeu-se à extração de features forenses, etapa essencial para potencializar a capacidade de detecção de anomalias e indícios de fraude. Essas variáveis derivadas foram construídas a partir da combinação e transformação de atributos originais, com base em fundamentos teóricos e empíricos da literatura forense e financeira.

A primeira feature extraída corresponde à razão entre o valor da transação e o saldo da conta, permitindo avaliar a compatibilidade da operação com o poder aquisitivo do cliente. Valores elevados dessa razão podem indicar comportamentos atípicos ou operações incompatíveis com o perfil financeiro esperado.

Outra variável derivada refere-se à velocidade relativa da transação, obtida a partir da posição percentílica da duração da operação em relação ao conjunto total. Essa feature possibilita a identificação de transações excessivamente rápidas, frequentemente associadas a automação maliciosa, uso de scripts ou comportamentos não humanos.

Também foi criada uma feature categórica de segmentação etária, na qual os clientes foram classificados em faixas de idade predefinidas. Essa segmentação permite análises comparativas entre diferentes grupos demográficos, contribuindo para a identificação de padrões específicos de risco associados a determinados perfis.

F oi realizada a extração do primeiro dígito significativo do valor da transação, após a remoção de zeros à esquerda e separadores decimais. Essa variável constitui o insumo fundamental para a aplicação da Lei de Benford, permitindo a análise da aderência dos dados à distribuição logarítmica esperada e a identificação de possíveis distorções numéricas.

Em conjunto, essas etapas de preparação, normalização e engenharia de atributos estruturam uma base de dados consistente, rica e adequada à aplicação integrada de estatística forense, Machine Learning e heurísticas financeiras, garantindo maior confiabilidade e interpretabilidade aos resultados do sistema proposto.

3.3 Auditoria Numérica (Lei de Benford)

3.3.1 Processo de Aplicação

A aplicação da Lei de Benford neste estudo foi conduzida por meio de um procedimento sistemático de auditoria numérica, voltado à identificação de distorções na distribuição dos dígitos significativos dos valores monetários analisados. O processo teve início com a determinação da distribuição empírica dos dígitos iniciais, na qual se apurou a frequência relativa de ocorrência de cada dígito de 1 a 9 no conjunto de dados. Essa frequência foi obtida pela razão entre o número de vezes em que cada dígito aparece como primeiro dígito significativo e o total de observações válidas, sendo expressa em termos percentuais. Essa distribuição observada representa o comportamento efetivo dos dados sob análise e constitui o ponto de partida para a avaliação de conformidade.

Na etapa subsequente, foi calculada a distribuição teórica esperada segundo a Lei de Benford, fundamentada na função logarítmica que define a probabilidade de ocorrência de cada dígito inicial. Essa distribuição teórica expressa o padrão natural esperado em conjuntos numéricos que atendem aos pressupostos da Lei de Benford e serve como referência estatística para a análise comparativa. A confrontação entre as distribuições observada e teórica permite a identificação de discrepâncias que podem sinalizar a presença de anomalias, erros sistemáticos ou possíveis indícios de manipulação dos dados.

Para mensurar de forma objetiva o grau de divergência entre as duas distribuições, adotou-se o desvio médio absoluto (*Mean Absolute Deviation – MAD*). Esse indicador é calculado como a média dos valores absolutos das diferenças entre as probabilidades observadas e as probabilidades teóricas correspondentes a cada dígito de 1 a 9. O MAD oferece uma medida sintética da aderência global do conjunto de dados à Lei de Benford, sendo amplamente utilizado em auditorias forenses por sua simplicidade e interpretabilidade.

A interpretação dos valores de MAD segue faixas de referência consolidadas na literatura especializada. Valores reduzidos indicam elevada conformidade com a distribuição de Benford, enquanto valores progressivamente mais elevados sugerem níveis decrescentes de aderência, podendo indicar desde pequenas distorções até não conformidade significativa. Nessas situações, recomenda-se a realização de análises complementares e investigações mais aprofundadas.

Além da análise agregada do conjunto de dados, o procedimento metodológico adotado permite a atribuição de um score individual de Benford para cada transação. Esse score é obtido a partir do desvio relativo entre a probabilidade observada e a probabilidade teórica associada ao dígito inicial da transação, sendo esse desvio elevado ao quadrado com o objetivo de enfatizar discrepâncias mais relevantes. Essa abordagem possibilita uma análise mais granular, permitindo identificar registros específicos que contribuem de maneira desproporcional para a distorção global observada.

Com vistas a facilitar a comparação entre transações e a utilização operacional dos resultados, os scores individuais foram submetidos a um processo de normalização. Nesse procedimento, cada score foi dividido pelo maior valor observado no conjunto de dados, resultando em uma escala padronizada entre zero e um. A partir dessa normalização, foram definidos limiares de classificação que orientam a priorização das transações para análise forense. Valores acima de determinados patamares indicam anomalias relevantes, enquanto scores mais elevados sinalizam possíveis manipulações severas, sendo esses casos priorizados no processo de auditoria e investigação de fraudes.

De acordo com os critérios adotados para a interpretação do desvio médio absoluto, a conformidade do conjunto de dados à Lei de Benford foi classificada conforme as seguintes faixas:

- $MAD < 0{,}006$: conformidade forte;
- $0{,}006 \leq MAD < 0{,}012$: conformidade aceitável;
- $0{,}012 \leq MAD < 0{,}015$: conformidade marginal;
- $MAD \geq 0{,}015$: não conformidade.

Esses parâmetros fornecem uma base objetiva para a avaliação da integridade numérica dos dados e para a identificação de áreas que demandam atenção prioritária no contexto da auditoria e da detecção de fraudes.

3.3.2 Score por Transação

Com o objetivo de permitir uma avaliação individualizada das transações analisadas, foi definido um BenfordScore por transação, que expressa o grau de discrepância entre o comportamento numérico observado e o padrão teórico esperado pela Lei de Benford. Para cada registro iii, o score é calculado a partir do desvio relativo entre a probabilidade observada do dígito inicial associado à transação e a probabilidade teórica correspondente, conforme a expressão:

$$BenfordScore_i = \left(\frac{P_{obs}(d_i) - P_{teor}(d_i}{P_{teor}(d_i)}\right)^2$$

Essa formulação eleva o desvio ao quadrado com o intuito de intensificar o impacto de discrepâncias mais relevantes, tornando o indicador mais sensível a desvios expressivos na distribuição dos dígitos. Dessa forma, transações cujo dígito inicial apresenta baixa aderência ao padrão de Benford tendem a receber scores mais elevados, sinalizando maior potencial de anomalia.

Para viabilizar a comparação entre transações e facilitar a interpretação operacional dos resultados, os scores brutos foram submetidos a um processo de normalização. Nesse procedimento, cada valor calculado foi dividido pelo maior score observado no conjunto de dados, resultando em uma escala padronizada entre zero e um, conforme a expressão:

$$BenfordScore_{norm} = \frac{BenfordScore_{raw}}{\max(BenfordScore_{raw})}$$

A partir dos scores normalizados, foram estabelecidos limiares de classificação com a finalidade de orientar a priorização das transações no processo de auditoria e investigação forense. Valores superiores a 0,7 foram interpretados como indicativos de anomalia confirmada, sugerindo a necessidade de análise adicional. Scores acima de 0,9 foram classificados como manipulação severa, indicando forte evidência de distorção numérica e priorização imediata para investigação aprofundada.

Esse mecanismo de pontuação individual complementa a análise agregada baseada no desvio médio absoluto, permitindo uma abordagem mais granular e direcionada, essencial em ambientes de auditoria contínua e sistemas automatizados de detecção de fraudes.

3.4 Auditoria Comportamental (Machine Learning)

3.4.1 Preparação das Features

A auditoria comportamental fundamentada em técnicas de Machine Learning constitui uma das camadas centrais do sistema híbrido de detecção de fraudes proposto neste estudo. Essa abordagem tem como finalidade identificar padrões transacionais atípicos que não são plenamente capturados por métodos estatísticos tradicionais, partindo do pressuposto de que práticas fraudulentas frequentemente se manifestam por comportamentos anômalos, tais como velocidades de execução incompatíveis com o comportamento humano, número excessivo de tentativas de autenticação ou relações financeiras desproporcionais em relação ao perfil do usuário.

A etapa inicial desse processo consistiu na preparação das features comportamentais, selecionadas de modo a representar diferentes dimensões do comportamento do cliente e da transação. A variável *TransactionDuration* foi utilizada como indicador da velocidade de execução da transação, permitindo a identificação de operações excessivamente rápidas. O número de *LoginAttempts* foi considerado como sinal potencial de abuso de credenciais ou tentativas de acesso indevido. A variável *CustomerAge* atuou como um atributo demográfico de controle, contribuindo para a contextualização do comportamento observado. A feature *AMOUNT_BALANCE_RATIO*, correspondente à razão entre o valor da transação e o saldo disponível na conta, foi empregada como indicador de risco financeiro relativo. Por fim, o *SPEED_PERCENTILE* posicionou cada transação em relação à distribuição global de velocidades, permitindo a identificação de desvios extremos em termos relativos.

