Revista Ibero-Americana de Humanidades, Ciências e Educação — REASE

doi.org/10.51891/rease.v12i11.33448

DETECÇÃO DE PADRÕES ATÍPICOS EM CONTRATAÇÕES PÚBLICAS DO ESTADO DO PIAUÍ: UMA ABORDAGEM COM APRENDIZADO DE MÁQUINA E SHAP

DETECTION OF ATYPICAL PATTERNS IN PUBLIC PROCUREMENT IN THE STATE OF PIAUÍ, BRAZIL: AN APPROACH WITH MACHINE LEARNING AND SHAP

DETECCIÓN DE PATRONES ATÍPICOS EN CONTRATACIONES PÚBLICAS DEL ESTADO DE PIAUÍ, BRASIL: UN ENFOQUE CON APRENDIZAJE AUTOMÁTICO Y SHAP

Vinícius Henrique Albino Andrade[1]
Joselito Mendes de Sousa Junior[2]

RESUMO: A auditoria algorítmica de licitações é empregada pelos órgãos federais de controle no Brasil, mas alcança poucos municípios. Este trabalho avalia se é possível identificar padrões atípicos em contratações públicas do Estado do Piauí sem casos rotulados, com uma explicação para cada sinalização, e quanto o resultado resiste às escolhas de construção. Com 25.757 registros do Portal Nacional de Contratações Públicas (PNCP), quatro padrões adaptados da literatura de controle (valor atípico para o tipo de objeto, concentração de vitórias, empresa recém-constituída e hiperfoco do fornecedor) formaram as variáveis de um modelo Isolation Forest, cujas decisões foram explicadas pelo SHapley Additive exPlanations (SHAP). A avaliação usou variáveis neutras, troca de sementes, testes de sensibilidade, estudos de caso e cadastros de sanções. O conjunto sinalizado é estável entre sementes, e as variáveis de padrão respondem pela maior parte da importância. Sem as transformações adotadas no valor atípico, porém, só 60,4% dos sinalizados se mantêm. Os sinalizados têm fornecedor sancionado com menos frequência (2,48%) que os demais (5,43%), diferença que fica incerta com o valor controlado. Conclui-se que a triagem sem rótulos é possível e explicável, mas depende fortemente das escolhas de construção, e a única validação externa disponível não a confirma.

Palavras-chave: Contratações públicas. Detecção de anomalias. Explicabilidade.

ABSTRACT: Algorithmic auditing of public procurement is used by Brazilian federal oversight bodies but reaches few municipalities. This study assesses whether atypical patterns can be identified in public procurement in the state of Piauí, Brazil, without labeled cases and with an explanation for each flag, and how far the result withstands construction choices. Using 25,757 records from the National Public Procurement Portal (PNCP), four patterns adapted from the oversight literature (atypical value for the type of object, concentration of awards, newly incorporated company and supplier hyperfocus) were represented as features of an Isolation Forest model, whose decisions were explained with SHapley Additive exPlanations (SHAP). The evaluation used neutral features, seed changes, sensitivity tests, case studies and registries of sanctioned companies. The flagged set is stable across seeds, and the pattern features account for most of the importance. Without the adopted transformations of the atypical value, however, only 60.4% of the flagged records remain. Flagged records have a sanctioned supplier less often (2.48%) than the rest (5.43%), a gap that becomes uncertain once contract value is controlled. Screening without labels is possible and explainable, but it depends strongly on construction choices, and the only external validation available does not confirm it.

Keywords: Public procurement. Anomaly detection. Explainability.

[1] Graduando em Engenharia de Software. Instituto de Ensino Superior iCEV, Teresina, PI.
[2] Mestre em Ciência da Computação pela Universidade Federal do Piauí (UFPI). Professor. Orientador. Instituto de Ensino Superior iCEV, Teresina, PI.

RESUMEN: La auditoría algorítmica de contrataciones públicas la emplean los órganos federales de control en Brasil, pero alcanza a pocos municipios. Este trabajo evalúa si es posible identificar patrones atípicos en contrataciones públicas del estado de Piauí, Brasil, sin casos etiquetados, explicando cada señalamiento, y en qué medida el resultado resiste las decisiones de construcción. Con 25.757 registros del Portal Nacional de Contrataciones Públicas (PNCP), cuatro patrones de la literatura de control (valor atípico para el tipo de objeto, concentración de adjudicaciones, empresa recién constituida e hiperfoco del proveedor) formaron las variables de un modelo Isolation Forest, cuyas decisiones se explicaron con SHapley Additive exPlanations (SHAP). La evaluación utilizó variables neutras, cambio de semillas, pruebas de sensibilidad, estudios de caso y registros de sanciones. El conjunto señalado es estable entre semillas, y las variables de patrón concentran la mayor parte de la importancia. Sin las transformaciones del valor atípico, solo se mantiene el 60,4% de los señalados. Los señalados tienen proveedor sancionado con menor frecuencia (2,48%) que los demás (5,43%), diferencia incierta al controlar el valor. El cribado sin etiquetas es posible y explicable, pero depende fuertemente de las decisiones de construcción, y la única validación externa disponible no lo confirma.

Palabras clave: Contrataciones públicas. Detección de anomalías. Explicabilidad.

1 INTRODUÇÃO

As licitações são o principal instrumento pelo qual a administração pública brasileira contrata bens, obras e serviços. A Lei nº 14.133/2021, que substituiu a Lei nº 8.666/1993, tornou regra a contratação eletrônica e criou o Portal Nacional de Contratações Públicas (PNCP). O portal é o sítio oficial em que editais, contratos e atas de registro de preços de todos os entes federativos devem ser divulgados de forma centralizada (Brasil, 2021). Com o portal, as compras de municípios, estados e União passaram a estar disponíveis em um único repositório, acessível por interfaces de programação de aplicações (APIs) públicas.

Essa disponibilidade de dados abre espaço para a chamada auditoria algorítmica de licitações: o uso de técnicas computacionais, como regras estatísticas, mineração de textos e aprendizado de máquina, para examinar a totalidade das contratações. O objetivo é priorizar as que apresentam indícios de risco (Borges *et al.*, 2026; Maia *et al.*, 2020). No Brasil, a experiência mais conhecida é a do *software* Analisador de Licitações, Contratos e Editais (ALICE), desenvolvido pela Controladoria-Geral da União (CGU). O *software* faz uma auditoria contínua das contratações por mineração de textos e cruzamento de dados e emite alertas aos auditores (Borges *et al.*, 2026). Somente em 2024, o *software* ALICE subsidiou 212 auditorias sobre compras que somavam R$ 30,57 bilhões (Brasil. Controladoria-Geral da União, 2024a). O Tribunal de Contas da União (TCU) também emprega soluções de inteligência artificial em seus processos de auditoria, como o Sistema de Orientação sobre Fatos e Indícios para o Auditor (SOFIA) e o Monitoramento Integrado para o Controle de Aquisições (MONICA) (Neves; Silva; Carvalho, 2019).

Essas soluções, porém, foram concebidas para o controle de contratações federais. A disponibilização do *software* ALICE a estados e municípios foi anunciada apenas em maio de 2024 (Brasil. Controladoria-Geral da União, 2024b), e, naquele ano, dos 126 entes cadastrados para receber os alertas, somente 61 eram municipais (Brasil. Controladoria-Geral da União, 2024a). Esses números indicam que a maior parte dos municípios brasileiros ainda não recebe os alertas do *software*.

Na literatura acadêmica brasileira, os trabalhos de Morais *et al.* (2024) e de Scoralick, Brandão e Belloze (2024) dependem de rótulos (multas aplicadas ou casos de conluio identificados), que não existem de forma sistemática para licitações municipais. Os de Ramalho, Almeida e Fraga (2020) e de Maia *et al.* (2020) dispensam rótulos, mas atribuem o risco a empresas, com critérios e pesos definidos por regra, sem medir o peso de cada sinal em cada processo sinalizado. Nenhum dos trabalhos consultados examina municípios piauienses.

No Estado do Piauí, essa lacuna é especialmente relevante. Em 2022, o Produto Interno Bruto (PIB) *per capita* do Estado equivalia a cerca de 45% da média nacional (Instituto Brasileiro de Geografia e Estatística, 2024). Além disso, em 211 dos seus 224 municípios (94,2%), mais de 80% da receita corrente provém de repasses da União e do Estado (Piauí. Tribunal de Contas do Estado, 2025). Nesse contexto, em que a maior parte dos recursos aplicados nas compras municipais têm origem externa e a capacidade local de fiscalização pode ser limitada, uma triagem automatizada e explicável das contratações pode ter utilidade prática.

Este trabalho investiga a seguinte pergunta: é possível identificar padrões atípicos em contratações públicas do Estado do Piauí, sem dispor de casos rotulados, de forma que cada sinalização seja acompanhada de uma explicação que um auditor possa examinar? Para responder a ela, foram definidos quatro padrões suspeitos, adaptados de sinais de alerta da literatura de controle. Os padrões são valor atípico para o tipo de objeto (P1), indício indireto de sobrepreço; concentração de vitórias de um fornecedor (P2); contratação de empresa recém-constituída (P3); e hiperfoco do fornecedor em um único tipo de objeto (P4). Os padrões foram operacionalizados como variáveis de um modelo Isolation Forest, algoritmo de detecção de anomalias que dispensa rótulos (Liu; Ting; Zhou, 2008). As decisões do modelo são decompostas pelo método SHapley Additive exPlanations (SHAP), que atribui a cada variável a sua contribuição para o resultado (Lundberg; Lee, 2017). Neste artigo, "Estado do Piauí" designa o território: a base reúne órgãos municipais, estaduais e federais ali sediados. O termo

"contratações" abrange as licitações e as contratações diretas (dispensa e inexigibilidade) divulgadas no PNCP, e "registro" designa o vínculo entre um fornecedor e uma contratação.

