



Administração Pública e Gestão Social  
ISSN: 2175-5787  
apgs@ufv.br  
Universidade Federal de Viçosa  
Brasil

# Análise Comparativa da Eficiência Alocativa das Despesas Públicas Municipais utilizando Mineração de Dados

**Guelman Davis, Paula; de Souza, Antônio Artur**

Análise Comparativa da Eficiência Alocativa das Despesas Públicas Municipais utilizando Mineração de Dados

Administração Pública e Gestão Social, vol. 13, núm. 4, 2021

Universidade Federal de Viçosa, Brasil

Disponível em: <https://www.redalyc.org/articulo.oa?id=351568433009>



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivar 3.0 Internacional.

# Análise Comparativa da Eficiência Alocativa das Despesas Públicas Municipais utilizando Mineração de Dados

Comparative Analysis of the Allocative Efficiency of Municipal Public Expenditures using Data Mining

Análisis comparativo de la eficiencia asignada del gasto público municipal mediante minería de datos

Paula Guelman Davis

Fundação Hospitalar do Estado de Minas Gerais  
(FHEMIG), Brasil  
paulagdavis94@gmail.com

Redalyc: <https://www.redalyc.org/articulo.oa?id=351568433009>

Antônio Artur de Souza

Universidade Federal de Minas Gerais (UFMG), Brasil  
antonioarturdesouza@gmail.com

Recepción: 08 Octubre 2020

Aprobación: 12 Abril 2021

Publicación: 01 Octubre 2021

## RESUMO:

**Objetivo da pesquisa:** Analisar comparativamente a eficiência alocativa das despesas públicas dos municípios brasileiros utilizando técnicas de mineração de dados. Entende-se como eficiente a melhor relação entre o volume de despesas e resultados nos indicadores socioeconômicos, ou seja, menores despesas e melhores resultados quando comparado aos outros municípios.

**Enquadramento teórico:** Teorias da Federação, especificamente o Federalismo Fiscal, na qual uma das principais proposições é a de que tributos e despesas públicas descentralizadas são mais eficientes.

**Metodologia:** Coleta e a integração de dados financeiros e operacionais dos anos mais atuais das fontes utilizadas (2010, 2011, 2015, 2016 e 2017). Transformação em indicadores. Limpeza e seleção de indicadores para transformação em atributos categóricos (discretização).

**Resultados:** Viabilidade e importância do uso da mineração de dados sobre conjuntos abrangentes de dados. Geração de árvores de decisão a partir de algoritmos de mineração de dados, permitindo identificar características em comum dos municípios com desempenho destacado. Foram identificados padrões relevantes nas áreas de educação, saúde e desenvolvimento, como a importância de se despendar recursos na formação dos docentes, e caracterizadas discrepâncias regionais em termos do desempenho do Sistema Único de Saúde (SUS).

**Originalidade:** Nos estudos anteriores, foi possível constatar uma repetição da metodologia *Data Envelopment Analysis* (DEA) para estudo da eficiência das despesas públicas. Em contraste com esses estudos, utilizamos grandes conjuntos de dados municipais de múltiplas fontes e áreas, aplicando técnicas de mineração de dados para encontrar padrões relevantes.

**Contribuições teóricas e práticas:** Proposição de uma metodologia de estudo da eficiência da despesa pública municipal, válida do ponto de vista teórico e prático, em múltiplas áreas de atuação, com as técnicas e ferramentas da Mineração de Dados.

**PALAVRAS-CHAVE:** Finanças Públicas, Planejamento Governamental, Eficiência das despesas públicas, Mineração de Dados.

## ABSTRACT:

**Research purpose:** Analyze the allocative efficiency of public expenditures of Brazilian municipalities using data mining techniques. The municipalities considered efficient were those that had the best relation between their expenses and their results in the socioeconomic indicators, that is lower expenses and better results when compared to other cities.

**Theoretical framework:** Theories of the Federation, specifically Fiscal Federalism, in which one of the main propositions is that decentralized public taxes and expenses are more efficient.

**Methodology:** Collection and integration of financial and operational data of the years 2010, 2011, 2015, 2016 and 2017. Transformation into indicators. Cleaning and selection of indicators for transformation into categorical attributes (discretization).

**Results:** Feasibility and importance of using data mining over comprehensive data sets. Generation of decision trees, based on data mining algorithms, allowing the identification of common characteristics of municipalities with outstanding performance. Relevant patterns were identified in the areas of education, health and development, such as the importance of spending resources

on teacher training and characterized regional discrepancies in terms of the performance of the Brazilian Unified Health System (SUS).

**Originality:** The study fills the gap in the use of Data Mining techniques in national works, aimed at comparing cities and respecting the reality in Brasil. In previous studies, it was possible to observe a repetition of the Data Envelopment Analysis (DEA) methodology for studying the efficiency of public expenditures.