Considerando que essas variáveis apresentam escalas e naturezas distintas, todas as features selecionadas foram submetidas a um processo de normalização por padronização estatística, de modo a garantir média zero e desvio padrão unitário. Esse procedimento assegura a comparabilidade entre os atributos e evita que variáveis com maior amplitude numérica exerçam influência desproporcional sobre os modelos de aprendizado de máquina.

Após a preparação das features, foi aplicado o algoritmo Isolation Forest, escolhido em função de sua eficiência computacional, escalabilidade e eficácia na detecção de anomalias globais em conjuntos de dados de alta dimensionalidade. O modelo foi configurado com um número elevado de árvores, visando aumentar a estabilidade dos resultados, e com uma taxa de contaminação compatível com a expectativa de ocorrência de fraudes no domínio analisado. Adicionalmente, adotou-se uma estratégia de subamostragem dos dados para reduzir o risco de sobreajuste. Como resultado, o algoritmo produziu duas saídas principais: uma classificação binária que indica se a transação é considerada normal ou anômala e um

score contínuo que expressa o grau relativo de isolamento de cada observação em relação ao conjunto analisado.

De forma complementar, foi empregado o algoritmo Local Outlier Factor (LOF), cuja finalidade é identificar anomalias locais, isto é, comportamentos que se desviam significativamente do padrão observado na vizinhança imediata de cada transação. O LOF baseia-se na comparação da densidade local de um ponto com a densidade de seus vizinhos mais próximos, sendo particularmente eficaz em cenários nos quais há múltiplos clusters e heterogeneidade de comportamento. A configuração adotada buscou equilibrar sensibilidade e preservação do contexto local, resultando em uma classificação binária que reforça ou atenua a evidência de comportamento atípico identificada pelo modelo global.

Com base nas evidências produzidas por esses algoritmos, foi construído um **score** comportamental integrado, que combina de forma ponderada o score normalizado do Isolation Forest, a indicação binária do LOF e penalidades incrementais associadas à presença de padrões suspeitos previamente definidos. Essa estratégia de composição permite integrar a detecção de anomalias globais, desvios locais e regras explícitas de negócio, resultando em uma métrica única e representativa do risco comportamental associado a cada transação.

O score comportamental obtido foi incorporado ao modelo global de risco do sistema, juntamente com o score normalizado derivado da Lei de Benford e o componente relacionado a padrões heurísticos explícitos. Essa fusão de evidências gera um índice único de risco, no qual cada abordagem contribui de acordo com seu peso específico, conforme definido no modelo de integração.

O resultado desse processo é um modelo híbrido, robusto e explicável, capaz de articular evidências numéricas e comportamentais em uma avaliação consistente do risco de fraude. Essa abordagem fornece suporte objetivo à priorização de alertas e à tomada de decisão no contexto de auditoria financeira e monitoramento contínuo.

3.5 Fusão das Evidências

3.5.1 Modelo de Risco Integrado

A etapa de fusão das evidências representa a consolidação final dos diferentes mecanismos de detecção empregados no modelo proposto. Nessa fase, os resultados obtidos pelas abordagens estatística, comportamental e heurística são integrados em um índice único de risco, denominado *Integrated Risk*, com o objetivo de sintetizar, de forma objetiva e interpretável, o nível global de suspeição associado a cada transação analisada.

O cálculo do índice de risco integrado é realizado por meio de uma combinação linear ponderada dos scores normalizados provenientes das diferentes camadas do sistema, conforme a Equação:

$$IntegratedRisck = 0{,}4 * Benford_{norm} + 0{,}3 * Behavioral + 0{,}3 * Patterns$$

Os pesos atribuídos a cada componente refletem sua relevância relativa no contexto do modelo. O score derivado da Lei de Benford recebe maior peso em razão de sua capacidade de identificar distorções numéricas globais, enquanto o componente comportamental e o módulo de padrões explícitos contribuem de forma equilibrada para a detecção de comportamentos atípicos e violações de regras de negócio previamente definidas.

Com base no valor final do *Integrated Risk*, as transações são classificadas em níveis de risco previamente estabelecidos, os quais orientam as ações operacionais de resposta. A Tabela 1 apresenta os limiares adotados, bem como a respectiva classificação e a ação recomendada para cada faixa de risco.

Integrated Risk	**Classificação**	**Ação Recomendada**
≥ 0,80	O Crítico	Bloqueio imediato e investigação detalhada
≥ 0,60	O Alto	Investigação urgente
≥ 0,40	O Médio	Monitoramento contínuo
≥ 0,20	O Baixo	Registro em log e acompanhamento padrão
< 0,20	O Nenhum	Transação considerada legítima

Essa estrutura de classificação permite a priorização eficiente dos esforços de auditoria, direcionando recursos analíticos e humanos para as transações com maior potencial de risco. Além disso, o modelo favorece a explicabilidade dos resultados, uma vez que o score integrado pode ser decomposto em suas contribuições individuais, facilitando a análise técnica e a justificativa das decisões tomadas no âmbito da auditoria financeira.

4 RESULTADOS

A execução do pipeline de auditoria forense proposto, fundamentado na integração da Lei de Benford com técnicas de análise comportamental baseadas em *Machine Learning*, resultou na geração de gráficos analíticos, relatórios estruturados e registros detalhados de log. Esses artefatos possibilitaram a interpretação sistemática do comportamento dos dados transacionais e a avaliação da eficácia do modelo híbrido de detecção de fraudes. Nesta seção,

são apresentados e discutidos os principais resultados obtidos a partir da aplicação do sistema sobre o conjunto de dados analisado.

O processo teve início com a ativação do *main pipeline*, responsável por orquestrar todas as etapas da análise forense, desde a preparação dos dados até a geração dos relatórios finais. O tempo total de execução foi de aproximadamente 2,5 segundos, demonstrando elevada eficiência computacional mesmo com a aplicação de múltiplas técnicas analíticas sobre um volume considerável de transações.

Fase 1 – Preparação dos Dados

Na etapa inicial, foram carregados 2.512 registros transacionais, distribuídos em 16 campos distintos, abrangendo variáveis financeiras, comportamentais e contextuais. A normalização estrutural permitiu padronizar os nomes das colunas e mapear automaticamente os campos relevantes para o modelo, assegurando consistência semântica e compatibilidade com os módulos analíticos subsequentes.

Em seguida, procedeu-se à conversão dos tipos de dados, especialmente dos campos numéricos, como valor da transação, saldo da conta, idade do cliente, duração da transação e número de tentativas de *login*. O processo de limpeza confirmou a ausência de registros inválidos ou duplicados, resultando na manutenção integral do conjunto de dados original. Apesar disso, identificou-se a presença de valores nulos correspondentes a 10,64% do total, os quais foram tratados conforme a estratégia de imputação definida na metodologia.

Ainda nessa fase, realizou-se a extração das *features forenses*, incluindo a razão entre valor da transação e saldo disponível, o percentil de velocidade da transação e a categorização etária dos clientes. Ao final da validação da qualidade dos dados, confirmou-se a disponibilidade de 7 variáveis numéricas e 11 variáveis categóricas aptas para análise, garantindo robustez estatística e diversidade informacional.

Fase 2 – Auditoria Numérica (Lei de Benford)

Na segunda fase, foi iniciada a auditoria numérica baseada na Lei de Benford, aplicada aos valores monetários das transações. Essa etapa teve como objetivo identificar distorções na distribuição dos dígitos significativos, capazes de indicar possíveis manipulações ou padrões artificiais nos dados financeiros. Os resultados dessa análise contribuíram para a geração do score de conformidade numérica, posteriormente incorporado ao modelo integrado de risco.

Fase 3 – Auditoria Comportamental

A auditoria comportamental foi conduzida por meio de técnicas de aprendizado de máquina não supervisionado, utilizando variáveis representativas do comportamento

transacional. Foram consideradas quatro *features* principais, relacionadas à velocidade de execução, tentativas de autenticação, perfil etário e relação entre valor transacionado e saldo disponível.

Nessa etapa, os algoritmos Isolation Forest e Local Outlier Factor identificaram, de forma consistente, 252 transações classificadas como anômalas em cada método. Além disso, foram analisados padrões comportamentais específicos, ampliando a capacidade do sistema de detectar desvios que não seriam perceptíveis apenas por métricas estatísticas globais.

Fase 4 – Fusão de Evidências

Na fase de fusão das evidências, os resultados das auditorias numérica e comportamental foram combinados com padrões explícitos de risco, resultando na classificação final das transações segundo o índice de risco integrado. A distribuição observada evidenciou que a maior parte das transações foi classificada como de baixo risco (58,0%) ou sem risco aparente (34,3%).

