

Inteligência de Dados Aplicada à Governança das Contratações Públicas: Evidências de Concentração de Mercado e Modelagem Preditiva de Riscos no Estado do Rio de Janeiro

Data Intelligence Applied to Public Procurement Governance: Evidence of Market Concentration and Predictive Risk Modeling in the State of Rio de Janeiro

Rafael Cocchiarelli

UERJ – Universidade do Estado do Rio de Janeiro
cocchiarelli@bol.com.br

José Francisco Moreira Pessanha

UERJ – Universidade do Estado do Rio de Janeiro
jose.pessanha@uerj.br

Nathalia Cordeiro

UFRJ – Universidade Federal do Rio de Janeiro
nathalia.cordeiro@gmail.com

Antonio Ranha

UERJ – Universidade do Estado do Rio de Janeiro
ranha@bol.com.br

Submissão: 10/01/2026

Aprovação: 27/05/2026

RESUMO

Este artigo analisa a governança das contratações públicas no Estado do Rio de Janeiro por meio da integração de indicadores econômicos clássicos e modelos preditivos de ciência de dados e inteligência artificial. Com base em dados reais de contratações públicas realizadas entre 2022 e 2025, o estudo investiga padrões de concentração de mercado e desigualdade na distribuição dos recursos públicos utilizando o Índice de Herfindahl-Hirschman (IHH) e o Coeficiente de Gini. Adicionalmente, emprega modelos de regressão linear e o algoritmo de aprendizagem de máquina *XGBoost* para classificação preditiva de contratações com maior probabilidade de risco. Os resultados revelam estruturas de mercado altamente concentradas em determinados serviços, bem como assimetrias significativas na alocação dos recursos. A modelagem preditiva demonstrou elevada capacidade de identificar *outliers* e contratações inconsistentes e, reforçando seu potencial como instrumento de fiscalização preventiva. As evidências indicam que a inteligência de dados pode fortalecer a governança das contratações públicas, promovendo decisões mais eficientes, transparentes e alinhadas aos princípios da Lei nº 14.133/2021.

Palavras-chave: Governança Pública; Contratações Públicas; Inteligência de Dados; Concentração de Mercado; Inteligência Artificial.

ABSTRACT

This article examines the governance of public procurement in the State of Rio de Janeiro through the integration of classical economic indicators and predictive models from data science and artificial intelligence. Based on real data from public procurement processes conducted between 2022 and 2025, the study investigates patterns of market concentration and inequality in the allocation of public resources using the Herfindahl–Hirschman Index (HHI) and the Gini coefficient. Additionally, linear regression models are employed as a statistical reference, and the XGBoost machine learning algorithm is applied for the predictive classification of procurements with a higher probability of risk. The results reveal highly concentrated market structures in specific services, as well as significant asymmetries in resource allocation. The predictive modeling demonstrated a strong capacity to identify outliers and inconsistent procurement processes, reinforcing its potential as a tool for preventive oversight. The evidence indicates that data intelligence can strengthen public procurement governance by promoting more efficient, transparent, and evidence-based decision-making aligned with the principles of Law No. 14,133/2021.

Keywords: Public Governance; Public Procurement; Data Intelligence; Market Concentration; Artificial Intelligence.

1. Introdução

Nas últimas décadas, a administração pública tem enfrentado crescentes pressões por maior eficiência, transparência e *accountability*, especialmente em contextos de restrição fiscal e complexidade institucional. As contratações públicas, responsáveis por parcela significativa do gasto governamental, tornaram-se um dos principais focos de atenção de gestores, órgãos de controle e da sociedade. (CUNHA, 2024)

A literatura contemporânea aponta que a governança pública deixou de se restringir ao cumprimento formal de normas, passando a incorporar práticas orientadas por evidências, inovação e aprendizado organizacional (OLIVEIRA, 2024). Nesse cenário, a inteligência de dados emerge como um vetor estratégico, ao permitir a transformação de grandes volumes de dados administrativos em conhecimento útil para a tomada de decisões (SANTOS, 2023).

No Brasil, a promulgação da Lei nº 14.133/2021 reforçou a centralidade do planejamento, da gestão de riscos e da governança nas contratações públicas. Entretanto, persistem desafios estruturais, como concentração de mercado, baixa competitividade e assimetrias na distribuição dos recursos, que comprometem a eficiência do gasto público (MIOTTO, 2024).

Diante disso, este artigo busca responder à seguinte questão: como a aplicação integrada de métodos estatísticos e algoritmos de aprendizado de máquina pode contribuir para a detecção de riscos, inconsistências e ineficiências nas contratações públicas? Para tanto, analisa-se o caso do Estado do Rio de Janeiro, utilizando uma abordagem empírica baseada em inteligência de dados.

2. Referencial Teórico

2.1 Governança Pública e Valor Público

A Nova Governança Pública desloca o foco da administração pública do controle burocrático para a geração de valor público, enfatizando eficiência, transparência e participação social (CAVALCANTE, 2024; CUNHA, FERREIRA & MURARI, 2024). Nesse modelo, a tomada de decisão baseada em dados torna-se essencial para reduzir assimetrias informacionais e fortalecer a *accountability*.

A Lei nº 14.133/2021 introduziu instrumentos como o Plano Anual de Contratações, a gestão de riscos e a responsabilização da alta administração, reforçando a governança das contratações. Contudo, a efetividade desses instrumentos depende da capacidade analítica do Estado em monitorar padrões de mercado, preços e comportamentos atípicos.

2.2 Inteligência de Dados

A inteligência de dados configura-se como um elemento estratégico central na Quarta Revolução Industrial, ao integrar tecnologias como Big Data, Inteligência Artificial (IA), aprendizado de máquina e *Business Intelligence* para transformar grandes volumes de dados em conhecimento acionável para a tomada de decisões organizacionais. Conforme destacam Góis e Carvalho (2025), essa abordagem permite não apenas a automação de processos e a

análise preditiva, mas também a identificação de padrões ocultos capazes de orientar estratégias em tempo real, ampliando a eficiência operacional e a personalização de serviços.