**Theoretical and practical contributions:** Proposition of a methodology for studying the efficiency of municipal public expenditure, valid from a theoretical and practical point of view, in multiple areas of activity, with the techniques and tools of Data Mining.

**KEYWORDS:** Public finances, Government Planning, Efficiency of public expenditure, Data Mining.

## RESUMEN:

**Objetivo de la investigación:** Analizar comparativamente la eficiencia asignativa del gasto público en los municipios brasileños utilizando técnicas de minería de datos. Se entiende como eficiente la mejor relación entre el volumen de gastos y resultados en los indicadores socioeconómicos, es decir, menores gastos y mejores resultados en comparación con otros municipios.

**Marco teórico:** Teorías de la federación, específicamente el federalismo fiscal, en las que una de las principales proposiciones es que los impuestos y gastos públicos descentralizados son más eficientes.

**Metodología:** Recolección e integración de datos financieros y operativos de los años más actuales de las fuentes utilizadas (2010, 2011, 2015, 2016 y 2017). Transformación en indicadores. Limpieza y selección de indicadores para su transformación en atributos categóricos (discretización).

**Resultados:** viabilidad e importancia de utilizar la minería de datos en conjuntos de datos completos. Generación de árboles de decisión, basados en algoritmos de minería de datos, que permitan la identificación de características comunes de municipios con desempeño sobresaliente. Se identificaron patrones relevantes en las áreas de educación, salud y desarrollo, como la importancia del gasto de recursos en la formación docente, y se caracterizaron las discrepancias regionales en cuanto al desempeño del Sistema Único de Salud Brasileño (SUS).

**Originalidad:** En estudios previos se pudo observar una repetición de la metodología del Análisis Envoltante de Datos (DEA) para estudiar la eficiencia del gasto público. En contraste con estos estudios, usamos grandes conjuntos de datos municipales de múltiples fuentes y áreas, aplicando técnicas de minería de datos para encontrar patrones relevantes.

**Aportes teóricos y prácticos:** Proposición de una metodología para el estudio de la eficiencia del gasto público municipal, válida desde un punto de vista teórico y práctico, en múltiples áreas de actividad, con las técnicas y herramientas de Data Mining.

**PALABRAS CLAVE:** Finanzas Públicas, Planificación gubernamental, Eficiencia del gasto público, Procesamiento de datos.

## 1. INTRODUÇÃO

O Estado moderno, desde a sua formação, não parou de crescer. Nas fases iniciais, desfrutava de um poder absoluto, mas com estrutura material, institucional e financeira pouco expressiva. Nos períodos seguintes, ampliou seu domínio, exercendo maior controle sobre a sociedade civil em todos os âmbitos da vida econômica e social, de forma a ser legitimado como instrumento de organização e realização da humanidade (Oliveira, 2012).

Independentemente das discussões sobre o tamanho do aparato estatal, é o Estado o responsável por atender necessidades que os indivíduos não são capazes de satisfazer sozinhos. Isso ocorre por meio das despesas públicas, foco de análise deste trabalho, que correspondem aos dispêndios realizados pelo Estado ou outras pessoas de direito público para garantir o funcionamento dos serviços públicos.

No orçamento, as despesas abarcam as autorizações de gastos com as funções e atribuições do governo. Segundo Giacomoni (2017), essas podem ser intituladas **funções fiscais**, considerando que esse é o principal instrumento da ação estatal na economia. Essas funções são a **alocativa**, que diz respeito ao fornecimento de bens e serviços públicos; a **distributiva**, associada a distribuição de renda de forma mais equitativa; e a **estabilizadora**, ligada à necessidade de uso da política econômica visando a elevado nível de emprego, estabilidade dos preços e crescimento econômico. Destas, destaca-se a **alocativa**, que se apoia no reconhecimento da inabilidade do mercado de prover bens e serviços para consumo coletivo.

Para cumprir essas funções, o Estado deve apresentar contas públicas equilibradas. No entanto, atualmente, ocorre o movimento contrário, ou seja, as despesas públicas vêm aumentando, fazendo com que os entes frequentemente incorram em *déficits* orçamentários. Esta realidade fez com que, nos últimos anos, a eficiência no gasto público e a relação desta com o desenvolvimento socioeconômico se tornassem objetos de estudos.

Em 2016, o Programa das Nações Unidas para o Desenvolvimento (PNUD) publicou o Relatório de Desenvolvimento Humano, que considera o desenvolvimento como um conjunto de dimensões cujo objetivo é aumentar os níveis de bem-estar social e qualidade de vida da população (Costa, Ferreira, Braga, & Abrantes, 2015). Nesse sentido, cabe ao governo criar um espaço fiscal para mobilizar recursos para instituir reformas capazes de melhorar a governança, as instituições e a economia. A construção desse espaço possui quatro pilares: assistência ao desenvolvimento, receitas públicas, financiamento do *déficit* público e eficiência das despesas públicas nas áreas consideradas prioritárias pelos chefes de Estado.