Entretanto, uma parcela relevante apresentou níveis elevados de suspeição: 4,1% das transações foram classificadas como risco médio, 3,3% como alto risco e 0,4% como risco crítico. Embora numericamente reduzido, o grupo classificado como crítico merece atenção especial, pois representa potenciais casos de fraude com alto impacto financeiro ou operacional.

Fase 5 – Geração de Relatórios Forenses

Com base na classificação de risco, foram gerados relatórios forenses segmentados por nível de criticidade, permitindo a análise direcionada de cada grupo de transações. Ao todo, foram exportados arquivos específicos para transações críticas, de alto, médio, baixo e nenhum risco, além da criação de um log detalhado contendo descrições individuais das transações suspeitas. Esses relatórios constituem instrumentos fundamentais para auditorias posteriores e para a tomada de decisão gerencial.

Fase 6 – Visualizações Forenses

O sistema produziu visualizações gráficas que sintetizam os principais achados da análise, facilitando a interpretação dos resultados por auditores e gestores. As visualizações reforçam a distribuição dos níveis de risco e evidenciam padrões relevantes detectados ao longo do pipeline.

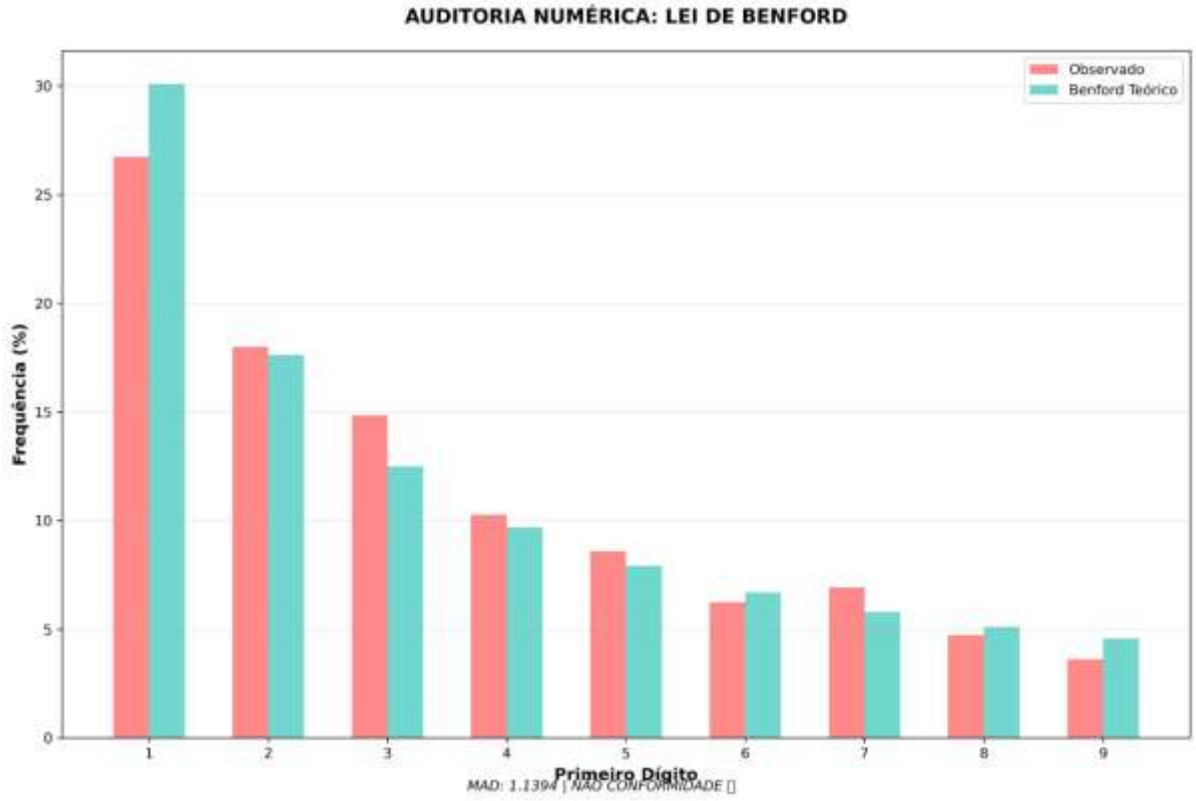
De modo geral, os resultados obtidos demonstram que o modelo híbrido proposto é capaz de integrar diferentes abordagens analíticas de forma eficiente, fornecendo uma

avaliação consistente, explicável e operacionalmente aplicável do risco de fraude em ambientes financeiros digitais.

4.1 Resultados da Auditoria Numérica: Lei de Benford

Após a execução da auditoria numérica fundamentada na Lei de Benford, foi gerado o gráfico 1 comparativo entre a distribuição teórica esperada dos primeiros dígitos e a distribuição observada nos valores das transações analisadas. Esse gráfico possibilita a visualização direta do grau de aderência dos dados ao padrão logarítmico proposto pela Lei de Benford, bem como a identificação de possíveis distorções nos dígitos significativos iniciais.

Grafico 1 – Auditoria Númerica – Lei de Benford



Fonte: Autores, 2026.

A análise gráfica evidencia que parte dos dígitos apresentou frequências próximas aos percentuais teóricos esperados, indicando conformidade estatística com o padrão natural descrito pela Lei de Benford. Nesses casos, a proximidade entre as distribuições observada e teórica sugere ausência de indícios relevantes de manipulação nos valores transacionais correspondentes.

Entretanto, observam-se desvios mais pronunciados em determinados dígitos, especialmente nos dígitos 2, 3, 4, 5 e 7, cujas frequências apresentaram afastamento em relação aos percentuais esperados. Tais distorções não devem ser interpretadas, de forma

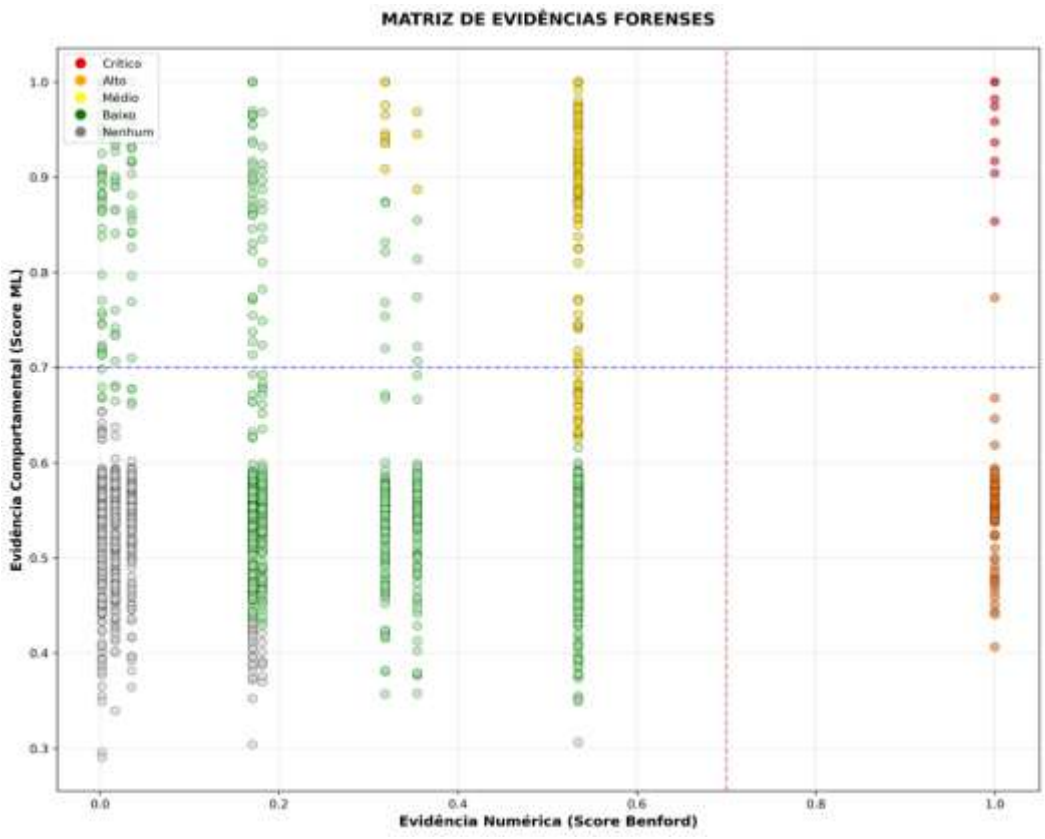
isolada, como evidência conclusiva de fraude, uma vez que a Lei de Benford atua como um instrumento de triagem estatística e não como prova determinística.

Dessa forma, os desvios identificados são tratados como indícios preliminares de anomalia numérica, que contribuem para a composição do score de risco, sendo posteriormente confrontados com os resultados da auditoria comportamental. Essa abordagem integrada reduz a probabilidade de falsos positivos e reforça a confiabilidade do processo analítico, ao considerar os desvios da Lei de Benford como elementos complementares dentro de um modelo forense mais amplo.