O trabalho testa três hipóteses. A primeira (H1) é que as variáveis que representam os padrões P1 a P4 respondem pela maior parte da explicação dos *scores* de anomalia, mesmo quando variáveis neutras são incluídas no modelo. Em outras palavras, o modelo sinalizaria registros pelos padrões que a pesquisa pretende investigar, mais do que por características neutras dos dados. A segunda (H2) é que contratações com empresas de menos de 12 meses de existência são sinalizadas em proporção superior à média do conjunto. Como a variável de idade foi construída para produzir esse efeito, H2 funciona como checagem de coerência, não como teste independente. A terceira (H3) é que os registros sinalizados concentram, em proporção superior à do restante dos dados, fornecedores inscritos no Cadastro de Empresas Inidôneas e Suspensas (CEIS) ou no Cadastro Nacional de Empresas Punidas (CNEP). Se confirmada, H3 funcionaria como uma validação externa parcial.

O objetivo geral é avaliar se é possível identificar padrões atípicos em dados públicos de contratações do Estado do Piauí, sem rótulos, com um modelo Isolation Forest cujas sinalizações são interpretadas pelo método SHAP. O mesmo objetivo inclui avaliar em que medida os resultados resistem às escolhas de construção do modelo. Os objetivos específicos são: (i) coletar dados de contratações do PNCP vinculadas ao Estado; (ii) tratar os dados e construir as variáveis que representam os padrões P1 a P4; (iii) aplicar um algoritmo de detecção de anomalias aos dados; (iv) interpretar os resultados com SHAP, nos níveis global e individual; (v) analisar os registros sinalizados à luz de indicadores conhecidos de irregularidade, incluindo o cruzamento com cadastros de sanções; e (vi) testar a sensibilidade dos resultados às escolhas de construção do modelo. Como os resultados são indicadores estatísticos de atipicidade, que não equivalem a acusações (Seção 2.2), os estudos de caso não identificam empresas nem órgãos.

Como contribuição, o trabalho oferece um *pipeline* de coleta, tratamento e modelagem construído apenas sobre dados públicos e organizado em etapas com testes automatizados. Oferece também um protocolo para avaliar o modelo sem rótulos, que combina verificações de coerência, estabilidade e sensibilidade e o cruzamento com cadastros de sanções.

2 REFERENCIAL TEÓRICO

2.1 Contratações públicas, PNCP e padrões de irregularidade

A transição para a Lei nº 14.133/2021 foi escalonada. A Lei nº 8.666/1993 foi definitivamente revogada em 30 de dezembro de 2023 (Brasil, 2023). Já os municípios com até 20 mil habitantes têm seis anos, contados da publicação da lei, para cumprir exigências como a licitação eletrônica e a divulgação em sítio oficial (Brasil, 2021). Esse cronograma explica por que os dados de anos anteriores a 2024 são escassos.

Os órgãos de controle trabalham com sinais de alerta, que apontam onde olhar sem afirmar que houve irregularidade. O Referencial de Combate a Fraude e Corrupção do TCU traz exemplos recorrentes em compras públicas: propostas todas bem acima do orçamento, no conluio entre licitantes; especificação que identifica o produto do vencedor, no conluio entre comprador e licitante; e, no caso de fornecedor único, o favorecimento de uma empresa e a empresa sem histórico no ramo (Brasil. Tribunal de Contas da União, 2018). Os quatro padrões investigados neste trabalho adaptam esses sinais ao que os dados do PNCP permitem medir. O sobrepreço é definido em lei como preço orçado ou contratado em valor expressivamente superior aos preços referenciais de mercado (Brasil, 2021). Como o PNCP não traz preços de referência, o valor atípico (P1) mede o quanto o valor homologado total se afasta do típico de contratos de objeto semelhante: aproxima o sinal de propostas acima do orçamento, sem medi-lo diretamente.

A concentração de vitórias (P2) e o hiperfoco do fornecedor em um único tipo de objeto (P4) aproximam-se do favorecimento de uma empresa e do direcionamento da especificação. Em P4, porém, o indício é ambíguo: a especialização total também é característica de um fornecedor típico de um ramo. A contratação de empresa recém-constituída (P3) aproxima-se da empresa sem histórico no ramo. Nenhum desses padrões é, por si, uma irregularidade. Na literatura internacional, Fazekas e Kocsis (2020) usam dados de 2,8 milhões de contratos de 28 países europeus para construir indicadores de risco de corrupção no nível do contrato, entre eles um índice composto de sinais de alerta (*red flags*); este estudo usa variáveis semelhantes como entrada de um modelo de detecção de anomalias explicado caso a caso.

2.2 Anomalias e aprendizado sem rótulos

Um *outlier*, termo usado muitas vezes como sinônimo de anomalia (Chandola; Banerjee; Kumar, 2009), é, na definição clássica de Hawkins (1980, p. 1, tradução nossa), "uma observação que se desvia tanto das demais observações a ponto de despertar suspeitas de que foi gerada por um mecanismo diferente". Detectar anomalias é, portanto, encontrar padrões que não se conformam ao comportamento esperado (Chandola; Banerjee; Kumar, 2009). Fraude, por outro lado, é um conceito jurídico, que em geral pressupõe intenção e tipificação legal, e passar da anomalia detectada computacionalmente à constatação de irregularidade exige análise documental e, muitas vezes, processo administrativo ou judicial.

Essa distinção orienta todo o trabalho. Um contrato pode ser atípico por razões legítimas (objeto raro, obra de grande porte, situação emergencial) ou por simples erro de registro. Os modelos de detecção de anomalias servem para ordenar a atenção do auditor; não substituem o seu julgamento.

Na abordagem supervisionada, o modelo aprende a separar classes a partir de exemplos rotulados. A limitação dessa abordagem é a dependência de rótulos confiáveis: para licitações municipais, não existe uma base pública de processos confirmados como irregulares, e os cadastros de sanções (CEIS e CNEP) cobrem apenas parte das irregularidades, com defasagem temporal. Por isso, este trabalho adota a abordagem não supervisionada, em que o modelo não recebe rótulos e procura observações que se afastam da estrutura geral dos dados (Chandola; Banerjee; Kumar, 2009).

2.3 Isolation Forest e explicabilidade com SHAP

O Isolation Forest, proposto por Liu, Ting e Zhou (2008), parte de uma premissa simples: anomalias são poucas e diferentes e, por isso, mais fáceis de isolar. O algoritmo constrói um conjunto de árvores binárias; em cada nó, escolhe ao acaso uma variável e um ponto de corte entre o mínimo e o máximo dela. Um registro típico precisa de muitos cortes para ser separado dos demais, enquanto um registro atípico fica isolado em poucos. O *score* de anomalia é derivado da profundidade média em que cada registro é isolado: quanto menor a profundidade, mais atípico o registro (Liu; Ting; Zhou, 2012).

Em licitações brasileiras, Abreu, Pereira e Gomes Junior (2024) compararam o Isolation Forest ao Local Outlier Factor (LOF) e aos mapas auto-organizáveis, em dados de notas fiscais, e o consideraram o mais efetivo em localizar possíveis casos de fraude. O algoritmo é adequado

a este estudo porque não exige rótulos, tem custo computacional baixo (Liu; Ting; Zhou, 2008) e, por ser baseado em árvores, admite o cálculo exato dos valores SHAP (Lundberg *et al.*, 2020). Por seu funcionamento, ele isola os dois extremos de cada variável; se apenas um deles representa risco, a variável precisa ser construída de modo que o outro deixe de se destacar (Seção 3.3).

O Isolation Forest, porém, não explica as próprias decisões: produz um número (o *score*) sem informar quais características do registro levaram a ele. Por isso, exige técnicas de explicação *a posteriori*, mais necessárias quando as decisões afetam pessoas (Barredo Arrieta *et al.*, 2020), como a escolha de quais processos auditar.

O método SHAP, de Lundberg e Lee (2017), responde a essa necessidade com base nos valores de Shapley da teoria dos jogos cooperativos. Para cada registro, reparte entre as variáveis a diferença entre a saída do modelo e a saída média, e a soma das contribuições reproduz exatamente essa diferença. No nível global, a média dos valores absolutos indica quais variáveis mais influenciam o modelo (Lundberg *et al.*, 2020); no individual, os valores de um registro mostram quais características o tornaram atípico. Para modelos de árvores, o algoritmo TreeExplainer calcula esses valores de forma exata (Lundberg *et al.*, 2020). O SHAP explica por que o modelo atribuiu determinado *score* a um registro; ele não diz se o registro corresponde a uma irregularidade.

2.4 Trabalhos relacionados

Os estudos brasileiros mais próximos e o *software* ALICE diferem deste trabalho quanto à dependência de rótulos, à explicação de cada sinalização e ao âmbito dos dados. O ALICE dispensa rótulos, emite alertas por regras e mineração de textos e processa cerca de 500 licitações por dia (Borges *et al.*, 2026), mas foi concebido para contratações federais. Ramalho, Almeida e Fraga (2020) aplicaram regras de associação a cerca de 104 mil licitações municipais da Paraíba, de 2005 a 2016, e ranquearam as empresas suspeitas de conluio por um índice de suspeição; a explicação é dada por empresa ou grupo de empresas, não por processo. Maia *et al.* (2020), também sem rótulos, combinaram padrões frequentes, correlação temporal e múltiplos critérios e identificaram, em compras de órgãos federais em Minas Gerais, um conluio prolongado entre cinco fornecedores, com explicação por critério definido por regra.

Com rótulos, Morais *et al.* (2024) usaram um Random Forest com licitações de 2022 e multas como rótulo e alcançaram *F1-score* de 80%, e Scoralick, Brandão e Belloze (2024), com

licitações federais de 2013 a 2018 e casos de conluio como rótulo, obtiveram ganhos de 1% a 4% com variáveis estatísticas; nenhum dos dois explica cada caso, e o segundo traz só a importância global. Abreu, Pereira e Gomes Junior (2024) usaram dados de notas fiscais, sem explicação por caso. Nos trabalhos consultados, selecionados de forma exploratória e sem revisão sistemática, não foi encontrada a combinação de detecção não supervisionada por um modelo multivariado com explicação individual, aplicada a dados do PNCP e ao contexto piauiense. O presente estudo examina essa combinação.