Ela combina estatística, ciência de dados e aprendizado de máquina para apoiar decisões em ambientes complexos. Na administração pública pode ser compreendida como parte de um processo histórico mais amplo de evolução das redes de informação, no qual a capacidade de coletar, processar e interpretar dados molda a forma como as sociedades tomam decisões. Como observa Harari (2024), as redes de informação sempre desempenharam um papel central no desenvolvimento humano, e na era da inteligência artificial, essa dinâmica atinge um novo patamar, permitindo análises preditivas e decisões mais precisas baseadas em grandes volumes de dados.

No setor público, essas técnicas têm sido utilizadas para detecção de fraudes, avaliação de políticas e análise de riscos, ainda de forma incipiente no campo das contratações públicas brasileiras. (COCCHIARELLI, 2024)

2.3 Índice Herfindahl-Hirschman e o Coeficiente de Gini

O Índice de Herfindahl-Hirschman (IHH), amplamente utilizado em economia e políticas de defesa da concorrência, é uma métrica destinada a avaliar o grau de concentração de mercado a partir da soma dos quadrados das participações relativas de cada agente. Originalmente desenvolvido por Orris C. Herfindahl (1950), valores próximos a zero indicam maior dispersão e competição, enquanto índices mais elevados sinalizam concentração e até mesmo situações de oligopólio ou monopólio.

$$IHH = \left(\sum_{i=1}^n s_i^2 \right) 10000 \quad (1)$$

em que:

s_i corresponde a participação relativa do agente i no total.

n corresponde ao número de agentes, nesse caso, fornecedores.

O índice varia entre 0 (mercado perfeitamente pulverizado) e 10.000 (monopólio absoluto, quando apenas um agente detém 100% do mercado).

O valor amplamente aceito como limite de concentração é 2.500.

Embora elevados níveis de concentração possam representar potenciais riscos concorrenciais, é importante reconhecer que a concentração de mercado, por si só, não implica necessariamente falhas de governança ou irregularidades nas contratações públicas. Em determinados setores da economia, características estruturais como economias de escala, barreiras tecnológicas de entrada, especialização produtiva ou exigências regulatórias podem naturalmente resultar em mercados com menor número de fornecedores (PORTER, 2008).

No contexto das compras públicas, serviços altamente especializados (como tecnologia da informação, equipamentos hospitalares ou determinados serviços técnicos) frequentemente apresentam estruturas de oferta naturalmente mais concentradas. Nessas situações, níveis elevados de concentração podem refletir condições estruturais do mercado e não necessariamente práticas anticompetitivas ou falhas institucionais.

Dessa forma, os indicadores de concentração empregados neste estudo devem ser interpretados como sinais analíticos ou indicadores de alerta, e não como evidência direta de irregularidade. A identificação de estruturas altamente concentradas deve ser compreendida como um ponto de partida para análises adicionais, envolvendo avaliação institucional, características do setor econômico e análise qualitativa dos processos de contratação.

Essa abordagem é consistente com a literatura de governança pública e análise de compras governamentais, que destaca a importância de combinar indicadores quantitativos com análise contextual e institucional para interpretação adequada dos resultados (OECD, 2019).

O Coeficiente de Gini, desenvolvido originalmente como medida de desigualdade de renda (GINI, 1912), consolidou-se como uma métrica universal para avaliar a concentração e a distribuição de recursos em diferentes domínios. Seu valor varia entre 0 e 1: quanto mais próximo de 0, maior a igualdade; quanto mais próximo de 1, maior a desigualdade.

$$Gini = n \sum i = 1 n x_i \sum i 1 n (2i - n - 1) x_i \quad (2)$$

onde:

x_i é o valor contratado por cada fornecedor.

n é o número total de fornecedores.

No âmbito das contratações públicas, o Coeficiente de Gini pode ser aplicado tanto à distribuição dos contratos entre fornecedores quanto à alocação dos recursos entre órgãos ou regiões. Essa adaptação metodológica permite visualizar em que medida o Estado distribui seus gastos de forma equitativa ou, ao contrário, concentra investimentos em poucos agentes ou setores, criando distorções socioeconômicas e riscos de dependência administrativa.

2.3 Regressão Linear e o Algoritmo XGBOOST

A regressão linear é uma das técnicas estatísticas mais consagradas e amplamente aplicadas na análise de dados quantitativos, especialmente em contextos onde se busca compreender relações entre variáveis e realizar previsões baseadas em padrões observados. Sua aplicação perpassa áreas como economia, administração pública, saúde, engenharia e educação, configurando-se como uma ferramenta fundamental para o diagnóstico, a modelagem e a tomada de decisões baseadas em evidências.

Para Chein (2019), a lógica central do modelo de regressão linear está associada ao princípio do *ceteris paribus*, expressão latina que significa “tudo o mais constante”. Isso implica que, ao estimar o efeito de uma variável independente sobre uma variável dependente, o modelo estima que todas as demais variáveis se mantêm constantes. Essa premissa é essencial para isolar e interpretar o impacto individual de cada fator no fenômeno observado, o que o torna particularmente útil em contextos complexos como as contratações públicas.

De forma geral, o modelo clássico de regressão linear assume uma relação linear entre as variáveis explicativas ($X_1, X_2, \dots, X_n$) e a variável dependente (Y), representada pela equação:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (3)$$

O avanço das tecnologias de inteligência artificial (IA) e do aprendizado de máquina tem revolucionado o modo como organizações públicas e privadas tomam decisões baseadas em dados. Entre as diversas técnicas disponíveis, o *XGBoost* (*eXtreme Gradient Boosting*) tem se consolidado como uma das ferramentas mais eficazes e versáteis para a análise preditiva e a detecção de anomalias.