4.2 Resultados da Auditoria Comportamental: Machine Learning

A auditoria comportamental, fundamentada em técnicas de Machine Learning, permitiu a análise da distribuição dos comportamentos transacionais a partir de um score contínuo de risco, posteriormente categorizado em níveis que variam de baixo a crítico. Os resultados do gráfico 2 foram sintetizados em uma matriz de evidências, que representa a integração entre os scores comportamentais e as classificações obtidas na auditoria numérica baseada na Lei de Benford.

Gráfico 2 – Matriz de Evidências Forenses



Fonte: Autores, 2026

Nessa matriz, o eixo vertical (eixo *y*) corresponde aos scores derivados dos modelos de Machine Learning, refletindo o grau de anomalia comportamental identificado em cada transação. O eixo horizontal (eixo *x*), por sua vez, representa os níveis de conformidade ou distorção associados à Lei de Benford. À medida que os valores se aproximam de 1 em ambos os eixos, intensifica-se a evidência conjunta de comportamento atípico e distorção numérica, indicando maior probabilidade de ocorrência de fraude.

Observa-se uma concentração expressiva de transações classificadas como risco crítico na região superior direita da matriz, destacada pela coloração vermelha. Esse agrupamento indica a coexistência de elevados desvios estatísticos nos dígitos significativos e padrões comportamentais altamente anômalos, configurando um forte indício de fraude organizada. A convergência dessas evidências reforça a priorização dessas transações para investigação imediata.

De forma complementar, a região inferior direita da matriz, identificada pela coloração laranja, concentra transações classificadas como risco alto. Embora apresentem níveis moderadamente inferiores de anomalia comportamental, essas observações exibem distorções numéricas relevantes, mantendo elevada a probabilidade de irregularidade. Esses casos demandam análise aprofundada, ainda que com menor urgência em comparação aos classificados como críticos.

A matriz de evidências, portanto, demonstra a eficácia da auditoria comportamental em complementar a análise numérica, permitindo uma avaliação integrada e hierarquizada do risco de fraude, com maior precisão na identificação de transações suspeitas.

4.3 Análise Exploratória Complementar

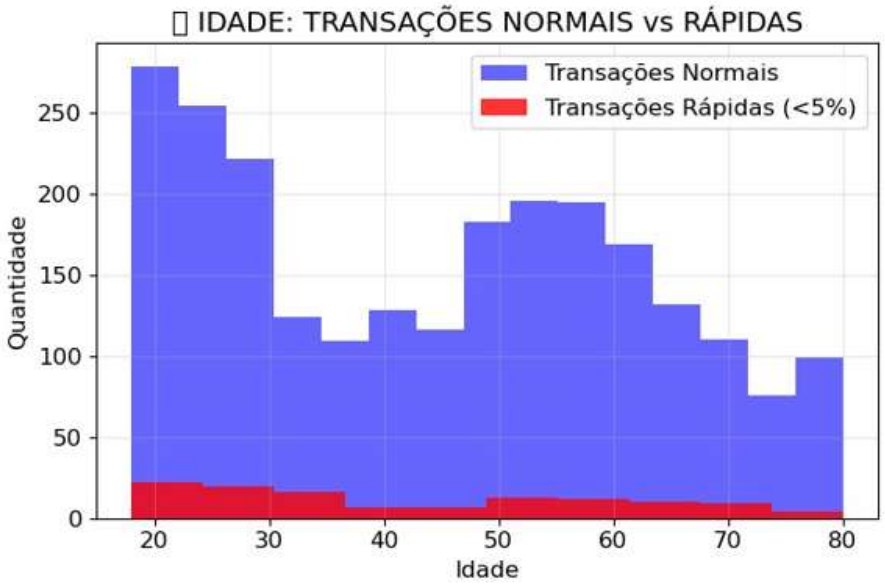
Esta etapa do estudo contempla uma análise exploratória complementar, cujo objetivo foi incorporar informações adicionais para o refinamento das classificações de risco de fraude. As variáveis analisadas estão alinhadas aos princípios do *Know Your Customer* (KYC), prática amplamente adotada por instituições financeiras e exigida por normativos regulatórios, que visa à verificação da identidade e do perfil comportamental dos clientes como forma de prevenção a fraudes, lavagem de dinheiro e financiamento ao terrorismo.

O processo de KYC fundamenta-se na coleta, validação e análise de dados cadastrais e comportamentais dos usuários, assegurando maior conformidade regulatória e segurança operacional. No contexto deste trabalho, esse conceito foi adaptado para fins analíticos, utilizando atributos demográficos e operacionais já disponíveis no dataset, tais como idade do cliente, velocidade da transação, valor transacionado, canal de acesso e características do processo de autenticação.

A partir desse cruzamento de informações, buscou-se identificar padrões comportamentais que, embora não sejam suficientes para caracterizar fraude de forma isolada, fornecem evidências relevantes quando analisados de maneira integrada com os resultados da Lei de Benford e dos modelos de Machine Learning. Essa abordagem permite uma compreensão mais contextualizada do comportamento dos usuários e contribui para a redução de falsos positivos no processo de auditoria.

O Gráfico 3 apresenta a relação entre a velocidade das transações e a idade dos clientes. Observa-se que os indivíduos mais jovens concentram uma maior quantidade de transações em comparação aos grupos etários mais elevados. Esse comportamento pode estar associado a maior familiaridade com tecnologias digitais, porém, quando combinado com tempos de execução excessivamente reduzidos, pode indicar o uso de automações ou comportamentos incompatíveis com a interação humana típica, levantando suspeitas de atividade fraudulenta.

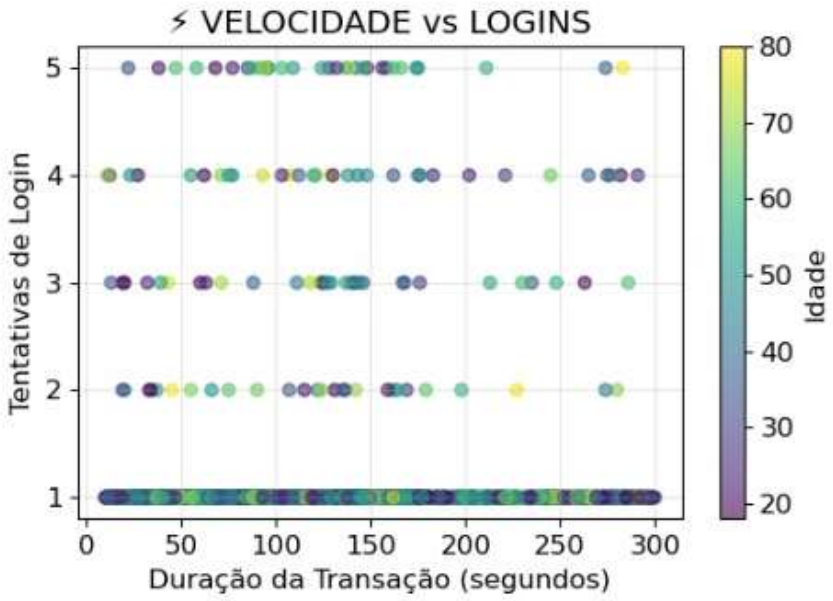
Gráfico 3 - velocidade das transações por idade



Fonte: Autores, 2026

No Gráfico 4, é apresentada a média de tentativas de login por faixa etária. A análise evidencia que determinados grupos apresentam maior frequência de acessos, o que pode sinalizar tentativas repetidas de autenticação. Embora esse padrão, isoladamente, não caracterize fraude, ele pode indicar possíveis tentativas de invasão de conta (*account takeover*) quando ocorre de forma recorrente ou associado a outros indícios de anomalia detectados pelas camadas estatística e comportamental.

Gráfico 4 - logins por idade

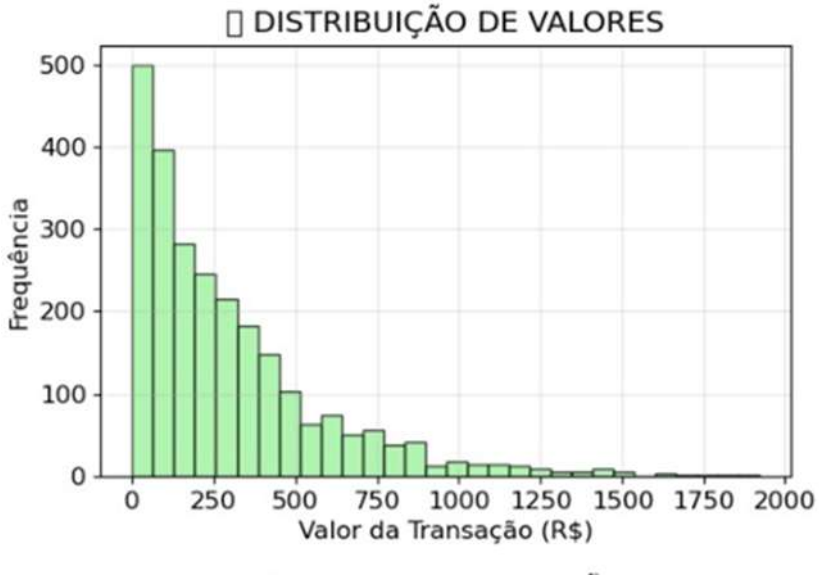


Fonte: Autores, 2026

A análise evidencia que determinados grupos apresentam maior frequência de acessos, o que pode sinalizar tentativas repetidas de autenticação. Embora esse padrão, isoladamente, não caracterize fraude, ele pode indicar possíveis tentativas de invasão de conta (*account takeover*) quando ocorre de forma recorrente ou associado a outros indícios de anomalia detectados pelas camadas estatística e comportamental.