3 METODOLOGIA

Esta pesquisa é aplicada quanto à natureza, quantitativa quanto à abordagem e exploratória quanto aos objetivos. Utiliza exclusivamente dados secundários públicos e delimita-se às contratações vinculadas ao Estado do Piauí registradas no PNCP. O procedimento seguiu cinco etapas: coleta (PNCP, BrasilAPI e cadastros CEIS e CNEP), pré-processamento e formação das categorias, construção das variáveis P1 a P4, modelagem com Isolation Forest e SHAP e avaliação sem rótulos. O código foi escrito em Python, com as bibliotecas pandas, *scikit-learn* e SHAP.

A parte teórica baseou-se em pesquisa bibliográfica exploratória, feita em março de 2026, sobre detecção de anomalias, corrupção e transparência em compras públicas, aprendizado de máquina em gastos municipais, explicabilidade com SHAP e avaliação sem rótulos. Não se aplicou um critério sistemático de exclusão, e as referências que têm DOI (*Digital Object Identifier*) foram conferidas no Crossref.

3.1 Fonte e coleta de dados

Os dados foram coletados em 4 e 5 de novembro de 2026 de duas APIs públicas do PNCP (Portal Nacional de Contratações Públicas, 2026). Primeiro, foram listadas as contratações publicadas por órgãos do Piauí de 2020 a 2025: 29.114, de 458 órgãos, das quais 20.187 municipais. Em seguida, foram obtidos os itens de cada contratação, com o valor estimado, e os fornecedores vencedores, com o valor homologado. Por fim, foram coletados os contratos dos 409 órgãos estaduais e municipais da listagem, mantidos os de unidades situadas no Piauí, num total de 10.676. Como a API de contratos não filtra por Estado e os contratos dos órgãos federais abrangem o país inteiro, estes ficaram de fora, e a esfera federal reúne apenas os fornecedores vencedores de contratações.

A consulta de itens foi feita sem paginação, e a API devolveu no máximo 10 itens por contratação. Das 28.406 contratações de editais e atas com itens, 6.459 (22,7%) têm exatamente 10, sinal de que a lista foi cortada. Nelas, o valor homologado considera só os 10 primeiros itens, e os fornecedores que venceram apenas itens posteriores não aparecem na base. Na base do modelo (Seção 3.2), 4.971 registros (19,3%) vêm dessas contratações, e a Seção 4.4 verifica o efeito de retirá-los. A recoleta paginada fica como trabalho futuro.

A cobertura tem limites: 222 dos 224 municípios do Estado têm ao menos uma contratação municipal publicada, mas o ano de 2020 não apresentou registros, e os anos anteriores a 2024 são escassos, reflexo da transição prevista na Lei nº 14.133/2021 (arts. 176 e 191) (Brasil, 2021; Brasil, 2023). A base aproxima, portanto, um censo das contratações publicadas no PNCP, não das realizadas.

Como o PNCP não informa a data de abertura das empresas, necessária para P3, os identificadores dos fornecedores, em sua maioria do Cadastro Nacional da Pessoa Jurídica (CNPJ), foram consultados na BrasilAPI (BrasilAPI, 2026), serviço aberto que replica os dados cadastrais da Receita Federal. Os identificadores de 11 dígitos têm o formato do Cadastro de Pessoas Físicas (CPF) e correspondem, em geral, a pessoas físicas. A validação externa usou os arquivos do CEIS e do CNEP do Portal da Transparência (Brasil. Controladoria-Geral da União, 2026), obtidos em 15 de julho de 2026.

3.2 Pré-processamento

A unidade de análise é o registro, vínculo entre um fornecedor e uma contratação. Os valores homologados e estimados dos itens que um fornecedor venceu no mesmo processo foram somados, para que a razão entre eles compare grandezas equivalentes. A base completa reúne 31.687 registros, e a base do modelo, sem as pessoas físicas e as possíveis duplicatas, 25.757.

Saíram 2.920 registros de pessoas físicas: para elas, P3 não se aplica, e o cruzamento com o CEIS e o CNEP só aceita CNPJ (Seção 3.5). Também saíram as possíveis duplicatas: registros que coincidem em órgão, fornecedor, ano e valor homologado com registros de outro tipo de documento e podem corresponder ao mesmo processo publicado como edital, ata e contrato. De cada grupo ficou um registro, com preferência pelo contrato. As possíveis duplicatas somam 3.151 registros, mas 141 deles já haviam saído como pessoas físicas, de modo que esse critério remove outros 3.010.

A base do modelo tem 415 órgãos e 216 municípios, e 90,5% dos registros são de contratações de 2024 e 2025. Dos 25.757 registros, 12.800 são de órgãos municipais (49,7%), 7.856 de estaduais (30,5%) e 4.941 de federais (19,2%), e 160 estão classificados como "Não se aplica". Por tipo de documento, 58,0% são editais, 38,0% contratos e 4,0% atas de registro de preços. Os valores homologados são muito assimétricos: a mediana é de R$ 58.695,00, e a média, de R$ 1.877.571,36. Como as médias e dispersões por categoria de objeto dependem de bases amplas, optou-se por modelar todas as esferas juntas e analisar a esfera municipal nos resultados.

Para comparar valores entre contratos semelhantes, a descrição do objeto de cada contratação foi convertida em uma representação *term frequency–inverse document frequency* (TF-IDF) (Salton; Buckley, 1988), que pondera cada palavra pela frequência na descrição e pela raridade no conjunto. Em seguida, o algoritmo K-Means (Lloyd, 1982), que agrupa as descrições de representações mais próximas, formou 25 categorias, número que é uma escolha operacional, sem otimização formal. A categorização é aproximada: a coesão média das categorias, medida pela similaridade de cosseno entre cada descrição e o centro da sua categoria, é de 0,39, contra 0,19 de uma partição aleatória com os mesmos tamanhos. Três categorias, porém, têm coesão próxima da aleatória ou inferior e reúnem 46,4% dos registros da base do modelo (Seção 4.4).

3.3 Variáveis do modelo

O Quadro 1 apresenta as seis variáveis do modelo. Cinco representam um padrão suspeito adaptado de sinais de alerta da literatura de controle (Brasil. Tribunal de Contas da União, 2018), e uma é auxiliar, a razão entre o valor homologado e o estimado.

Quadro 1 – Variáveis do modelo e padrões suspeitos que representam

Variável	Definição
P1 global: valor atípico	z-score de log(1+valor homologado total) em relação à média e ao desvio-padrão da categoria de objeto; valores negativos passam a 0
P1 municipal: valor atípico	O mesmo z-score, calculado no grupo formado pelo município do órgão e pela categoria de objeto e multiplicado por n/(n+10), em que n é o número de registros do grupo; valores negativos passam a 0; indefinido em grupos de um registro
Razão homologado/estimado (auxiliar)	Logaritmo do valor homologado dividido pelo valor estimado pelo órgão para os mesmos itens; valores negativos passam a 0; indefinida quando o estimado é zero, negativo ou ausente
P2: concentração de vitórias	Proporção dos registros do órgão no conjunto em que o fornecedor é o vencedor; indefinida em órgãos com menos de 20 registros
P3: idade da empresa	Meses entre a abertura do CNPJ e a data de referência do registro, limitados a 60; a data de referência é a assinatura, nos contratos, e a data do resultado do item, nos editais e nas atas

Variável	Definição
P4: hiperfoco	Índice de Herfindahl-Hirschman (HHI) da distribuição dos registros do fornecedor entre as categorias de objeto; valores abaixo da mediana (0,50) passam à mediana; indefinido para fornecedores com um único registro

Fonte: Elaboração própria (2026).

Nota: Valores indefinidos recebem a mediana da variável, e todas as variáveis são padronizadas antes do treinamento.

Os dois *z-scores* medem o quanto o valor total do registro se afasta do típico de contratos de objeto semelhante. P1 é, portanto, um indício indireto de sobrepreço e não o mede, pois o valor total mistura preço unitário, quantidade e escopo. A razão homologado/estimado, mais próxima da definição legal, foi mantida como auxiliar porque um homologado abaixo do estimado resulta também de disputa de preços legítima. Um quinto padrão considerado inicialmente, a baixa concorrência, foi descartado porque o PNCP não informa o número de licitantes.

As variáveis seguem a mesma regra: cada uma destaca só o polo de risco, e grupos pequenos não devem gerar valores extremos. Como o Isolation Forest isola com a mesma facilidade os valores muito altos e os muito baixos (Seção 2.3), o lado sem risco foi achatado: os valores abaixo de um piso recebem o próprio piso, e só a cauda de risco continua extrema. Quanto aos grupos pequenos, uma média calculada sobre poucos registros é instável, e um valor extremo diz mais sobre o tamanho do grupo do que sobre o registro.

Em P1, o *z-score* usa o logaritmo do valor, porque os valores homologados são muito assimétricos (Seção 3.2), e o corte em o mantém só os valores acima da média da categoria, pois um contrato barato não é sinal de risco. No P1 municipal, o grupo não distingue a esfera do órgão, e os tamanhos dos grupos diferem muito: a mediana é de 819 registros por grupo nos órgãos estaduais e de 21 nos municipais do interior. O fator n/(n+10), chamado aqui de encolhimento, aproxima de zero o *z-score* dos grupos pequenos e quase não muda o dos grandes. A constante 10 é uma escolha operacional, sem validação empírica, e a Seção 4.4 mede só o efeito de retirar o encolhimento.

Em P2, num órgão com poucos registros, uma única vitória já produz uma proporção alta, sem que haja concentração de fato. Por isso, P2 ficou indefinida nos órgãos com menos de 20 registros, escolha operacional que atinge 197 órgãos, com 1.390 registros.