Desenvolvido por Tianqi Chen e Carlos Guestrin (2016), o *XGBoost* foi concebido como um sistema de *tree boosting* escalável, capaz de lidar com grandes volumes de dados de maneira eficiente e precisa. Desde sua criação, o algoritmo vem alcançando resultados expressivos em desafios de ciência de dados e em aplicações práticas nos setores financeiro, energético e governamental (CHEN; GUESTRIN, 2016).

Além de seu desempenho superior, o *XGBoost* é amplamente reconhecido pela robustez, velocidade de processamento e capacidade de generalização, o que o torna adequado para aplicações complexas, como previsão de comportamento, detecção de fraudes e identificação de padrões ocultos em bases de dados extensas e desbalanceadas.

O *XGBoost* pertence à família dos métodos de *boosting*, cuja lógica baseia-se em treinar modelos de forma sequencial, cada novo modelo busca corrigir os erros cometidos pelos anteriores. Esse processo gera um modelo final mais robusto e preciso, capaz de capturar relações não lineares entre as variáveis.

De acordo com Chen e Guestrin, o *XGBoost* incorpora uma série de inovações que o diferenciam de versões anteriores do *gradient boosting*, entre as quais se destacam:

- Regularização explícita, que reduz o risco de sobreajuste (*overfitting*);
- Otimização paralela e distribuída, aumentando a escalabilidade;
- Tratamento de dados esparsos, garantindo eficiência mesmo em bases com variáveis faltantes; e o uso de um esquema de quantil ponderado, que acelera a busca por divisões ideais nas árvores de decisão.

Essas características conferem ao *XGBoost* um equilíbrio notável entre precisão e eficiência computacional, tornando-o uma das principais escolhas entre cientistas de dados e analistas de políticas públicas.

3. Metodologia

A escolha dos métodos analíticos empregados neste estudo baseia-se na necessidade de combinar indicadores estruturais de mercado com técnicas de modelagem preditiva, permitindo uma análise simultaneamente descritiva e prospectiva das contratações públicas.

Inicialmente, foram utilizados o Índice de Herfindahl-Hirschman (IHH) e o Coeficiente de Gini, amplamente empregados na literatura econômica e de políticas públicas para mensurar, respectivamente, níveis de concentração de mercado e desigualdade na distribuição de recursos (HIRSCHMAN, 1945; HERFINDAHL, 1950; COWELL, 2011). O IHH é frequentemente utilizado por autoridades antitruste e por organismos reguladores para avaliar estruturas competitivas e identificar possíveis riscos associados à dominância de fornecedores (ESTADOS UNIDOS, 1997). No contexto das contratações públicas, esse indicador permite avaliar se a participação de fornecedores se encontra excessivamente concentrada em determinados agentes econômicos.

O Coeficiente de Gini, por sua vez, complementa essa análise ao capturar a assimetria na distribuição dos valores contratados de forma direta, evidenciando possíveis desequilíbrios na alocação de recursos públicos entre fornecedores. Embora tradicionalmente utilizado em

estudos de desigualdade de renda, sua aplicação tem se expandido para análises de distribuição de contratos públicos e concentração de fornecedores em compras governamentais.

A utilização conjunta desses dois indicadores permite uma leitura estrutural do mercado de contratações, identificando padrões de concentração e desigualdade que podem sinalizar riscos potenciais à concorrência e à eficiência da política pública.

Em complemento à análise estrutural, foi empregada a regressão linear como ferramenta de modelagem estatística para identificar tendências e relações entre variáveis associadas às contratações públicas. A regressão linear constitui um dos métodos mais consolidados em análise econométrica e permite estimar relações entre variáveis explicativas e indicadores de desempenho ou risco (WOOLDRIDGE, 2016).

Por fim, adotou-se o algoritmo XGBoost (*Extreme Gradient Boosting*) para fins de modelagem preditiva. O XGBoost é amplamente reconhecido na literatura de ciência de dados por sua elevada capacidade preditiva e robustez em conjuntos de dados estruturados, sendo frequentemente superior a métodos tradicionais em tarefas de classificação e regressão (CHEN & GUESTRIN, 2016). O algoritmo baseia-se em técnicas de *gradient boosting*, nas quais múltiplas árvores de decisão são combinadas de forma iterativa para reduzir erros de previsão e capturar relações não lineares complexas.

A escolha desse modelo justifica-se pela necessidade de identificar padrões complexos de risco nas contratações públicas, que frequentemente envolvem múltiplas variáveis institucionais, econômicas e operacionais. Diferentemente de modelos estatísticos lineares, algoritmos de *boosting* possuem maior capacidade de lidar com interações não lineares, heterogeneidade de dados e *outliers*, características frequentemente presentes em bases administrativas de grande escala.

Assim, a integração entre indicadores econômicos tradicionais e técnicas modernas de aprendizado de máquina permite uma abordagem analítica mais abrangente, combinando diagnóstico estrutural do mercado com mecanismos de previsão de risco, o que contribui para o fortalecimento da governança das contratações públicas.

Para a identificação preditiva de contratações com maior probabilidade de risco, foi aplicado o algoritmo XGBoost, amplamente utilizado em tarefas de classificação em bases de dados estruturadas.

O modelo foi treinado utilizando variáveis relacionadas às características das contratações, incluindo valor do contrato, modalidade de contratação, quantidade de fornecedores participantes, tipo de objeto contratual e histórico de participação dos fornecedores.

Para garantir a robustez do modelo, foi adotado o procedimento de validação cruzada (*cross-validation*), técnica amplamente utilizada em aprendizado de máquina para avaliar a capacidade de generalização dos modelos preditivos (HASTIE, TIBSHIRANI & FRIEDMAN, 2009). Esse procedimento consiste na divisão do conjunto de dados em subconjuntos de treinamento e teste, permitindo avaliar o desempenho do modelo em diferentes amostras.

O desempenho do modelo foi avaliado por meio de métricas clássicas de classificação, incluindo:

- Acurácia, que mede a proporção de classificações corretas;
- Precisão, que avalia a proporção de previsões positivas corretamente identificadas;
- Recall, que indica a capacidade do modelo de identificar casos relevantes;
- F1-score, que combina precisão e recall em uma métrica única de desempenho.