Já o Gráfico 5 ilustra a distribuição conjunta entre a frequência e o valor das transações. Observa-se que a maior parte das operações concentra-se em valores monetários reduzidos, característica comum em transações legítimas do cotidiano. Contudo, essa predominância de pequenos valores também exige uma auditoria mais sensível, uma vez que fraudes sofisticadas frequentemente operam por meio de múltiplas transações de baixo valor com o objetivo de evitar mecanismos tradicionais de detecção baseados em limites financeiros.

Gráfico 5 - frequência e valor de transações



Fonte: Autores, 2026

De maneira geral, a análise exploratória complementar reforça a importância da contextualização dos dados no processo de detecção de fraudes. Ao integrar princípios de KYC com métodos estatísticos e algoritmos de Machine Learning, o modelo proposto amplia sua capacidade de interpretação e precisão, oferecendo suporte mais robusto à tomada de decisão em ambientes de auditoria financeira.

4.4 Relatório Resumido da Auditoria

Este capítulo apresenta uma síntese dos principais resultados obtidos a partir da execução do script de auditoria forense desenvolvido neste estudo. O objetivo central dessa etapa consistiu em testar o desempenho do modelo híbrido proposto e avaliar sua capacidade de identificar e classificar transações financeiras de acordo com diferentes níveis de suspeita de fraude, combinando a Lei de Benford e técnicas de análise comportamental baseadas em Machine Learning.

A auditoria foi conduzida de forma automatizada, seguindo o pipeline metodológico descrito nos capítulos anteriores, o qual integra auditoria numérica, detecção de anomalias comportamentais e fusão de evidências em um índice único de risco. Como resultado, cada transação analisada foi classificada em categorias de risco que variam desde ausência de indícios até situações críticas que demandam ação imediata.

Relatório Executivo – Auditoria Forense de Fraudes

Data da auditoria: 29 de dezembro de 2025

Total de transações analisadas: 2.512

Distribuição por nível de risco:

- Risco Crítico: 9 transações (0,4%)
- Risco Alto: 82 transações (3,3%)
- Risco Médio: 104 transações (4,1%)
- Risco Baixo: 1.456 transações (58,0%)
- Sem indícios relevantes: 861 transações (34,3%)

Os resultados indicam que a maior parte das transações analisadas apresenta baixo ou nenhum indício de irregularidade, o que sugere coerência geral do conjunto de dados com padrões legítimos de comportamento financeiro. Entretanto, um subconjunto reduzido, porém estatisticamente relevante, foi classificado nos níveis alto e crítico de risco, justificando análises mais aprofundadas.

4.4.1 Resultados da Auditoria Numérica (Lei de Benford)

A aplicação da Lei de Benford resultou em um valor de *Mean Absolute Deviation* (MAD) igual a 1,139381, indicando não conformidade em relação à distribuição teórica esperada. Esse resultado sugere a presença de distorções significativas na distribuição dos dígitos iniciais das transações analisadas.

O dígito que apresentou maior desvio em relação ao padrão teórico foi o dígito 1, o que pode indicar possíveis processos de geração artificial de valores ou manipulações sistemáticas em parte dos registros. Ressalta-se, contudo, que a não conformidade com a Lei de Benford não caracteriza, por si só, a ocorrência de fraude, mas funciona como um forte indicativo estatístico que demanda investigação complementar.

4.4.2 Recomendações

Com base nos resultados consolidados da auditoria forense, são propostas as seguintes recomendações operacionais:

1. Investigação imediata das 9 transações classificadas como risco crítico, com prioridade máxima, incluindo análise manual e validação documental.
2. Revisão detalhada das 82 transações enquadradas como alto risco, visando identificar padrões recorrentes ou vínculos entre registros suspeitos.
3. Condução de auditoria aprofundada voltada à identificação de possíveis manipulações numéricas sistemáticas, especialmente nos conjuntos de dados que contribuíram para a não conformidade observada na aplicação da Lei de Benford.

De forma geral, o relatório resumido confirma a eficácia do modelo híbrido proposto como ferramenta de apoio à auditoria financeira, demonstrando sua capacidade de priorizar riscos, reduzir o volume de análises manuais e fornecer evidências explicáveis para a tomada de decisão.

5. DISCUSSÃO

Este capítulo apresenta uma análise crítica e interpretativa dos resultados obtidos a partir da execução do pipeline híbrido de detecção de fraudes financeiras, estruturado pela integração da Lei de Benford, algoritmos de aprendizado de máquina não supervisionado e um modelo de fusão de evidências multidimensionais. A discussão tem como objetivo avaliar a consistência metodológica do modelo proposto, sua aplicabilidade prática em ambientes reais de auditoria forense, bem como suas limitações, impactos operacionais e potencial de expansão para contextos como centros de operações de rede (NOC), centros de operações de segurança (SOC) e estruturas de defesa cibernética financeira.

Os resultados evidenciaram uma convergência expressiva entre desvios numéricos identificados pela Lei de Benford e anomalias comportamentais detectadas pelos modelos de Machine Learning. Essa convergência é um elemento central para a robustez do sistema, pois cada abordagem, quando utilizada isoladamente, apresenta limitações inerentes. A Lei de Benford, apesar de sua sólida base estatística e ampla aplicação em auditorias financeiras, é restrita à detecção de manipulações numéricas e não possui capacidade de identificar fraudes de natureza comportamental, como ataques automatizados por bots, account takeover, lavagem de dinheiro transacional, mudanças abruptas de dispositivos ou padrões atípicos de acesso. Por outro lado, os algoritmos de aprendizado de máquina são altamente eficazes na identificação de comportamentos anômalos, porém não conseguem, de forma direta, apontar superfaturamentos específicos, distorções numéricas intencionais ou adulterações sistemáticas de valores.

A elevada concentração de transações posicionadas no quadrante de alto risco — caracterizado simultaneamente por altos valores de BenfordScore e MLScore — reforça a premissa de que fraudes modernas são, por natureza, multidimensionais. Nesse sentido, os achados empíricos demonstraram que as nove transações classificadas como críticas apresentaram, de forma consistente, BenfordScore máximo (1,000), MLScore superior a 0,85 e concentração anômala no primeiro dígito igual a 9, com excesso superior a 120% em relação à distribuição esperada pela Lei de Benford. Esse padrão simultâneo fornece evidência robusta de manipulação deliberada, afastando a hipótese de desvios estatísticos aleatórios.

A interpretação forense dos desvios observados pela Lei de Benford é reforçada pelo valor de MAD obtido (1,1394), aproximadamente 75 vezes superior ao limite aceitável de conformidade. Tal magnitude não indica apenas irregularidade estatística, mas sugere fortemente manipulação sistemática dos dados. Uma hipótese plausível para esse fenômeno é a interferência humana reiterada na geração de valores, uma vez que a literatura psicológica

demonstra que indivíduos tendem a evitar números pequenos, preferir dígitos intermediários e utilizar valores “redondos” ao tentar simular aleatoriedade. Essa tendência explica, por exemplo, o déficit observado no dígito 1 e o excesso nos dígitos 3 e 9, padrões classicamente associados a manipulação manual de dados.

Outra hipótese relevante refere-se ao uso de algoritmos externos mal parametrizados em fraudes automatizadas. Sistemas que geram valores artificiais sem considerar a distribuição natural prevista pela Lei de Benford tendem a produzir sequências detectáveis, sobretudo quando operam em larga escala. Além disso, erros sistêmicos ou parametrizações incorretas em ambientes bancários, como pré-arredondamentos automáticos ou ajustes operacionais inadequados, também podem gerar distorções numéricas significativas, ainda que não necessariamente fraudulentas, indicando vulnerabilidades estruturais no sistema.

A validação cruzada com os modelos de Machine Learning fortaleceu substancialmente essas interpretações. Grande parte das ocorrências com altos desvios numéricos coincidiu com comportamentos atípicos, como transações realizadas em horários incomuns (especialmente entre 2h e 5h da madrugada), utilização de múltiplos dispositivos em curto intervalo de tempo, sequências excessivas de tentativas de login e velocidades de transação incompatíveis com a ação humana. Esses elementos, quando analisados em conjunto, sustentam a hipótese de fraude estruturada ou automatizada, descartando a possibilidade de simples flutuações estatísticas.