Em P4, o HHI, medida tradicional de concentração de mercado, é a soma dos quadrados das participações de cada parte. Com um único registro, o índice valeria 1 por definição, sem

haver especialização, e por isso ficou indefinido; com dois registros, porém, o HHI só pode valer 0,50 ou 1. Os valores baixos, de fornecedores diversificados, foram levados à mediana, de 0,50.

Na razão homologado/estimado, o logaritmo reduz o efeito dos valores extremos, e o corte em 0 mantém só o polo de risco: dos 25.136 registros com razão definida, 25.003 ficam em 0, e só 133 têm homologado acima do estimado. Em P3, a idade foi limitada a 60 meses pela mesma regra, pois sem o teto empresas com décadas de existência se destacariam tanto quanto as recém-abertas. O limiar de 12 meses, usado nos resultados, é uma definição operacional de empresa recém-constituída, sem validação empírica.

No P1 global, 12.440 registros ficam no piso; no P1 municipal, 12.751 ficam no piso e 399 são indefinidos; em P3, a idade é indefinida em 14 registros. Os valores indefinidos foram preenchidos com a mediana de cada variável, o que concentra muitos registros em um mesmo valor e pode alterar as divisões feitas pelo Isolation Forest; o efeito fica registrado como limitação.

P2 e P4 foram calculadas com os registros de todo o período, assim como as médias e os desvios de P1. Um registro de 2023 usa, portanto, vitórias de 2025, informação que um auditor não teria na data da contratação. Os valores também não foram corrigidos pela inflação, e as duas simplificações ficam como limitações do estudo.

3.4 Modelo e explicabilidade

O Isolation Forest foi configurado com 300 árvores, subamostras de 256 registros por árvore, valor recomendado pelos autores do algoritmo (Liu; Ting; Zhou, 2008), e semente aleatória fixa. O parâmetro de contaminação foi fixado em 5%: os 5% de registros com *scores* mais atípicos são classificados como sinalizados. Esse valor é uma escolha operacional de capacidade de análise, não uma estimativa da prevalência de irregularidades. Como o aprendizado é não supervisionado, o modelo foi ajustado sobre toda a base do modelo, sem divisão entre treino e teste.

Os valores SHAP foram calculados com o TreeExplainer (Lundberg *et al.*, 2020). No Isolation Forest, eles decompõem a profundidade média de isolamento de cada registro: valores negativos encurtam o caminho até o isolamento e empurram o registro em direção à anomalia.

A especificação descrita até aqui resulta de várias revisões. As primeiras versões usavam coletas menores e outros conjuntos de variáveis; numa delas, o conjunto de variáveis foi escolhido por superar uma meta de razão SHAP de 0,60, medida que tende a ser alta quando

quase todas as variáveis são de padrão (Seção 3.5). A versão imediatamente anterior, chamada neste artigo de especificação anterior, usava a base completa e as variáveis sem as transformações da Seção 3.3. Os seus resultados mostraram artefatos de construção: o *z-score* sobre o valor bruto era dominado por extremos e destacava também contratos baratos; o *z-score* municipal perdia sentido em grupos pequenos; P2 era alta por construção nos órgãos pequenos; P4 e a razão destacavam os dois extremos; e a base incluía as possíveis duplicatas e as pessoas físicas. A especificação atual foi definida para corrigir esses artefatos, mas as correções foram escolhidas depois de conhecidos os resultados da anterior, o que deixa margem de decisão ao pesquisador. Por isso, a especificação anterior é reportada inteira como sensibilidade, e o efeito de cada correção é medido em separado (Seção 4.4).

3.5 Avaliação sem rótulos

Na ausência de rótulos, métricas como precisão, revocação e *F1-score* não podem ser calculadas. Obter rótulos confiáveis para treinar ou validar detectores de anomalias costuma ser caro ou inviável (Chandola; Banerjee; Kumar, 2009). A avaliação reúne, por isso, evidências de naturezas diferentes, cada uma com o critério de leitura definido a seguir.

Hipótese H1. A medida prevista para H1 era a razão SHAP, a parcela da importância total (média do valor SHAP absoluto) atribuída às variáveis de padrão, isto é, a todas exceto a auxiliar. Como a auxiliar quase não varia depois do corte (Seção 3.3), a razão SHAP fica perto de 1 por construção e é apenas reportada.

O teste que pode refutar H1 acrescenta ao modelo cinco variáveis neutras. São elas o logaritmo do valor estimado, a natureza do objeto (serviço ou material), o regime legal e o tipo de documento, este codificado por duas variáveis indicadoras (contrato e edital). H1 é considerada sustentada se, com elas, as variáveis de padrão responderem por mais de 50% da importância. Elas também precisam superar o esperado com importância igual entre as onze variáveis (45,5%). O critério foi formulado depois de dois resultados da especificação anterior, um em que H1 não teria sido sustentada e outro em que seria, e foi aplicado sem alteração à especificação atual, cujo resultado ainda não era conhecido. Como o limiar de 50% foi escolhido com aqueles resultados à vista, a sustentação de H1 deve ser lida com essa ressalva.

O logaritmo do valor estimado não é de fato neutro: acompanha de perto o valor homologado (correlação de Spearman de 0,923) e carrega parte da informação de P1, o que pesa contra H1. Por isso, o teste também é feito sem essa variável (esperado de 50,0%) e sem P4

(esperado de 40,0%), para verificar se a sustentação depende do hiperfoco. Já as variáveis binárias com categoria rara podem comandar a sinalização de alguns registros, mas o efeito se dilui na importância média, o que favorece H1; não se sabe, assim, se a construção do teste, no conjunto, favorece ou prejudica a H1.

Hipótese H2. H2 é uma checagem de coerência: compara a taxa de sinalização das empresas com menos de 12 meses com a taxa geral e é sustentada se a primeira for superior. Como o resultado depende de como a idade foi construída, H2 vem acompanhada de uma sensibilidade ao teto de idade (36, 60 e 120 meses e sem teto).

Hipótese H3. H3 usa os cadastros de sanções como referência externa: um registro é marcado quando o CNPJ do fornecedor aparece no arquivo do CEIS ou do CNEP, qualquer que seja a data da sanção. H3 é sustentada se a proporção de marcados for maior entre os sinalizados do que entre os demais, com intervalo de confiança (IC) de 95% que exclua zero. Uma variante considera só as sanções vigentes na data de referência do registro. Como os sinalizados tendem a ser contratos maiores, uma análise complementar controla o valor por padronização direta: a base é dividida em quintis do valor homologado, e o número de sinalizados com fornecedor sancionado é comparado ao esperado se cada um tivesse a taxa dos demais registros da mesma faixa.

Comparação de conjuntos. Várias verificações comparam dois conjuntos de sinalizados pelo índice de Jaccard (Jaccard, 1912), a razão entre o tamanho da interseção e o da união: $J=|A \cap B|/|A \cup B|$. Como referência, dois conjuntos sorteados ao acaso, cada um com 5% dos registros, teriam índice de cerca de 0,026.

Estabilidade e linha de base. A estabilidade é medida pelo Jaccard entre o conjunto sinalizado com a semente original e os obtidos com outras dez sementes. A linha de base são os 5% de registros com maior P1 global; a interseção dela com o conjunto sinalizado, comparada com a esperada ao acaso, indica se o modelo faz algo além de apontar contratos caros para a categoria.

Dependência das escolhas de construção. Na ablação, o modelo é refeito sem uma das correções de cada vez: nas variáveis, o logaritmo e o corte de P1, o encolhimento do P1 municipal, o mínimo de 20 registros em P2, o corte de P4 e o logaritmo e o corte da razão; na base, a remoção das possíveis duplicatas e a das pessoas físicas. Outras verificações trocam a forma de P1 (versões bicaudais e com mediana e desvio absoluto mediano) e ajustam um modelo só com os registros municipais e outro sem os registros de contratações com 10 itens.

Injeção e incerteza. A sensibilidade por injeção altera as variáveis de um padrão, ou de dois padrões combinados, em 1.000 registros não sinalizados e mede quantos o modelo treinado passa a sinalizar; ela mede o que o modelo foi desenhado para ver e nada diz sobre irregularidades reais. A estabilidade e as sensibilidades são descritivas: não se fixou um limiar de estabilidade. Os IC de 95% vêm de *bootstrap* por fornecedor, com 2.000 repetições (Field; Welsh, 2007), porque os registros de um mesmo fornecedor não são independentes. O *bootstrap* trata o conjunto sinalizado como dado, de modo que os intervalos refletem só a amostragem de fornecedores, não a incerteza do próprio modelo.

Casos. Dois registros sinalizados, escolhidos por critério objetivo, são analisados individualmente com os valores SHAP. O Caso A é o registro mais atípico de todos, e o Caso B é o registro municipal sinalizado mais atípico cuja contribuição mais negativa vem da idade da empresa.

3.6 Aspectos éticos e transparência

Todos os dados são públicos e referem-se, na maior parte, a pessoas jurídicas e a órgãos no exercício de suas funções. Na base completa, 9,2% dos registros têm uma pessoa física como fornecedor; eles ficaram fora do modelo, e os dados dessas pessoas, publicados pelo próprio PNCP, não são divulgados no artigo. A decisão segue o art. 7º, § 3º, da Lei Geral de Proteção de Dados Pessoais (Brasil, 2018), que manda considerar a finalidade, a boa-fé e o interesse público que justificaram a disponibilização dos dados. Como uma sinalização estatística associada ao nome de uma empresa pode prejudicar a sua imagem, os estudos de caso também são apresentados sem identificação.

O *pipeline* foi organizado em etapas separadas de coleta, processamento, modelagem e verificações. Testes automatizados conferem as funções que constroem as variáveis, e um *script* confere os números do texto contra os arquivos de resultado. Com a semente fixa, duas execuções sobre os mesmos dados brutos produzem os mesmos resultados. O código é mantido em repositório versionado, disponível mediante solicitação aos autores. Como as APIs do PNCP e a BrasilAPI mudam com o tempo, repetir a análise depende da data de corte da coleta, informada na Seção 3.1, ou do arquivamento dos dados brutos.