Essas métricas são amplamente utilizadas na literatura de aprendizado de máquina para avaliação de modelos de classificação e permitem avaliar não apenas a capacidade preditiva geral do modelo, mas também sua sensibilidade na identificação de casos potencialmente críticos (CHEN & GUESTRIN, 2016).

A metodologia de desenvolvimento do algoritmo XGBoost foi estruturada segundo o modelo CRISP-DM (CHAPMAN, 2000) adaptado ao contexto das contratações públicas.

- Entendimento do Negócio - Etapa inicial dedicada à compreensão institucional sobre as contratações públicas
- Entendimento dos Dados - Etapa de compreensão das bases que serão utilizadas nas análises.
- Transformação dos Dados - Etapa de limpeza, cruzamento de bases, enriquecimento dos dados, padronização de variáveis e eliminação de inconsistências nos dados.
- Modelagem
 - a) Análise Exploratória de Dados e Treinamento do Modelo XGBoost
 - b) Aplicação dos Índices IHH e Gini
 - c) Aplicação da Regressão Linear
 - d) Aplicação do Algoritmo XGBoost nos processos outliers
- Avaliação - Etapa de interpretação e análise dos resultados obtidos

Os dados foram coletados exclusivamente por meio de portais de transparência estaduais, garantindo rastreabilidade e conformidade legal.

É importante destacar que foram excluídos os processos de inexigibilidade da base de dados utilizada neste estudo, não abrangendo processos nos quais a legislação prevê a existência de fornecedor único e, portanto, inviabiliza qualquer análise de competição ou estrutura de mercado. Assim, todos os registros analisados pertencem a categorias de contratação nas quais, há pluralidade potencial de fornecedores, permitindo a aplicação adequada de métricas como o Índice de Herfindahl-Hirschman (IHH) e o Coeficiente de Gini.

Ademais, o modelo preditivo foi treinado com dados do período 2022 a 2025, recorte estabelecido por dois motivos principais: (i) garantir o uso de informações mais recentes, especialmente após a consolidação das normas e práticas introduzidas pela Lei nº 14.133/2021; e (ii) viabilizar o processamento computacional, uma vez que o volume total de dados históricos é elevado e poderia comprometer tempo de execução, capacidade de memória e reprodutibilidade dos modelos. Esse recorte temporal preserva a atualidade, melhora a consistência das variáveis e permite análises mais alinhadas ao cenário vigente das contratações públicas do Estado.

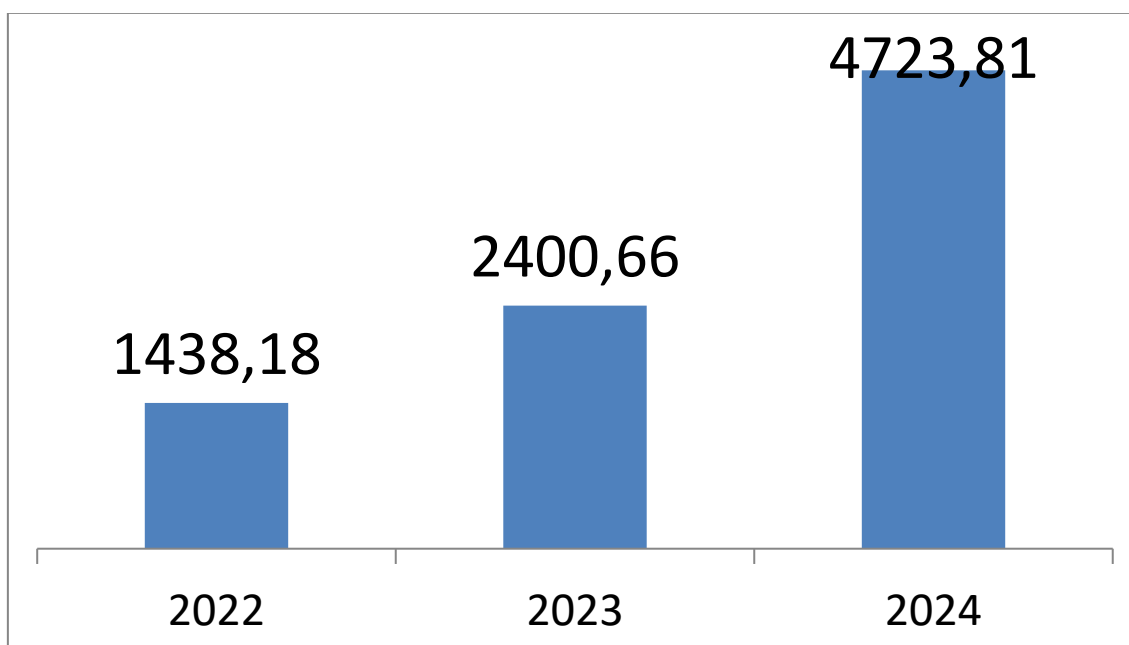
4. Resultados

Os resultados empíricos evidenciam padrões estruturais de elevada concentração de mercado em serviços estratégicos da administração pública estadual, com impactos diretos sobre a competitividade, a eficiência alocativa e os riscos à governança das contratações. A análise integrada dos indicadores IHH e Gini revela não apenas concentração de fornecedores, mas também forte assimetria na distribuição dos recursos públicos, configurando um ambiente institucional marcado por dependência excessiva de poucos agentes econômicos em setores sensíveis.

o serviço de apoio hospitalar (Gini = 0,72), o fornecimento de refeições (Gini = 0,87) e a limpeza predial escolar (Gini = 0,99) apresentam concentração extrema, indicando que a quase totalidade dos recursos está direcionada a um número muito reduzido de fornecedores. Esses resultados sugerem riscos relevantes à competitividade, à eficiência alocativa e à governança das contratações públicas, podendo sinalizar dependência excessiva de determinados fornecedores, barreiras à entrada de novos competidores e maior vulnerabilidade a práticas oportunistas.