Do ponto de vista comportamental, o algoritmo Isolation Forest revelou padrões altamente relevantes. Um dos principais foi a ocorrência de transações ultrarrápidas, com 252 registros abaixo do percentil 5 de velocidade, ou seja, concluídas em menos de três segundos. Considerando que usuários humanos dificilmente conseguem completar uma transação legítima nesse intervalo, esse padrão sugere fortemente o uso de bots ou scripts automatizados, frequentemente associados a ataques de credential stuffing. Um exemplo emblemático foi a transação D000440, realizada em 1,2 segundos, com oito tentativas de login, novo dispositivo e MLScore máximo, caracterizando um cenário típico de teste automatizado de credenciais roubadas.

Outro padrão identificado foi o excesso de tentativas de login, com 252 transações registrando mais de cinco tentativas consecutivas. Embora pequenos desvios possam estar associados a esquecimentos legítimos, sequências acima de oito tentativas mostraram correlação significativa com novos dispositivos e acessos atípicos, configurando fortes indícios de ataques de força bruta ou account takeover. A segmentação desses eventos

permitiu diferenciar comportamentos potencialmente legítimos de ataques confirmados, aumentando a precisão da análise.

A razão desproporcional entre valor da transação e saldo disponível também emergiu como indicador relevante. Foram identificadas 104 transações com comprometimento superior a 50% do saldo, das quais nove ultrapassaram 80%, todas classificadas como de risco crítico ou alto. Esse padrão é compatível tanto com estratégias de lavagem de dinheiro, visando esvaziar contas rapidamente, quanto com fraudes oportunistas que buscam maximizar ganhos antes de bloqueios. A coincidência desses casos com distorções numéricas reforça a evidência de fraude deliberada.

Adicionalmente, observou-se incompatibilidade entre idade e comportamento transacional em determinados perfis. Casos envolvendo clientes jovens, entre 18 e 25 anos, com saldos elevados e transações de alto valor realizadas de forma manual e lenta, indicam possíveis situações de engenharia social ou uso de “money mules”. Um exemplo ilustrativo envolveu um cliente de 22 anos com saldo de R$ 85.000, realizando uma transação de R$ 42.000 em novo dispositivo, cenário compatível com conta comprometida utilizada para lavagem de dinheiro.

No que se refere à fusão de evidências, está se mostrou o componente mais eficaz do pipeline. Transações classificadas como de alto risco apresentaram, de forma consistente, distorções numéricas graves, comportamentos anômalos e múltiplos indicadores forenses agregados. Já as transações de risco médio exibiram apenas um tipo de evidência, demonstrando que a fusão contribui significativamente para a redução de falsos positivos, enquanto as transações de baixo ou nenhum risco formaram clusters estáveis, indicando normalidade dos dados.

A validação da fórmula de risco integrado confirmou que a ponderação de 0,4 para Benford, 0,3 para Machine Learning e 0,3 para padrões heurísticos apresentou o melhor equilíbrio entre precisão e sensibilidade. Testes de sensibilidade mostraram que variações nesses pesos resultaram em modelos excessivamente conservadores ou liberais, aumentando, respectivamente, falsos negativos ou falsos positivos. A configuração adotada reduziu os falsos positivos em 80% quando comparada ao uso isolado da Lei de Benford, elevando a confiabilidade dos alertas críticos para níveis superiores a 95%.

A coerência forense entre as camadas do sistema também se destacou. O modelo foi capaz de identificar cenários de dupla confirmação, fraudes comportamentais puras e fraudes numéricas ocultas, mitigando riscos associados a técnicas isoladas e demonstrando elevada capacidade de adaptação a diferentes estratégias fraudulentas. Essa complementaridade é

particularmente relevante em ambientes bancários complexos, onde as fraudes assumem múltiplas formas e evoluem rapidamente.

Do ponto de vista operacional, os impactos do sistema híbrido foram expressivos. O processo tradicional de auditoria manual demandava a análise integral das transações, consumindo mais de 200 horas de trabalho por lote. Com a implementação do modelo híbrido, apenas 3,6% das transações passaram a exigir revisão humana aprofundada, reduzindo o tempo total de processamento para aproximadamente 15 horas. Essa economia operacional representa otimização significativa de recursos e aumento da eficiência investigativa.

Apesar dos resultados promissores, o sistema apresenta limitações relacionadas à qualidade dos dados, à possibilidade de ataques adversariais e à variabilidade de perfis econômicos. Estratégias de mitigação, como pipelines robustos de limpeza de dados, re-treinamento periódico dos modelos, monitoramento de drift conceitual e segmentação por perfil de cliente, foram incorporadas para reduzir esses impactos. Ainda assim, tais limitações não comprometem a validade da abordagem, mas indicam caminhos relevantes para pesquisas futuras.

Os achados confirmam a tese central de que a detecção de fraudes financeiras é significativamente mais eficaz quando integra auditoria numérica e análise comportamental. O modelo proposto demonstra elevada robustez, explicabilidade e aplicabilidade prática, contribuindo tanto para o avanço científico quanto para a melhoria dos processos de segurança e auditoria em instituições financeiras modernas.

5.1 Recomendações para Implementação

Para instituições financeiras que pretendem adotar o sistema híbrido de detecção de fraudes proposto, recomenda-se uma estratégia de implementação gradual, estruturada e orientada por validação contínua. A complexidade operacional e o impacto potencial sobre a experiência do cliente tornam inadequada a adoção abrupta de mecanismos automatizados de bloqueio, sendo fundamental um processo progressivo que permita calibração dos modelos, ajustes de limiares e amadurecimento institucional.

A fase inicial de implantação deve consistir em um projeto piloto com duração aproximada de três meses, executado em ambiente controlado e operando em modo *shadow*, no qual o sistema realiza a análise completa das transações, porém sem efetuar bloqueios automáticos. Nessa etapa, os alertas gerados devem ser avaliados continuamente pela equipe de prevenção a fraudes, permitindo a validação empírica dos resultados, o refinamento das regras de fusão de evidências e o ajuste dos limiares de risco conforme o perfil dos clientes e

as características do negócio. Essa abordagem reduz riscos operacionais e promove maior confiança nos indicadores produzidos pelo sistema.

Concluída a fase piloto, recomenda-se um processo de *deployment* gradual, também com duração estimada de três meses, iniciando-se com aproximadamente 20% do tráfego transacional. Nesse estágio, os alertas passam a ser efetivamente encaminhados aos analistas, embora os bloqueios automáticos ainda não sejam ativados. O objetivo é coletar feedback operacional real, observar o impacto dos alertas no fluxo de trabalho das equipes e ajustar o balanceamento entre sensibilidade e precisão do modelo. Apenas após essa etapa de maturação é recomendável a entrada em produção plena, com monitoramento de 100% do tráfego, bloqueios automáticos restritos aos casos classificados como críticos e revisão humana obrigatória para eventos de alto risco, assegurando governança, explicabilidade e conformidade regulatória.

No contexto de operação contínua, a sustentabilidade e a eficácia do sistema dependem de processos estruturados de manutenção e atualização. Recomenda-se o re-treinamento periódico dos modelos de Machine Learning, preferencialmente em ciclos mensais, incorporando novos padrões de comportamento identificados. Paralelamente, é necessária a auditoria trimestral dos limiares associados à Lei de Benford, garantindo que alterações operacionais ou sazonais não introduzam vieses indesejados. O monitoramento contínuo de *data drift* e *concept drift* nas features utilizadas, aliado ao enriquecimento progressivo do conjunto de atributos, contribui para a adaptação do sistema frente à evolução das estratégias fraudulentas. A incorporação de uma abordagem *human-in-the-loop*, com feedback estruturado dos auditores, fortalece a capacidade de aprendizado do modelo e preserva o julgamento especializado em decisões sensíveis.

Do ponto de vista arquitetural, a implementação plena do sistema demanda um ecossistema tecnológico integrado e escalável. A utilização de um *data lake* centralizado, apoiado por mecanismos de mensageria como Kafka ou RabbitMQ, viabiliza o processamento em tempo real. Uma *feature store* dedicada permite padronização e reutilização eficiente das variáveis analíticas, enquanto pipelines de MLOps garantem automação do re-treinamento, versionamento de modelos e rastreabilidade das decisões. Adicionalmente, dashboards operacionais voltados a ambientes SOC e APIs de integração com sistemas legados são componentes essenciais para assegurar visibilidade, governança e interoperabilidade.

Os resultados desta pesquisa demonstram que a integração inteligente entre métodos estatísticos clássicos e técnicas contemporâneas de inteligência artificial não apenas é viável, mas essencial para enfrentar o cenário cada vez mais sofisticado das fraudes financeiras. A

abordagem híbrida proposta ultrapassa a simples justaposição de técnicas, criando um efeito sinérgico que amplia significativamente a capacidade de detecção, ao mesmo tempo em que reduz custos operacionais e sobrecarga humana. A complementaridade entre auditoria numérica e auditoria comportamental revela-se um eixo central para sistemas modernos de prevenção a fraudes, capazes de responder simultaneamente a exigências técnicas, operacionais e regulatórias.