O Elicit, ferramenta de pesquisa assistida por inteligência artificial, foi usado na busca bibliográfica, em março de 2026. O assistente Claude (Anthropic) foi usado na revisão do

código, na proposição e na execução de análises de robustez e na revisão do texto. Todo o conteúdo foi revisado pelos autores, que respondem pelo texto, pelos dados e pelas conclusões.

4 RESULTADOS E DISCUSSÃO

4.1 Visão geral

O modelo sinalizou 1.288 dos 25.757 registros da base (5,0%). O *score* médio dos sinalizados foi −0,622, contra −0,416 dos demais (*scores* menores indicam registros mais atípicos), uma separação apenas descritiva, pois decorre da definição do limiar. Os sinalizados ocupam a cauda esquerda da distribuição, sem fronteira natural entre os grupos: o limiar de 5% é uma escolha de capacidade de análise.

Os sinalizados distribuem-se entre órgãos estaduais (609), municipais (451) e federais (225), e 3 vêm dos 160 registros classificados como "Não se aplica". Os sinalizados são, em geral, contratos maiores: o valor homologado mediano é de R$ 688.732,00 entre eles e de R$ 55.095,00 entre os demais.

O conjunto sinalizado é concentrado em poucos órgãos e fornecedores. O órgão com mais sinalizações responde por 240 delas (18,6% do total), e os dez primeiros, por 50,4%. Um único fornecedor, que atua quase só nesse mesmo órgão estadual, tem 201 dos 1.288 sinalizados (15,6%), 196 deles com a concentração de vitórias (P2) como principal contribuição, isto é, a contribuição SHAP mais negativa. Assim, um só relacionamento entre fornecedor e órgão ocupa 200 posições da lista, já que apenas um dos sinalizados do fornecedor vem de outro órgão. O órgão e o fornecedor são os do Caso A (Seção 4.5).

4.2 Explicabilidade global e hipótese H1

A importância de cada variável foi medida pela média do valor SHAP absoluto. O hiperfoco (P4) é a variável mais importante (1,018), seguido pela idade da empresa (P3, 0,965), pelos dois *z-scores* de valor (P1 municipal, 0,630; P1 global, 0,586) e pela concentração de vitórias (P2, 0,539). A razão homologado/estimado quase não contribui (0,047).

O primeiro lugar de P4 decorre, em parte, do corte na mediana do HHI (0,50). Depois do corte, 12.315 registros ficam no piso, e os 5.043 sem HHI definido recebem o mesmo valor, de modo que a cauda acima da mediana se destaca com facilidade. O primeiro lugar também pode refletir os fornecedores com apenas dois registros, cujo HHI vale 0,50 ou 1 (Seção 3.3).

A razão homologado/estimado ficou quase inerte por construção: só 133 registros ficam acima do piso (Seção 3.3), e 39 deles foram sinalizados. Por isso, a razão SHAP chegou a 0,988, contra 0,833 esperados se as seis variáveis tivessem a mesma importância, e, como previsto na Seção 3.5, não serve como evidência para H1.

O teste que pode refutar H1 acrescenta cinco variáveis neutras ao modelo (Seção 3.5). Na versão completa, as variáveis de padrão respondem por 66,5% da importância, contra 45,5% esperados com importância igual; sem o logaritmo do valor estimado, por 72,6% (esperado de 50,0%); e sem P4, por 58,3% (esperado de 40,0%). Nas três versões, a parcela supera 50% e o esperado. H1 é, portanto, sustentada pelo critério definido. Na versão completa, 269 sinalizados (20,9%) têm uma variável neutra como principal contribuição, 220 deles pelo regime legal, pois o Isolation Forest isola com facilidade os poucos registros da lei antiga; ainda assim, o regime legal responde por só 4,9% da importância média, abaixo dos 9,1% de uma divisão igual.

A sustentação de H1 vale com ressalvas, pois o critério foi fixado depois de conhecidos os resultados da especificação anterior (Seção 3.5). A importância média também soma contribuições nos dois sentidos: em P3, quase metade dela (48,9%) vem das empresas com 60 meses ou mais, que o teto torna típicas e que o modelo empurra para a normalidade. Além disso, H1 trata da importância das variáveis, não de quais registros são sinalizados: o modelo com neutras mantém 74,0% dos sinalizados do principal (Jaccard de 0,587), e as seções seguintes analisam o conjunto do modelo principal.

Valores altos de P1, P2 e P4 e valores baixos de P3, de empresas jovens, produzem as contribuições negativas, que empurram o registro para a anomalia, enquanto os cortes deixam os valores baixos de P1 e de P4 em contribuições pequenas e positivas. Entre os sinalizados, a variável de contribuição mais negativa é P2 em 29,3% dos casos, P4 em 26,5%, P1 global em 19,1%, P3 em 18,1%, P1 municipal em 4,7% e a razão em 2,3%.

4.3 Idade da empresa, esfera e hipótese H2

Contratações com empresas de menos de 12 meses foram sinalizadas em 16,6% dos casos (IC de 95%: 12,6% a 21,2%). A taxa cai com a idade: 9,0% entre 12 e 36 meses, 4,2% entre 36 e 60 meses e 3,5% com 60 meses ou mais. A taxa das empresas com menos de 12 meses é 4,68 vezes a das empresas com 60 meses ou mais (IC de 95%: 2,63 a 8,61).

H2 se verifica, mas como checagem de coerência, pois o teto de 60 meses foi construído para que apenas a juventude da empresa se destacasse. Com o modelo refeito com outros tetos,

a taxa das empresas com menos de 12 meses é de 23,5% com teto de 36 meses, 16,6% com 60 e 11,1% com 120. Sem teto, ela cai para 3,8%, abaixo da média, e a das empresas com cinco anos ou mais sobe a 5,9%. H2 decorre, portanto, da construção da variável. O resultado mostra que a juventude pesa nas sinalizações, mas não mostra, por si, que empresas jovens cometem mais irregularidades.

A esfera não é objeto de hipótese, mas foi examinada porque a lacuna de cobertura dos municípios motivou o trabalho. Os órgãos municipais têm a menor taxa de sinalização: 3,5% (IC de 95%: 2,9% a 4,2%), contra 7,8% nos estaduais (4,6% a 12,1%) e 4,6% nos federais (3,0% a 6,5%). A taxa municipal equivale a 0,55 vez a das demais esferas (IC de 95%: 0,36 a 0,86).

Em todas as variações do modelo principal (Tabela 1), a razão entre a taxa municipal e a das demais esferas fica abaixo de 1, e a municipal fica abaixo da estadual. O IC dessa razão, porém, passa a incluir 1 quando se desfaz o encolhimento do *z-score* municipal ou o mínimo de 20 registros em P2: a conclusão sobre a esfera municipal depende das correções que tratam de grupos pequenos. A taxa municipal cresce com o tamanho do grupo usado no *z-score* municipal, de 1,8% nos grupos de 1 ou 2 registros a 4,3% nos grupos com mais de 20, porque, com 2 registros, esse *z-score* não passa de 0,12.

Com o modelo ajustado apenas aos 12.800 registros municipais, 399 dos 451 sinalizados municipais continuam entre os 451 mais atípicos. Nesse teste, porém, as variáveis continuam calculadas na base inteira. O resultado mostra que o ajuste das árvores depende pouco das demais esferas, mas variáveis como P1 global ainda comparam cada registro com contratos de todas elas.

Sem o fornecedor com mais sinalizações, as taxas ficam em 3,5% na esfera municipal, 5,6% na estadual e 4,5% na federal: a estadual cai bastante, mas continua à frente. A composição por tipo de documento também não explica a diferença entre as esferas. Dentro dos editais, a taxa é de 3,9% na esfera municipal, 8,5% na estadual e 4,7% na federal; dentro dos contratos, de 2,6% na municipal e 7,5% na estadual (não há contratos federais na base).

Na especificação anterior, P2 subia por construção nos órgãos com poucos registros. O mínimo de 20 registros deslocou o artefato: a maior taxa por tamanho de órgão aparece logo acima do corte, com 8,1% nos órgãos com 20 a 49 registros, contra 4,3% nos de 50 a 199. O corte em 20 é uma escolha sem validação empírica, e essa dependência fica registrada como limitação.

4.4 Robustez e validação externa: hipótese H3

O Quadro 2 resume as evidências de avaliação, detalhadas nos parágrafos seguintes.

Quadro 2 – Síntese da avaliação do modelo

Evidência	Resultado	Leitura
Razão SHAP	0,988 (esperado 0,833)	Alta por construção; não serve como evidência para H1
Variáveis de padrão no modelo com neutras	66,5% (esperado 45,5%); sem P4, 58,3% (esperado 40,0%)	H1 sustentada pelo critério, com ressalvas
Jaccard entre sementes (10 execuções)	0,892 (0,857 a 0,915)	Conjunto estável; 1.109 registros em todas as execuções
Linha de base de valor (5% de maior P1 global)	Jaccard 0,235 (ao acaso, 0,026); 38,0% dos sinalizados nela	O valor alto é parte importante do conjunto, mas não o explica sozinho
Empresas com menos de 12 meses sinalizadas	16,6% (teto de 60 meses); 3,8% sem teto	H2 se verifica apenas com o teto de idade
Correções desfeitas uma a uma	Jaccard de 0,433 a 0,869; de 60,4% a 93,0% dos sinalizados mantidos	O conjunto depende das escolhas, sobretudo da construção de P1
Especificação anterior	Jaccard 0,271; 43,4% dos sinalizados mantidos	As duas especificações apontam, em boa parte, registros diferentes
Injeção de padrões em 1.000 registros	100,0% (combinações); 13,7% (P4); 15,7% (razão de 3); 16,4% a 27,4% (P3)	P4, a razão e a idade, isolados, são detectados em menos de 30% dos casos
Registros com fornecedor no CEIS ou no CNEP	2,48% (sinalizados) e 5,43% (demais)	Menos sancionados entre os sinalizados (H3 não sustentada)
Sanções vigentes na data do registro	0,23% (sinalizados) e 0,78% (demais)	Mesmo sentido
Sancionados entre os sinalizados, com o valor controlado (quintis)	32 observados e 51,7 esperados	Diferença cai para cerca de metade, e o IC inclui zero

Fonte: Elaboração própria (2026), com dados do PNCP, do CEIS e do CNEP.