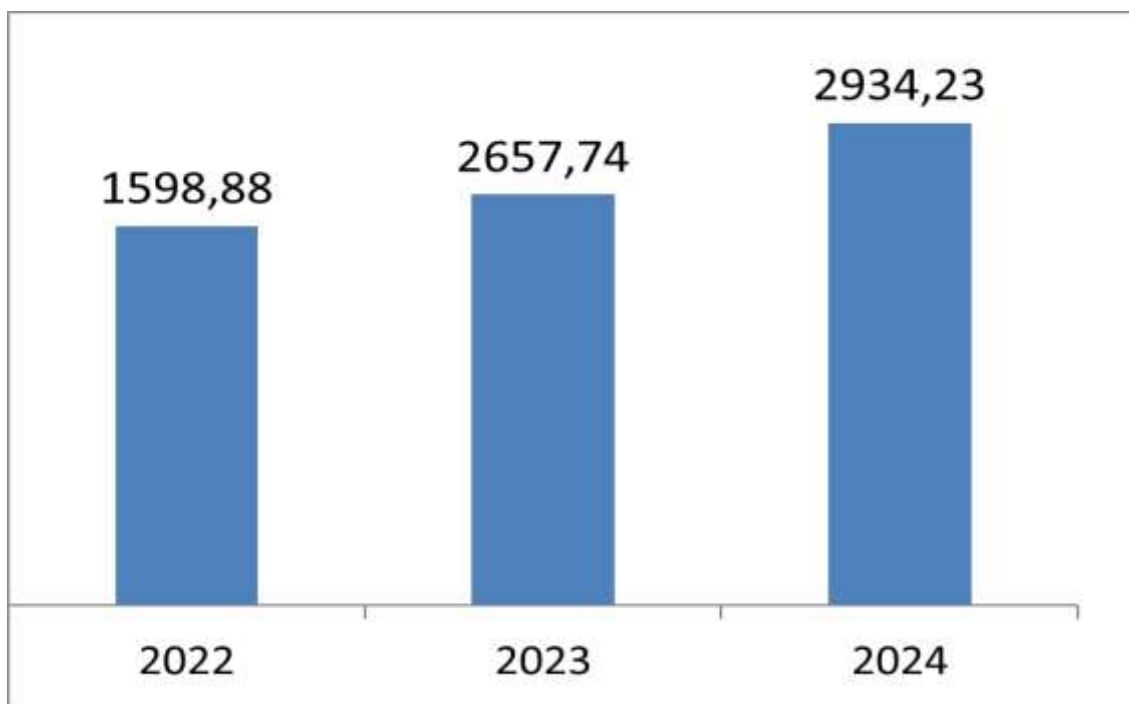
No contexto da gestão pública, tais níveis de concentração reforçam a necessidade de instrumentos de inteligência de dados e mecanismos de controle preventivo, capazes de subsidiar decisões estratégicas, promover maior diversificação de fornecedores e mitigar riscos estruturais nas contratações.

Figura 1 - IHH aplicado ao serviço de apoio hospitalar



Fonte: autoria própria, 2025

Figura 2 - IHH aplicado ao serviço de fornecimento de refeição



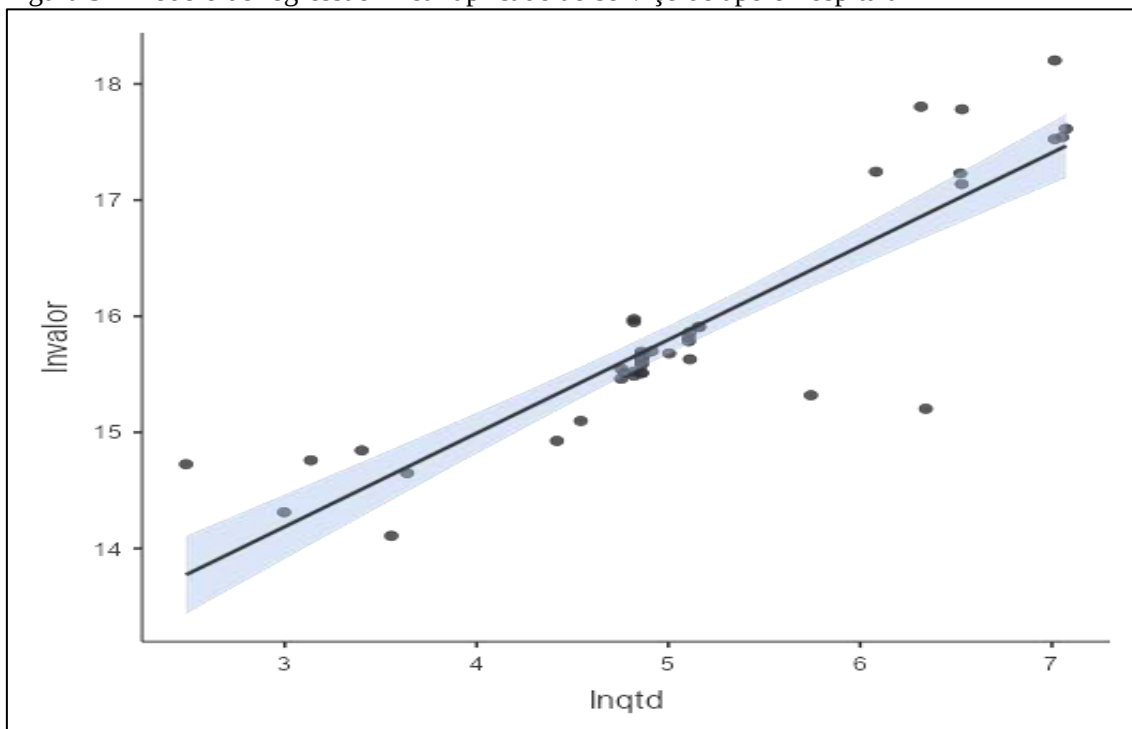
Fonte: autoria própria, 2025

Em contraste com os setores altamente concentrados, identificou-se o Serviço de Operação Geral de UTI como um caso de estrutura de mercado significativamente mais competitiva e institucionalmente mais consistente. Entre 2023 e 2024, esse serviço movimentou mais de R\$ 600 milhões em contratações, distribuídos entre mais de 30 fornecedores atuantes, configurando um ambiente de concorrência efetiva. Os indicadores quantitativos corroboram esse diagnóstico: o Índice de Herfindahl-Hirschman ($IHH = 0.490$) indica baixa concentração de mercado, enquanto o Coeficiente de Gini (0,21) revela uma distribuição equilibrada dos recursos públicos entre os fornecedores.