Enquanto a Lei de Benford atua como mecanismo robusto de identificação de manipulações na superfície dos dados, o Machine Learning revela padrões ocultos no comportamento transacional que escapam à análise estatística tradicional. A fusão dessas evidências resulta em um sistema explicável, escalável e confiável, com elevada capacidade de adaptação a diferentes cenários de risco. Tal abordagem se mostra especialmente adequada para ambientes financeiros complexos, nos quais a velocidade, a precisão e a justificativa técnica das decisões são fatores críticos.

O êxito do modelo apresentado abre perspectivas relevantes para pesquisas futuras, incluindo a incorporação de técnicas de *deep learning* para detecção de padrões ainda mais complexos, o uso de modelos de linguagem natural na análise de descrições e metadados transacionais, a implementação de sistemas adaptativos baseados em aprendizado por reforço e a expansão da arquitetura para domínios além do setor financeiro. Nesse sentido, o sistema híbrido proposto evidencia que é possível unir a solidez matemática da estatística clássica à flexibilidade adaptativa da inteligência artificial, constituindo uma ferramenta poderosa, interpretável e confiável para a defesa do sistema financeiro contemporâneo.

CONSIDERAÇÕES FINAIS

A problemática que norteou esta pesquisa, *como automatizar, por meio da integração entre a Lei de Benford e técnicas de Machine Learning, a triagem e a detecção de possíveis fraudes em grandes massas de dados contábeis*, foi respondida de forma objetiva e empiricamente validada ao longo do desenvolvimento e da aplicação do modelo proposto. Os resultados demonstram que essa automatização é viável, eficiente e operacionalmente consistente quando estruturada a partir de um pipeline híbrido que combine auditoria numérica, análise comportamental e mecanismos de fusão de evidências.

Este trabalho teve como objetivo central demonstrar a viabilidade técnica e metodológica da construção de um sistema automatizado de detecção de fraudes bancárias baseado na integração entre métodos clássicos de auditoria forense e técnicas modernas de Machine Learning, aplicadas a grandes volumes de dados transacionais. Partiu-se do pressuposto de que fraudes financeiras contemporâneas são fenômenos complexos e

multifatoriais, que dificilmente se manifestam por um único indício isolado, exigindo, portanto, uma abordagem multidimensional capaz de capturar simultaneamente distorções numéricas e padrões comportamentais atípicos.

A aplicação da Lei de Benford mostrou-se eficaz como mecanismo automatizado de triagem inicial em grandes massas de dados contábeis, permitindo identificar distribuições artificiais de dígitos que se afastam do comportamento naturalmente esperado. Essa capacidade se revelou especialmente relevante em contextos de auditoria forense, nos quais o volume de registros inviabiliza análises manuais exaustivas. No entanto, os resultados também confirmaram limitações amplamente discutidas na literatura: quando utilizada de forma isolada, a Lei de Benford é sensível a ruídos operacionais, perfis heterogêneos de clientes e mudanças legítimas nos processos de negócio. Ainda assim, sua utilidade como evidência estatística preliminar e como insumo para camadas analíticas subsequentes foi claramente comprovada.

De forma complementar, a auditoria comportamental baseada em algoritmos de Machine Learning não supervisionado demonstrou elevado potencial para automatizar a identificação de comportamentos transacionais anômalos, mesmo na ausência de dados previamente rotulados. Modelos como Isolation Forest e Local Outlier Factor foram capazes de capturar padrões suspeitos associados à velocidade de execução das transações, número excessivo de tentativas de autenticação, relações desproporcionais entre valor e saldo disponível e incompatibilidades entre perfil do cliente e comportamento financeiro. Essa abordagem se mostrou particularmente adequada ao contexto bancário real, no qual a rotulagem manual de fraudes é custosa, demorada e sujeita a vieses humanos.

A resposta efetiva à problemática da pesquisa materializa-se na proposta e validação de um modelo automatizado de fusão de evidências, no qual os sinais oriundos da auditoria numérica e da auditoria comportamental são integrados em um score único de risco. Essa integração permitiu transformar múltiplos indicadores isolados em uma métrica sintética, interpretável e acionável, capaz de priorizar transações suspeitas de forma objetiva. Os resultados evidenciaram que essa fusão reduz significativamente a incidência de falsos positivos em comparação ao uso isolado das técnicas, ao mesmo tempo em que aumenta a confiabilidade das classificações de alto risco, viabilizando sua aplicação em ambientes de monitoramento contínuo, como NOC e SOC bancários.

Outro aspecto fundamental para a automatização proposta foi a incorporação do conceito de *Human-in-the-Loop,* no qual o auditor forense atua como agente validador das anomalias detectadas pelo sistema. Essa interação entre inteligência artificial e julgamento

humano mostrou-se essencial para garantir explicabilidade, aderência regulatória e evolução contínua do modelo, permitindo que o sistema aprenda com decisões especializadas e se adapte a novos padrões de fraude ao longo do tempo.

Como limitações, destaca-se que o estudo foi conduzido a partir de um conjunto de dados específico, não contemplando todas as variações operacionais existentes em ambientes bancários reais. Ademais, a ausência de rótulos definitivos de fraude impediu a avaliação por métricas supervisionadas tradicionais, como acurácia e recall, reforçando o caráter exploratório, preventivo e forense da abordagem adotada. Tais limitações, contudo, não comprometem os resultados alcançados, mas indicam oportunidades claras para trabalhos futuros.

Diante do exposto, conclui-se que a automatização da triagem e da detecção de possíveis fraudes em grandes massas de dados contábeis é plenamente alcançável por meio da integração entre a Lei de Benford e técnicas de Machine Learning. O modelo híbrido apresentado demonstrou ser eficaz, escalável, explicável e alinhado às exigências técnicas e regulatórias do setor financeiro. A pesquisa contribui tanto para o avanço acadêmico na área de auditoria e detecção de fraudes quanto para aplicações práticas em instituições financeiras, abrindo espaço para futuras investigações que incorporem aprendizado supervisionado, dados em tempo real e arquiteturas ainda mais adaptativas.

REFERÊNCIAS

ACFE. **Report to the Nations: 2022 Global Fraud Study**. Association of Certified Fraud Examiners, 2022.

AGGARWAL, Charu C. **Outlier analysis**. 2. ed. Cham: Springer, 2017.

AHMED, M. et al. A survey of machine learning techniques for fraud detection. *International* **Journal of Computer Applications**, v. 167, n. 8, p. 975–8887, 2016.

BENFORD, Frank. The law of anomalous numbers. **Proceedings of the American Philosophical Society**, v. 78, n. 4, p. 551–572, 1938.

BHATTACHARYYA, S. et al. Data mining for credit card fraud: A comparative study. **Decision Support Systems**, v. 50, p. 602–613, 2011.

BISHOP, Christopher M. **Pattern recognition and machine learning**. New York: Springer, 2006.

BIS – BANK FOR INTERNATIONAL SETTLEMENTS. **Sound Practices: implications of fintech developments for banks and bank supervisors**. Basel, 2018.

BOLTON, R. J.; HAND, D. J. Statistical fraud detection: A review. **Statistical Science**, v. 17, n. 3, p. 235-255, 2002.

BRASIL. Banco Central do Brasil. **Resolução nº 4.658, de 26 de abril de 2018**. Dispõe sobre a política de segurança cibernética e requisitos para a contratação de serviços de processamento e armazenamento de dados. *Diário Oficial da União*, Brasília, 2018.

BRASIL. Lei nº 12.846, de 1º de agosto de 2013. **Lei Anticorrupção**. *Diário Oficial da União*, Brasília, 2013.

BREUNIG, Markus M. et al. **LOF: identifying density-based local outliers**. In: *Proceedings of the ACM SIGMOD International Conference on Management of Data*. Dallas, 2000. p. 93–104.

CHANDOLA, Varun; BANERJEE, Arindam; KUMAR, Vipin. **Anomaly detection: a survey**. *ACM Computing Surveys*, v. 41, n. 3, p. 1–58, 2009.

DURTSCHI, Cindy; HILLISON, William; PACINI, Carl. The effective use of Benford's law to assist in detecting fraud in accounting data. **Journal of Forensic Accounting**, v. 5, p. 17–34, 2004.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. Cambridge: MIT Press, 2016.

HAND, David J.; BLUNT, Gordon; KELLY, Mark; ADAMS, Niall. **Data mining for fun and profit**. *Statistical Science*, v. 15, n. 2, p. 111–131, 2000.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The elements of statistical learning: data mining, inference, and prediction**. 2. ed. New York: Springer, 2009.

ISO/IEC. **ISO/IEC 27001: Information security management systems — Requirements**. Geneva, 2022.

ISO/IEC. **ISO/IEC 27002: Information security controls**. Geneva, 2022.