Nota: Na linha das sementes, entre parênteses, os valores mínimo e máximo do índice de Jaccard nas dez execuções.

Trocar a semente aleatória quase não muda o conjunto. Com dez sementes, o índice de Jaccard médio é de 0,892 (mínimo de 0,857 e máximo de 0,915), e 1.109 dos 1.288 sinalizados aparecem em todas as execuções. A estabilidade mostra que o resultado se repete; não mostra que ele é válido.

A comparação com a linha de base de valor, os 5% de registros com maior P1 global, mostra que o valor pesa, mas não explica tudo. Os dois conjuntos têm 490 registros em comum, contra 64,4 esperados ao acaso, e índice de Jaccard de 0,235. Assim, 38,0% dos sinalizados estão entre os 5% de maior valor em relação à categoria.

A Tabela 1 mede a dependência das escolhas de construção. No primeiro bloco, cada linha desfaz uma correção das variáveis; no segundo, as linhas devolvem à base os registros

removidos ou retiram os de contratações com itens truncados; a última linha é a especificação anterior inteira.

Tabela 1 – Sensibilidade do conjunto sinalizado às escolhas de construção

Especificação	Reg.	Sinal.	Jaccard	Mant. (%)	Mun.	Est.	Fed.	Mun./demais (IC)
Modelo principal	25.757	1.288	1,000	100,0	3,5	7,8	4,6	0,55 (0,36 a 0,86)
P1 sem logaritmo nem corte	25.757	1.285	0,433	60,4	2,9	10,2	2,3	0,40 (0,19 a 0,98)
Z-score municipal sem encolhimento	25.757	1.288	0,869	93,0	4,1	7,2	4,0	0,68 (0,45 a 1,12)
P2 sem o mínimo de 20 registros	25.757	1.288	0,730	84,4	4,3	6,7	4,3	0,75 (0,54 a 1,09)
P4 sem corte	25.757	1.288	0,710	83,1	3,4	8,4	3,9	0,52 (0,35 a 0,78)
Razão sem logaritmo nem corte	25.757	1.288	0,696	82,1	3,0	6,8	7,3	0,44 (0,30 a 0,66)
Com as possíveis duplicatas	28.767	1.439	0,857	92,8	3,7	7,2	4,5	0,59 (0,40 a 0,91)
Com as pessoas físicas	28.536	1.427	0,827	89,7	3,6	7,4	4,9	0,59 (0,39 a 0,89)
Sem os registros de contratações com 10 itens	20.786	1.039	0,752	81,0	3,3	7,2	5,2	0,51 (0,33 a 0,82)
Especificação anterior	31.687	1.585	0,271	43,4	5,7	3,7	5,5	1,30 (1,09 a 1,57)

Fonte: Elaboração própria (2026), com dados do PNCP.

Nota: Reg.: registros da base. Sinal.: registros sinalizados. Mant.: parcela dos sinalizados do modelo principal que a especificação também sinaliza. Jaccard e Mant. são calculados sobre os registros presentes nas duas bases. Mun., Est. e Fed.: taxa de sinalização (%) por esfera. Mun./demais: razão entre a taxa municipal e a das demais esferas, com IC de 95%. A linha com as pessoas físicas devolve 2.779 dos 2.920 registros, pois 141 deles também são possíveis duplicatas e continuam fora.

O conjunto sinalizado depende das escolhas de construção, sobretudo da construção de P1: sem o logaritmo e o corte, só 60,4% dos sinalizados se mantêm (Jaccard de 0,433). O encolhimento do *z-score* municipal é a correção que menos muda o conjunto (93,0% mantidos, com Jaccard de 0,869). Entre a especificação anterior e a atual, o índice cai a 0,271, e só 43,4% dos sinalizados atuais também eram sinalizados na anterior: 559 registros, contra 66,7 esperados ao acaso. As duas especificações concordam muito mais que o acaso, mas apontam, em boa parte, registros diferentes.

Outras formas de calcular P1 confirmam o papel do corte em 0. Com mediana e desvio absoluto mediano do logaritmo, só no polo alto, 84,6% dos sinalizados se mantêm (Jaccard de 0,734). As versões que consideram os dois polos mantêm de 51,4% a 60,9% dos sinalizados (Jaccard de 0,346 a 0,438) e voltam a sinalizar contratos baratos: com o *z-score* do logaritmo nos dois polos, o valor mediano dos sinalizados cai para R$ 4.611,73, e 52,9% deles ficam abaixo da mediana da categoria.

Na injeção, as três combinações de dois padrões testadas (P2 de 60% com P3 de 6 meses, P1 de *z-score* 2 com P3 de 6 meses e P2 de 60% com P4 igual a 1) são detectadas em 100,0% dos casos. Isolados, o hiperfoco máximo (HHI igual a 1) e a razão de 3 são detectados em 13,7% e 15,7% dos casos, respectivamente, e a idade, de 16,4% (empresa com 12 meses) a 27,4% (1 mês).

Os 4.971 registros de contratações com exatamente 10 itens (Seção 3.1), de valor tendencialmente subestimado, são sinalizados em 2,3% dos casos, contra 5,6% nos demais; sem eles, 81,0% dos sinalizados se mantêm, e a ordem das esferas não muda. Devolver à base as possíveis duplicatas ou as pessoas físicas também muda pouco o conjunto (92,8% e 89,7% mantidos).

As três categorias de objeto de baixa coesão, que reúnem 46,4% dos registros, são um pouco menos sinalizadas (4,2% contra 5,9%). Ainda assim, a cauda alta de uma categoria heterogênea pode reunir os objetos caros por natureza e esconder o contrato caro para o seu tipo, o que fica como limitação.

A hipótese H3 não foi sustentada, e a comparação direta deu o inverso do esperado. Fornecedores presentes no CEIS ou no CNEP aparecem em 2,48% dos registros sinalizados e em 5,43% dos demais, diferença de −2,94 pontos percentuais (IC de 95%: −4,54 a −1,03) que mantém o sentido em todas as especificações listadas na Tabela 1. Só com as sanções vigentes na data do registro, a taxa é de 0,23% entre os sinalizados e de 0,78% entre os demais, num recorte pequeno (3 sinalizados com sanção vigente). Entre os 100 registros mais atípicos, a proporção de sancionados é de 3,0%, abaixo dos 5,28% da base.

Essa comparação, porém, não controla o valor dos contratos. Os sinalizados são, em geral, contratos maiores (Seção 4.1): 58,2% deles estão no quintil mais alto do valor homologado. Entre os demais registros, a proporção de sancionados cai à medida que o valor sobe, de 9,0% no primeiro quintil para 2,5% no último. Nesse último quintil, os sinalizados têm proporção um pouco maior que a dos demais (3,5% contra 2,5%).

Com a padronização pelos quintis de valor (Seção 3.5), seriam esperados 51,7 sinalizados com fornecedor sancionado, contra 32 observados (razão de 0,62; IC de 95%: 0,29 a 1,34). Em taxa, a diferença passa de −2,94 para −1,53 ponto percentual (IC de 95%: −3,49 a 1,01). Com o valor controlado, portanto, a diferença cai para cerca de metade e deixa de ser distinguível de zero. Boa parte do resultado inverso parece vir, assim, da concentração dos sinalizados em contratos grandes, que têm menos fornecedores sancionados. Isso não basta para apoiar H3: a estimativa continua abaixo do esperado, e o intervalo admite tanto menos quanto mais

sancionados entre os sinalizados. A comparação continua observacional, pois os quintis controlam o valor de forma grosseira.

Há outras explicações para a diferença, nenhuma delas testada. A primeira é o tipo de sanção: dos 1.360 registros com fornecedor nos cadastros, só 195 tinham a sanção vigente na data de referência, e em 1.162 ela começou depois. No arquivo do CEIS, 61,0% das sanções a pessoas jurídicas são impedimentos de licitar e contratar, fundamentados sobretudo na inexecução de contratos ou no descumprimento de obrigações do certame. Se for assim, os cadastros registram, em grande parte, falhas de execução posteriores à contratação, que os padrões modelados não descrevem. A segunda é a regra de cruzamento pelo CNPJ completo, que não conta a sanção registrada para outro estabelecimento da mesma empresa. Por fim, as empresas jovens, que o modelo prioriza, não concentram sanções: a proporção de sancionados entre as empresas com menos de 12 meses é de 5,4%, perto da base.

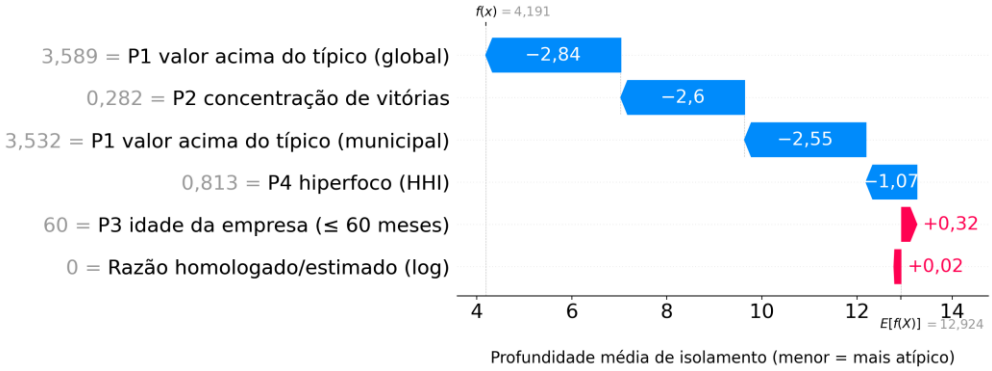
Os cadastros eram a única referência externa disponível, e ela não apoia a priorização feita pelo modelo. Com estes dados, não é possível saber se a falha está no modelo ou no uso dos cadastros como referência; separar as duas coisas exigiria outra fonte, como achados de auditoria dos tribunais de contas.