Em termos comparativos, esse arranjo contratual apresenta desempenho institucional superior aos demais serviços analisados, nos quais se observaram estruturas altamente concentradas e assimetrias severas na alocação dos recursos. Esse contraste evidencia que níveis mais elevados de competitividade são factíveis no âmbito das contratações públicas estaduais, reforçando a importância de políticas e práticas administrativas que favoreçam ambientes concorrenciais mais equilibrados e alinhados aos princípios da eficiência, isonomia e economicidade previstos na Lei nº 14.133/2021.

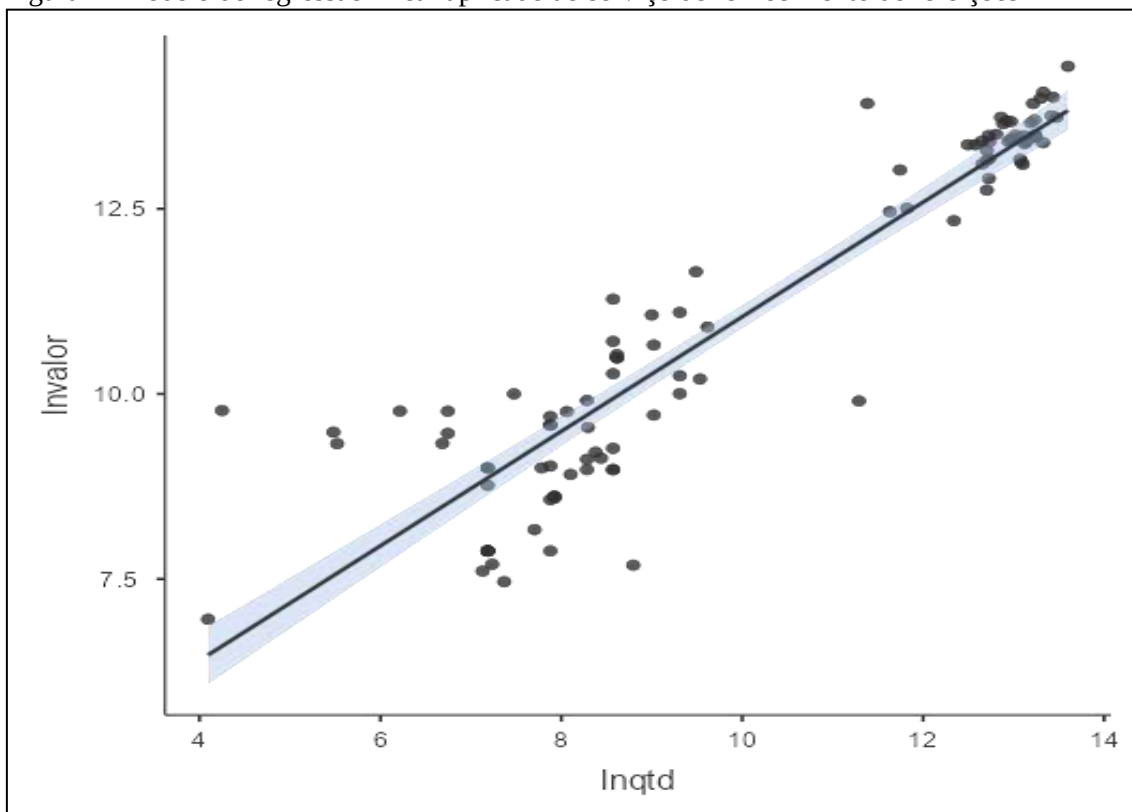
A regressão linear identificou *outliers* relevantes nos valores contratados, sugerindo potenciais inconsistências que demandam atenção dos órgãos de controle.