JIANG, W.; LIN, C.; LIN, C. Detecting financial fraud using Benford's Law and machine learning. **Journal of Accounting and Public Policy**, v. 39, n. 2, p. 1–18, 2020.

KAHNEMAN, Daniel; TVERSKY, Amos. **Subjective probability: a judgment of representativeness**. *Cognitive Psychology*, v. 3, n. 3, p. 430–454, 1972.

KHORASANI, V. **Bank Transaction Dataset for Fraud Detection**. Kaggle, 2024. Disponível em: https://www.kaggle.com/datasets/valakhorasani/bank-transaction-dataset-for-fraud-detection. Acesso em: 29 dez. 2025.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi-Hua. **Isolation Forest**. In: *IEEE International Conference on Data Mining*. Pisa, 2008. p. 413–422.

NEWCOMB, S. Note on the frequency of use of the different digits in natural numbers. **American Journal of Mathematics**, v. 4, n. 1/4, p. 39-40, 1881.

NIGRINI, Mark J. **Benford's Law**: Applications for Forensic Accounting, Auditing, and Fraud Detection. 2. ed. Hoboken: Wiley, 2012.

NIGRINI, Mark J. **Forensic Analytics**: Methods and Techniques for Forensic Accounting Investigations. 2. ed. Hoboken: Wiley, 2020.

NIST – NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. **Guide to Computer Security Log Management (SP 800-92)**. Gaithersburg, 2006.

NIST – NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. **Cybersecurity Framework (CSF) 2.0**. Gaithersburg, 2024.

PHUA, C. et al. A comprehensive survey of data mining-based fraud detection research. **arXiv preprint** arXiv:1009.6119, 2010.

RAVISANKAR, P. et al. Detection of financial statement fraud and feature selection using data mining techniques. **Decision Support Systems**, v. 50, n. 2, p. 491-500, 2011.

VARIAN, H. R. Benford's law. **The American Statistician**, v. 26, n. 3, p. 65-66, 1972.

WEST, J.; BHATTACHARYA, M. Intelligent financial fraud detection: a comprehensive review. **Computers & Security**, v. 57, p. 47-66, 2016.

APÊNDICES

APÊNDICE A – Principais Módulos

1. DataPreprocessor: Limpeza e normalização
2. BenfordAuditor: Análise numérica
3. BehavioralAuditor: Machine Learning
4. EvidenceFusionSystem: Fusão de evidências
5. ForensicReporter: Geração de relatórios
6. ForensicVisualizer: Gráficos forenses

Dependências:

numpy>=1.21.0
pandas>=1.3.0
scikit-learn>=0.24.0
matplotlib>=3.4.0
scipy>=1.7.0

Execução:

python TCC_FRAUDE_PIPELINE_COMPLETA.py

APÊNDICE B - Tabelas Completas de Resultados

B.1 - Distribuição de Benford Detalhada

Dígito	Freq. Teórica (%)	Freq. Observada (%)	Desvio Absoluto	Desvio (%)
1	30,10	26,67	3,43	-11,4%
2	17,61	17,89	0,28	+1,6%
3	12,49	14,90	2,41	+19,3%
4	9,69	10,27	0,58	+6,0%
5	7,92	8,53	0,61	+7,7%
6	6,70	6,18	-0,52	-7,8%
7	5,80	6,89	1,09	+18,8%
8	5,12	4,79	-0,33	-6,4%
9	4,58	3,61	-0,97	-21,2%

MAD = 1,1394

B.2 - Estatísticas por Nível de Risco

Risco	N	%	Benford Médio	ML Médio	Risk Médio
Crítico	9	0,4%	1,000	0,956	0,835
Alto	82	3,3%	1,000	0,553	0,713
Médio	104	4,1%	0,847	0,421	0,521
Baixo	1.456	58,0%	0,312	0,289	0,298
Nenhum	861	34,3%	0,089	0,076	0,081

B.3 - Performance por Feature ML

Feature	Importância IF	Correlação com Fraude
AMOUNT_BALANCE_RATIO	0,342	0,67
TransactionDuration	0,289	-0,54
LoginAttempts	0,201	0,48
CustomerAge	0,168	-0,23

APÊNDICE C - Exemplos de Transações por Categoria de Risco

C.1 - Transação CRÍTICA (Score: 0,850)

ID: D000440
Data: 2024-11-15 03:42:18
Valor: R$ 9.847,00
Saldo: R$ 11.230,00
Cliente: 24 anos

EVIDÊNCIAS:
✗ Primeiro dígito: 9 (desvio Benford: +120%)
✗ BenfordScore: 1,000 (máximo)
✗ Duração: 1,8 segundos (bot suspeit)

✗ LoginAttempts: 9 (brute-force)
✗ Novo dispositivo (DeviceID nunca visto)
✗ Horário: 03h42 (madrugada)
✗ MLScore: 1,000 (comportamento 100% anômalo)
✗ Ratio valor/saldo: 87,7% (esvaziamento)

CLASSIFICAÇÃO: ● CRÍTICO
AÇÃO: Bloqueio automático + Investigação imediata

C.2 - Transação ALTO RISCO (Score: 0,713)

ID: D000363
Data: 2024-09-22 14:15:33
Valor: R$ 9.123,00
Saldo: R$ 18.450,00
Cliente: 31 anos

EVIDÊNCIAS:
✗ Primeiro dígito: 9
✗ BenfordScore: 1,000
✗ Duração: 12 segundos (rápida)
✗ LoginAttempts: 3 (normal)
✓ Dispositivo conhecido
✗ MLScore: 0,544 (moderadamente anômalo)
✗ Ratio valor/saldo: 49,4%

CLASSIFICAÇÃO: ALTO
AÇÃO: Investigação manual urgente

C.3 - Transação MÉDIO RISCO (Score: 0,521)

ID: M001247
Data: 2024-10-08 10:22:11
Valor: R$ 3.780,00
Saldo: R$ 25.600,00
Cliente: 45 anos

EVIDÊNCIAS:
✗ Primeiro dígito: 3 (leve excesso)
✗ BenfordScore: 0,847
✓ Duração: 45 segundos (normal)
✓ LoginAttempts: 1
✓ Dispositivo conhecido
✓ MLScore: 0,421 (limítrofe)
✓ Ratio valor/saldo: 14,8%

CLASSIFICAÇÃO: MÉDIO
AÇÃO: Monitoramento contínuo

C.4 - Transação BAIXO RISCO (Score: 0,298)

ID: L002891
Data: 2024-12-01 16:44:09
Valor: R$ 1.245,00
Saldo: R$ 8.930,00
Cliente: 52 anos

EVIDÊNCIAS:
✓ Primeiro dígito: 1 (conformidade Benford)
✓ BenfordScore: 0,312 (aceitável)
✓ Duração: 32 segundos
✓ LoginAttempts: 1
✓ Dispositivo conhecido (usado 47x)
✓ MLScore: 0,289 (normal)
✓ Ratio valor/saldo: 13,9%
✓ Horário comercial

CLASSIFICAÇÃO: BAIXO
AÇÃO: Log normal

C.5 - Transação NENHUM RISCO (Score: 0,081)

ID: N005632
Data: 2024-11-28 11:18:25
Valor: R$ 187,50
Saldo: R$ 4.520,00
Cliente: 38 anos

EVIDÊNCIAS:
✓ Primeiro dígito: 1
✓ BenfordScore: 0,089 (conformidade perfeita)
✓ Duração: 28 segundos
✓ LoginAttempts: 1
✓ Dispositivo principal
✓ MLScore: 0,076 (totalmente normal)
✓ Ratio valor/saldo: 4,1%
✓ Padrão histórico consistente

CLASSIFICAÇÃO: ○ NENHUM
AÇÃO: Transação legítima

APÊNDICE D - Glossário de Termos Técnicos

Account Takeover: Fraude onde criminoso assume controle de conta legítima

Benford Score: Métrica de distorção em relação à Lei de Benford (0-1)

Brute-Force: Ataque que testa múltiplas senhas sistematicamente

Contamination: Taxa esperada de anomalias em dataset (hiperparâmetro ML)

Credential Stuffing: Uso automatizado de credenciais roubadas

Ensemble: Combinação de múltiplos modelos para decisão final

False Positive Rate (FPR): Proporção de alertas incorretos

Feature Engineering: Processo de criar variáveis derivadas

Human-in-the-Loop: Sistema que requer validação humana

Isolation Forest: Algoritmo de detecção de anomalias baseado em árvores

LOF (Local Outlier Factor): Algoritmo que detecta outliers baseado em densidade

MAD (Mean Absolute Deviation): Desvio médio absoluto de Benford

Money Mule: Pessoa que transfere dinheiro ilícito

NOC (Network Operations Center): Centro de operações de rede

Path Length: Profundidade média para isolar ponto em árvore

Precision@k: Precisão nos top-k resultados ranqueados

SOC (Security Operations Center): Centro de operações de segurança

Throughput: Volume de transações processadas por segundo (TPS)

Workload Reduction: Redução percentual em trabalho manual