4.5 Estudos de caso

Os dois casos foram escolhidos pelo critério descrito na Seção 3.5. Na Figura 1, o modelo parte da profundidade média de isolamento do conjunto, E[f(X)]=12,92, e cada barra mostra quanto uma variável aproximou ou afastou o registro do isolamento: uma profundidade menor corresponde a um registro mais atípico.

Caso A: um fornecedor que domina o órgão. Trata-se de um contrato estadual de 2025 para fornecimento de alimentação, no valor de cerca de R$ 42 mil, igual ao estimado pelo órgão (razão de 1,00). A empresa existe há 197,4 meses, e por isso a idade empurra o registro para a normalidade (+0,32). A Figura 1 mostra que a queda da profundidade vem sobretudo de três contribuições de tamanho parecido: P1 global (−2,84), P2 (−2,60) e P1 municipal (−2,55), com o hiperfoco em seguida (−1,07).

Figura 1 – Decomposição SHAP do Caso A



Fonte: Elaboração própria (2026), com dados do PNCP.

Nota: O eixo horizontal é a profundidade média de isolamento nas árvores; quanto menor, mais atípico o registro. À esquerda de cada variável, o seu valor no registro, antes da padronização.

A explicação revela ao auditor o que está por trás do caso. O fornecedor, o mesmo da Seção 4.1, venceu 28,2% dos 1.875 registros do órgão. Os dois *z-scores* de valor são quase iguais (3,59 e 3,53) porque medem a mesma coisa. A categoria do caso é formada só pelos 617 registros desse órgão, que coincidem com o grupo de município e categoria, e na prática o valor é contado duas vezes. Um contrato de cerca de R$ 42 mil é atípico apenas diante de uma categoria de fornecimentos pequenos, com mediana de cerca de R$ 2,4 mil. O caso ilustra a concentração do conjunto sinalizado e não constitui um indício claro de sobrepreço. A primeira providência do auditor seria examinar a relação entre o fornecedor e o órgão, antes de discutir o preço.

Caso B: empresa recém-aberta em obra de grande porte. Trata-se do resultado de um edital municipal de 2024 para a construção de uma escola, homologado em cerca de R$ 9,1 milhões, abaixo do valor estimado pelo órgão, na posição 49 do *ranking* geral. Na data do resultado, a empresa tinha 6,5 meses. A decomposição deste caso, dada aqui só em texto, tem o mesmo formato da Figura 1: as maiores contribuições vêm da idade (−2,17) e do hiperfoco (−2,09), seguidas dos dois *z-scores* de valor (P1 municipal, −1,86; P1 global, −1,73). A concentração de vitórias é baixa (2,0%) e empurra levemente para a normalidade (+0,16).

Um filtro de valor também teria apontado esse registro, pois seu P1 global, de 1,80, está acima do limiar de 1,52 da linha de base. A explicação acrescenta o motivo da posição alta no *ranking*: a empresa é muito jovem, e todas as suas vitórias estão em uma só categoria (HHI igual a 1). Ela mostra também a força de cada sinal: o hiperfoco máximo vem de um fornecedor com apenas 2 registros, um indício fraco de especialização. A combinação de empresa recém-aberta

com obra de grande porte aproxima-se de sinais de alerta da literatura de controle, como a empresa sem histórico no ramo (Brasil. Tribunal de Contas da União, 2018).

4.6 Discussão

Em relação aos trabalhos relacionados, o modelo não oferece medidas de desempenho comparáveis às de Morais *et al.* (2024), que dispunham de rótulos, mas pode ser aplicado onde os rótulos não existem, como nas contratações municipais. Scoralick, Brandão e Belloze (2024) usaram o SHAP para mostrar a contribuição global das variáveis; aqui, cada decisão vem acompanhada de uma explicação individual, que não substitui a validação que os rótulos permitiriam. Em Ramalho, Almeida e Fraga (2020) e Maia *et al.* (2020), que também dispensaram rótulos, a explicação se resume a pontuações por critério, definidas por regra e atribuídas a empresas. A escolha do Isolation Forest segue Abreu, Pereira e Gomes Junior (2024), e este trabalho acrescenta uma ressalva: com o mesmo algoritmo, o conjunto sinalizado muda bastante conforme a construção das variáveis. O *software* ALICE (Borges *et al.*, 2026) emite alertas por regra, enquanto o modelo reúne os padrões em um único *score* e mostra a contribuição de cada um; não se testou se as duas abordagens se complementam.

Para um órgão de controle, os resultados sugerem usar o modelo como fila de priorização: começar pelo topo do *ranking* e descer até onde a equipe conseguir, dando preferência aos 1.109 registros que aparecem com todas as sementes. Agrupar a fila por fornecedor e por órgão evita que um único relacionamento ocupe boa parte dela, como o do Caso A, com 200 posições. A primeira etapa, porém, deve ser a conferência dos próprios dados, pois parte do que aparece no topo pode ser erro de registro. Nada disso foi testado com auditores, e a utilidade prática da explicação continua a ser uma hipótese.

O resultado de H3 também tem uma implicação para trabalhos que usam sanções como rótulo. Morais *et al.* (2024) usaram as penalidades (multas) aplicadas pela CGU em 2022, sem informar se elas vêm do CEIS e do CNEP. Se as sanções decorrem sobretudo da inexecução de contratos, como parece ocorrer no CEIS, um classificador treinado com elas aprende o risco de execução, não o da contratação. Mesmo onde há rótulos, é preciso verificar se eles medem o risco que o modelo pretende detectar.

A principal lição do trabalho é que uma triagem sem rótulos depende muito das escolhas de construção. A especificação anterior e a atual partem da mesma coleta, do mesmo algoritmo e dos mesmos padrões, mas apontam, em boa parte, registros diferentes (Tabela 1), e a conclusão

sobre a esfera municipal se inverte entre elas, embora nenhuma correção desfeita sozinha produza essa inversão. A construção de P1, sozinha, muda boa parte do conjunto. Essas escolhas precisam ser reportadas junto com o resultado, acompanhadas das verificações que mostram o quanto ele muda.

Os resultados também revelaram problemas de qualidade nos dados públicos, como 621 registros com valor estimado zero, negativo ou ausente e anos inválidos, como 20024 e 20225. Nesta fase do PNCP, parte do que uma triagem automatizada encontra são falhas de registro.

5 CONSIDERAÇÕES FINAIS

Este trabalho partiu de uma lacuna: as ferramentas de auditoria algorítmica de licitações usadas no Brasil foram concebidas para contratações federais, e poucos municípios recebem esse tipo de triagem. Quanto à primeira parte do objetivo geral, a resposta é afirmativa: sem nenhum rótulo, o Isolation Forest sinalizou registros sobretudo pelas variáveis de padrão e detectou todas as combinações de dois padrões injetados, e o SHAP indicou, para cada sinalizado, quais variáveis levaram o modelo a pontuá-lo. Isso mostra que o modelo encontra os padrões para os quais foi desenhado, mas não que eles correspondam a irregularidades. Quanto à segunda parte, o resultado resiste só em parte às escolhas de construção: é estável entre sementes, mas muda bastante com a forma de calcular o valor atípico e, em menor grau, com as demais correções. Além disso, a única referência externa disponível não o confirma.

Os objetivos específicos foram cumpridos: os dados foram coletados e tratados, os padrões foram convertidos em variáveis, e o modelo foi ajustado, explicado e avaliado com estudos de caso, cruzamento com cadastros de sanções e testes de sensibilidade. A hipótese H1 foi sustentada pelo critério definido: mesmo com variáveis neutras no modelo, as variáveis de padrão respondem pela maior parte da importância, também sem P4 e sem o logaritmo do valor estimado. A ressalva é que o critério foi fixado depois de conhecidos os resultados da especificação anterior. Já a hipótese H2 vale apenas como checagem de coerência, pois depende do teto dado à idade da empresa (60 meses no modelo principal): sem teto, as empresas jovens ficam abaixo da média. No que diz respeito à hipótese H3, esta não encontrou apoio, já que os sinalizados têm proporcionalmente menos fornecedores sancionados que os demais registros, o contrário do que a hipótese previa. Quando o valor do contrato é controlado, essa diferença cai para cerca de metade e se torna incerta.

Os órgãos municipais ficaram com taxa de sinalização menor que a das demais esferas, mas a diferença deixa de ser clara quando se desfazem as correções para grupos pequenos; a pouca cobertura dos municípios continua a justificar a abordagem, embora os resultados não indiquem maior atipicidade neles. Nos estudos de caso, a explicação individual expôs, no primeiro, o domínio de um fornecedor sobre o órgão e um artefato de construção (o valor contado duas vezes) e, no segundo, uma empresa recém-aberta com indício fraco de especialização.

Como contribuição, o trabalho oferece um *pipeline* de coleta, tratamento e modelagem construído apenas sobre dados públicos e um protocolo para avaliar uma triagem sem rótulos: variáveis neutras, troca de sementes, correções desfeitas uma a uma, injeção de padrões e cruzamento com cadastros de sanções. Por fim, mede o quanto o conjunto sinalizado depende da construção das variáveis: sem o logaritmo e o corte no valor atípico, só 60,4% dos sinalizados se mantêm.