Figura 3 - Modelo de regressão linear aplicado ao serviço de apoio hospitalar



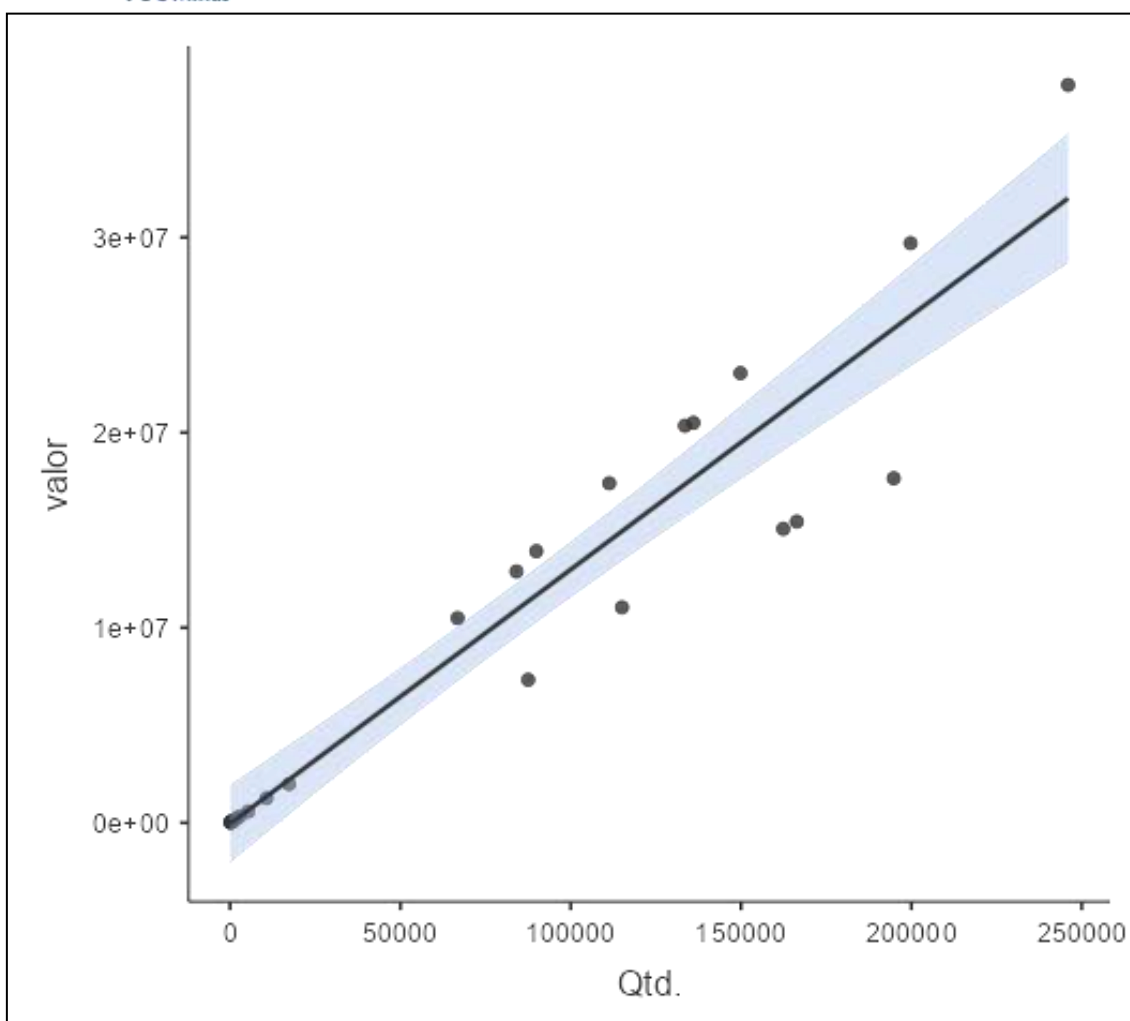
Fonte: autoria própria, 2025

Figura 4- Modelo de regressão linear aplicado ao serviço de fornecimento de refeições



Fonte: autoria própria, 2025

Figura 5- Modelo de regressão linear aplicado ao serviço de limpeza predial escolar



Fonte: autoria própria, 2025

A base utilizada para o treinamento do modelo preditivo *XGBoost* reúne 18.880 registros de contratações diretas, contendo informações estruturadas sobre fornecedores, itens, valores, unidades demandantes e, especialmente, o desfecho da execução contratual representado pela variável MOTIVO. Essa variável sintetiza o resultado da execução do contrato e assume quatro categorias:

- Contratação OK, contratações concluídas sem anomalias;
- Atraso injustificado na execução;
- Falha ou Possibilidade de ocorrência de Fraude na execução do contrato;
- Inexecução total ou parcial do contrato.

A categoria Contratação OK representa 52% da amostra (9.826 registros), indicando que pouco mais da metade das contratações ocorreu sem intercorrências relevantes. As categorias que representam algum nível de ocorrência somam 48%, distribuídas da seguinte forma:

- Atraso injustificado: 6% (1.184 registros);
- Falha ou Fraude na execução do contrato: 19% (3.572 registros);
- Inexecução total ou parcial: 23% (4.298 registros).

Essa composição inicial já oferece indícios relevantes para compreender o

funcionamento das contratações públicas no período observado. O fato de quase metade dos registros apresentar algum tipo de ocorrência (desde atrasos leves até falhas graves e indícios de fraude) aponta para um ambiente contratual suscetível a riscos operacionais e comportamentais, reforçando a importância de estratégias analíticas de monitoramento.

Quadro 1 - Resultados da aplicação do modelo XGBoost numa amostra de processos outliers

SEI-XX0001/0xxx01/2024	INEXECUÇÃO TOTAL OU PARCIAL
SEI-XX0007/0xxx22/2024	CONTRATAÇÃO OK
SEI-XX0005/0xxx9/2024	INEXECUÇÃO TOTAL OU PARCIAL
SEI-XX0007/0xxx04/2024	ATRASSO INJUSTIFICADO
SEI-XX0004/0xxx9/2024	CONTRATAÇÃO OK
SEI-XX0002/0xxx63/2024	INEXECUÇÃO TOTAL OU PARCIAL
SEI-XX0007/0xxx74/2024	CONTRATAÇÃO OK
SEI-XX0004/0xxx0/2024	POSSIBILIDADE DE FALHA OU FRAUDE
SEI-XX0007/0xxx9/2024	INEXECUÇÃO TOTAL OU PARCIAL
SEI-XX0007/0xxx5/2024	INEXECUÇÃO TOTAL OU PARCIAL
SEI-XX0007/0xxx9/2024	POSSIBILIDADE DE FALHA OU FRAUDE
SEI-XX0005/0xxx8/2024	CONTRATAÇÃO OK
SEI-XX0007/0xxx7/2024	CONTRATAÇÃO OK
SEI-XX0001/0xxx0/2024	INEXECUÇÃO TOTAL OU PARCIAL

Fonte: autoria própria, 2025

5. Conclusão

Este estudo demonstrou, com base em evidências empíricas oriundas de dados reais de contratações públicas no Estado do Rio de Janeiro, que a integração de indicadores econômicos clássicos e modelos analíticos de ciência de dados constitui um instrumento robusto para o fortalecimento da governança pública. Os resultados revelaram níveis críticos de concentração de mercado em serviços essenciais, com valores elevados do Índice de Herfindahl-Hirschman e coeficiente de Gini extremos, indicando assimetrias severas na distribuição dos recursos públicos e estruturas contratuais fortemente dependentes de poucos fornecedores.

Em contraste, a identificação do Serviço de Operação Geral de UTI como um setor com baixa concentração (IHH = 0.490), baixa desigualdade na alocação de recursos (Gini = 0,21) e elevada pulverização de fornecedores (mais de 30 prestadores atuantes) evidencia que arranjos contratuais mais competitivos são factíveis mesmo em mercados complexos e de alta relevância estratégica. Esse contraste empírico reforça que a concentração observada em outros serviços não decorre de uma condição estrutural inevitável do setor público, mas de escolhas institucionais, regulatórias e organizacionais passíveis de revisão e aprimoramento.

A análise de *outliers* por meio da regressão linear e a classificação preditiva de risco com o algoritmo XGBoost evidenciaram elevado potencial de uso dessas ferramentas como mecanismos de fiscalização preventiva e de apoio à tomada de decisão estratégica, contribuindo para a priorização de processos com maior probabilidade de inconsistência. Ao integrar métricas de concentração de mercado, desigualdade distributiva e modelagem preditiva, a abordagem proposta amplia a capacidade analítica dos órgãos de controle e gestores públicos,

promovendo um modelo de governança orientado por dados, mais eficiente, transparente e alinhado aos princípios da Lei nº 14.133/2021.

Em termos de implicações práticas, os resultados sugerem que políticas de contratação pública que ampliem a concorrência, reduzam barreiras de entrada e promovam maior pulverização de fornecedores tendem a produzir ambientes contratuais mais consistentes, com menor assimetria na alocação de recursos e menor exposição a riscos estruturais. Assim, o uso sistemático de inteligência de dados no monitoramento das contratações públicas emerge como um caminho promissor para o aprimoramento institucional da governança e do controle no setor público.

REFERÊNCIAS

- BRASIL. Lei nº 14.133, de 1º de abril de 2021. *Institui normas de licitações e contratos administrativos para a Administração Pública*. Diário Oficial da União: seção 1, Brasília, DF, 1º abr. 2021. Disponível em: <https://tinyurl.com/3349z4h7>. Acesso em: 30 dez. 2025.
- CHAPMAN, P. et al. *CRISP-DM 1.0: Step-by-step data mining guide*. The CRISP-DM Consortium, 2000. Disponível em: <https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf>. Acesso em: 5 jul. 2025.
- CHEN, T.; GUESTRIN, C. XGBoost: um sistema escalável de reforço de árvores. In: KDD '16. New York: ACM, 2016. p. 785-794. Disponível em: <https://doi.org/10.1145/2939672.2939785>. Acesso em: 18 jul. 2025.
- CHEIN, F. *Introdução aos modelos de regressão linear*. 2019.
- COCCHIARELLI, R. S.; PESSANHA, J. F. M. A Lei de Newcomb-Benford como ferramenta de controle interno nas contratações públicas do Estado do Rio de Janeiro. *Revista Delos*, v. 17, n. 62, p. e3272, 2024. Disponível em: <https://ojs.revistadelos.com/ojs/index.php/delos/article/view/3272>. Acesso em: 18 jun. 2025.
- COWELL, Frank. *Medindo a desigualdade*. Oxford University Press, 2011.
- CUNHA, H. S. da; FERREIRA, C. V.; MURARI, T. B. Governança nas contratações públicas: referencial teórico. *Revista de Gestão e Secretariado*, v. 15, n. 7, p. e4029, 2024. Disponível em: <https://ojs.revistagesec.org.br/secretariado/article/view/4029>. Acesso em: 21 jun. 2025.
- ESTADOS UNIDOS. DEPARTAMENTO DE JUSTIÇA. *Diretrizes para fusões horizontais*. Departamento de Justiça dos EUA, 1997.
- GINI, C. Concentration and dependency ratios. *Journal of the Royal Statistical Society*, v. 75, n. 3, p. 326-339, 1912.
- GOÍIS, David Xavier de Souza; CARVALHO, Zulmara Virgínia de. Inteligência de dados como catalisador do relacionamento e da experiência do cliente. *Revista Observatorio de la Economía Latinoamericana*, Curitiba, v. 23, n. 1, p. 1–19, 2025. <https://doi.org/10.55905/oelv23n1-224>
- GOVERNO DO ESTADO DO RIO DE JANEIRO. Portal de Compras Públicas – Extração de Dados. Disponível em: <https://www.compras.rj.gov.br/Principal/extracaoTotal.action>. Acesso em: 9 maio 2025.
- HARARI, Y. N. *Nexus: uma breve história das redes de informação*. Rio de Janeiro: Alta Books, 2024.
- HERFINDAHL, O. C. *Concentration in the steel industry*. 1950. Tese (Doutorado em Economia) – Columbia University, New York, 1950.
- HASTIE, T., TIBSHIRANI, R., & FRIEDMAN, J. (2009). *The Elements of Statistical Learning*. Springer.

MIOTTO, M. V. J. et al. Governança e controle das contratações públicas. *Conpedi Law Review*, v. 10, n. 1, 2024. Disponível em: <https://www.indexlaw.org/index.php/conpedireview/article/view/10705>. Acesso em: 21 jun. 2025.

OLIVEIRA, F. M. et al. Participação do cidadão na nova governança pública. *Revista Foco*, v. 17, n. 1, p. e4300, 2024. Disponível em: <https://ojs.focopublicacoes.com.br/foco/article/view/4300>. Acesso em: 21 jun. 2025.

PORTER, M. (2008). *The Five Competitive Forces That Shape Strategy*. Harvard Business Review.

SANTOS, M. C. P. dos; NEVES, W. R. das; PIFFER, D. M. Estratégias de incentivo à inovação tecnológica no setor público: uma abordagem sobre a gestão participativa. *Akrópolis: Revista de Ciências Humanas*, v. 31, n. 2, p. 194–219, 2023. Disponível em: <https://revistas.unipar.br/index.php/akropolis/article/view/10767>. Acesso em: 19 jul. 2025.

WOOLDRIDGE, JM (2016). *Econometria introdutória: uma abordagem moderna*. South-Western Cengage Learning.