A principal limitação é a ausência de rótulos, que impede medir a precisão do modelo. A utilidade das explicações não foi testada com auditores. A consulta de itens foi cortada em 10 por contratação, o que subestima o valor de parte dos registros, e as pessoas físicas ficaram fora do modelo. P2 e P4 usam informação de todo o período, que um auditor não teria na data da contratação, e o mínimo de 20 registros em P2 deslocou o efeito dos órgãos pequenos para os órgãos logo acima do corte. Em P4, o HHI de fornecedores com dois registros chega ao máximo com facilidade, e o peso desses fornecedores nas sinalizações não foi medido. Um único fornecedor ocupa 15,6% da lista, e a categorização dos objetos por texto é aproximada e pode esconder valores atípicos nas categorias heterogêneas.

Esses limites indicam os próximos passos. A recoleta paginada dos itens eliminaria o truncamento e permitiria comparar o preço unitário de cada item no lugar do valor total, e um teste com auditores mostraria se a lista e as explicações ajudam a escolher o que examinar. Falta, sobretudo, uma referência externa melhor que os cadastros de sanções, como achados de auditoria dos tribunais de contas.

Os registros sinalizados são indicadores estatísticos de atipicidade, não acusações. A apuração de eventuais irregularidades cabe aos órgãos de controle e de investigação. Dentro desses limites, a abordagem permite ordenar contratações segundo padrões definidos *a priori* e mostrar por que o modelo priorizou cada uma.

AGRADECIMENTOS E FINANCIAMENTO

Esta pesquisa não recebeu financiamento.

REFERÊNCIAS

ABREU, Breno M. de; PEREIRA, Thomaz H. S.; GOMES JUNIOR, Luiz. Detecção de fraudes em licitações públicas: uma comparação de modelos de detecção de anomalias. In: ESCOLA REGIONAL DE BANCO DE DADOS, 19., 2024, Farroupilha. *Anais [...]*. Porto Alegre: Sociedade Brasileira de Computação, 2024. p. 81-90. DOI: https://doi.org/10.5753/erbd.2024.238821.

BARREDO ARRIETA, Alejandro et al. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, v. 58, p. 82-115, 2020. DOI: https://doi.org/10.1016/j.inffus.2019.12.012.

BORGES, Alessandro de Oliveira et al. ALICE: análise automatizada de contratações públicas para prevenção de irregularidades e combate à corrupção. *Cadernos Técnicos da CGU*, v. 8, p. 11-16, 2026. Disponível em: https://revista.cgu.gov.br/Cadernos_CGU/article/view/867. Acesso em: 22 out. 2026.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). *Diário Oficial da União: seção 1*, Brasília, DF, 15 ago. 2018.

BRASIL. Lei nº 14.133, de 1º de abril de 2021. Lei de Licitações e Contratos Administrativos. *Diário Oficial da União: seção 1*, Brasília, DF, 1 abr. 2021. Edição extra-F.

BRASIL. Lei Complementar nº 198, de 28 de junho de 2023. Altera a Lei Complementar nº 91, de 22 de dezembro de 1997 [...]; e altera a Lei nº 14.133, de 1º de abril de 2021 (Lei de Licitações e Contratos Administrativos). *Diário Oficial da União: seção 1*, Brasília, DF, 28 jun. 2023. Edição extra.

BRASIL. Controladoria-Geral da União. *Alice: Analisador de Licitações, Contratos e Editais*. Brasília, DF: CGU, 2024a. Disponível em: https://www.gov.br/cgu/pt-br/assuntos/auditoria-e-fiscalizacao/alice. Acesso em: 22 out. 2026.

BRASIL. Controladoria-Geral da União. *CGU disponibiliza ferramenta Alice para estados e municípios identificarem licitações com suspeitas de irregularidades*. Brasília, DF: CGU, 2024b. Notícia publicada em 15 maio 2024. Disponível em: https://www.gov.br/cgu/pt-br/assuntos/noticias/2024/05/cgu-disponibiliza-ferramenta-alice-para-estados-e-municipios-identificarem-licitacoes-com-suspeitas-de-irregularidades. Acesso em: 22 out. 2026.

BRASIL. Controladoria-Geral da União. *Portal da Transparência: download de dados do CEIS e do CNEP*. Brasília, DF: CGU, 2026. Disponível em: https://portaldatransparencia.gov.br/download-de-dados/ceis. Acesso em: 15 jul. 2026.

BRASIL. Tribunal de Contas da União. *Referencial de combate a fraude e corrupção: aplicável a órgãos e entidades da Administração Pública*. 2. ed. Brasília, DF: TCU, 2018.

BRASILAPI. *API de consulta de CNPJ*. [*S. l.*]: BrasilAPI, 2026. Disponível em: https://brasilapi.com.br/docs. Acesso em: 5 nov. 2026.

CHANDOLA, Varun; BANERJEE, Arindam; KUMAR, Vipin. Anomaly detection: a survey. *ACM Computing Surveys*, v. 41, n. 3, p. 15:1-15:58, 2009. DOI: https://doi.org/10.1145/1541880.1541882.

FAZEKAS, Mihály; KOCSIS, Gábor. Uncovering high-level corruption: cross-national objective corruption risk indicators using public procurement data. *British Journal of Political Science*, v. 50, n. 1, p. 155-164, 2020. DOI: https://doi.org/10.1017/S0007123417000461.

FIELD, Christopher A.; WELSH, Alan H. Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v. 69, n. 3, p. 369-390, 2007. DOI: https://doi.org/10.1111/j.1467-9868.2007.00593.x.

HAWKINS, Douglas M. *Identification of outliers*. London: Chapman and Hall, 1980.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Sistema de contas regionais: Brasil 2022*. Rio de Janeiro: IBGE, 2024. Disponível em: https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9054-contas-regionais-do-brasil.html. Acesso em: 22 out. 2026.

JACCARD, Paul. The distribution of the flora in the alpine zone. *New Phytologist*, v. 11, n. 2, p. 37-50, 1912. DOI: https://doi.org/10.1111/j.1469-8137.1912.tb05611.x.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi-Hua. Isolation Forest. In: IEEE INTERNATIONAL CONFERENCE ON DATA MINING, 8., 2008, Pisa. *Proceedings [...]*. Los Alamitos: IEEE Computer Society, 2008. p. 413-422. DOI: https://doi.org/10.1109/ICDM.2008.17.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi-Hua. Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data*, v. 6, n. 1, p. 3:1-3:39, 2012. DOI: https://doi.org/10.1145/2133360.2133363.

LLOYD, Stuart P. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, v. 28, n. 2, p. 129-137, 1982. DOI: https://doi.org/10.1109/TIT.1982.1056489.

LUNDBERG, Scott M. et al. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, v. 2, n. 1, p. 56-67, 2020. DOI: https://doi.org/10.1038/s42256-019-0138-9.

LUNDBERG, Scott M.; LEE, Su-In. A unified approach to interpreting model predictions. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 31., 2017, Long Beach. *Advances in Neural Information Processing Systems 30*. Red Hook: Curran Associates, 2017. p. 4765-4774.

MAIA, Patrícia et al. Auditing government purchases with a multicriteria anomaly detection strategy. *Journal of Information and Data Management*, v. 11, n. 1, p. 50-65, 2020. DOI: https://doi.org/10.5753/jidm.2020.2029.

MORAIS, Vinícius et al. Previsão de fraude em licitações no Brasil. *Cadernos de Finanças Públicas*, v. 24, n. 3, p. 1-32, 2024. Disponível em: https://publicacoes.tesouro.gov.br/index.php/cadernos/article/view/256. Acesso em: 22 out. 2026.

NEVES, Fabrício Ramos; SILVA, Polyana Batista da; CARVALHO, Hugo Leonardo Menezes de. Mulheres artificiais contra a corrupção: em busca de legitimidade no Tribunal de Contas da União. *Revista de Contabilidade e Organizações*, v. 13, p. 31-50, 2019. DOI: https://doi.org/10.11606/issn.1982-6486.rco.2019.158530.

PIAUÍ. Tribunal de Contas do Estado. *Relatório de levantamento da gestão tributária dos municípios piauienses: exercícios de referência 2023 e 2024*. Teresina: TCE-PI, 2025. Processo TC/013031/2025. Disponível em: https://www.tcepi.tc.br/wp-content/uploads/2026/01/Levantamento-Gestao-Tributaria.pdf. Acesso em: 22 out. 2026.

PORTAL NACIONAL DE CONTRATAÇÕES PÚBLICAS. *APIs de dados abertos do PNCP*. Brasília, DF: PNCP, 2026. Compreende a API de consulta (contratações e contratos), no endereço indicado abaixo, e a API principal (itens e resultados), em https://pncp.gov.br/api/pncp/swagger-ui/index.html. Disponível em: https://pncp.gov.br/api/consulta/swagger-ui/index.html. Acesso em: 4 nov. 2026.

RAMALHO, Hilton Martins Brito; ALMEIDA, Aléssio Tony Cavalcanti de; FRAGA, Alcimar Alves. Detecção de casos suspeitos de conluio em licitações públicas: uma aplicação do algoritmo a priori de aprendizado de máquina para o estado da Paraíba. *Teoria e Prática em Administração*, v. 10, n. 2, p. 5-22, 2020. DOI: https://doi.org/10.21714/2238-104X2020v10i2-51526.

SALTON, Gerard; BUCKLEY, Christopher. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, v. 24, n. 5, p. 513-523, 1988. DOI: https://doi.org/10.1016/0306-4573(88)90021-0.

SCORALICK, Lucas D.; BRANDÃO, Diego N.; BELLOZE, Kele T. Aprimoramento de modelos para detecção de conluios em licitações públicas brasileiras com variáveis estatísticas e modelos explicáveis. In: SIMPÓSIO BRASILEIRO DE BANCOS DE DADOS, 39., 2024, Florianópolis. *Anais [...]*. Porto Alegre: Sociedade Brasileira de Computação, 2024. p. 680-686. DOI: https://doi.org/10.5753/sbbd.2024.243170.