A eficiência das despesas públicas pode ser interpretada como a melhor relação custo-benefício para alcançar os objetivos de determinada política, ação ou diretriz governamental a partir das demandas dos cidadãos (Costa & Castanhar, 2003). Quanto mais eficiente for o uso do recurso, maior será a sua capacidade de gerar recursos adicionais. No Brasil, os entes públicos que possuem maior contato com essas demandas são as prefeituras.

Considerando esse contexto, o objetivo deste trabalho é analisar a eficiência na alocação de recursos pelos municípios, a partir de dados financeiros e operacionais. A justificativa se dá a partir da Constituição da República que, em seu artigo 37, estabelece que a Administração Pública de qualquer um dos Poderes da União, estados, Distrito Federal e municípios deve seguir os princípios de legalidade, impessoalidade, moralidade, publicidade e eficiência.

O princípio da eficiência, segundo Di Pietro (2017), inclui dois aspectos: o modo de atuação dos agentes públicos e as práticas de organização, estruturação e disciplina da Administração Pública. Para ambos, espera-se o melhor desempenho para que sejam alcançados os melhores resultados possíveis na prestação de serviços à população. Isso inclui as decisões de direcionamento dos recursos governamentais, que devem visar a eficiência do dispêndio público, além do atendimento às necessidades da população. Essa eficiência deve ser mensurada principalmente em um contexto de orçamentos comprometidos com pagamento de juros e encargos da dívida e de despesas com pessoal.

Sendo assim, a presente pesquisa orienta-se em torno do seguinte problema: Quais são as similaridades na composição das despesas públicas dos municípios com desempenho destacado em termos dos indicadores socioeconômicos?

Para responder esse problema, utilizou-se a mineração de dados, ou seja, um conjunto de métodos e técnicas voltadas para a descoberta de conhecimento a partir de dados. Foram coletados dados disponíveis em meio digital, tanto nos portais de transparência dos entes públicos quanto nos portais de organizações governamentais e não governamentais dedicadas a organizar e prover dados aos cidadãos conforme a disponibilidade dos dados para os municípios nos anos mais atuais de cada fonte (2010, 2011, 2015, 2016 e 2017). A partir do tratamento desses dados, foram identificados, por meio de árvores de decisão, padrões que contribuíram para a formação do conhecimento sobre a eficiência das despesas públicas municipais.

O trabalho mostra a importância da metodologia da mineração de dados, pois além de permitir a realização da pesquisa com o maior número possível de municípios, possibilitou a identificação de municípios com experiências e resultados importantes, que não configuram na lista dos mais populosos ou mais ricos do país. Esses resultados contrariam a hipótese de que quanto maiores as despesas em determinada área, melhores serão os resultados. Além disso, foi possível expandir análises de trabalhos anteriores com uma metodologia diferenciada, uma vez que, até onde foi possível verificar, muitos utilizaram metodologias semelhantes, como a *Data Envelopment Analysis* (DEA).

Este artigo é composto por sete seções, incluindo esta introdução. A seção 2 traz um estudo sobre o federalismo fiscal e as teorias da federação. A seção 3 faz uma revisão da literatura com os principais trabalhos

que analisaram as despesas públicas sobre a ótica dos indicadores socioeconômicos. A seção 4 mostra a metodologia adotada. A seção 5 apresenta a caracterização dos dados, ou seja, as etapas que os dados passaram até a aplicação de logaritmos de mineração de dados. A seção 6 tem como objetivo a apresentação, análise e interpretação dos resultados. Por fim, a seção 7 apresenta as considerações finais e sugestões de trabalhos futuros.

## 2. TEORIAS DA FEDERAÇÃO: O FEDERALISMO FISCAL

A execução das funções alocativa, distributiva e estabilizadora do Estado pode ser realizada por meio de três formas de organização política: a total centralização, na qual todas as funções são desempenhadas pelo governo central; a absoluta descentralização, modelo em que os governos locais seriam autônomos e totalmente responsáveis por essas funções; e a forma federativa, que combina as outras duas, na qual as responsabilidades são divididas entre os entes governamentais, que é o caso do Estado brasileiro. Segundo Silva (2005), o sistema federativo requer uma análise dos graus de descentralização existentes, de forma a verificar o desenho organizacional mais adequado para o cumprimento das funções do Estado.

Segundo Oliveira (2007), quando o federalismo foi incorporado como tema relevante dentro dos estudos da Economia do Setor Público, passou-se a dar ênfase no campo fiscal, isto é, na repartição de receitas, encargos e competências entre os entes federados. Desde então, procura-se analisar os arranjos institucionais que promovem maior eficiência na alocação dos recursos públicos, considerando que todos os entes governamentais possuem essa mesma função. Por isso, é essencial que exista uma organização constitucional de competências fiscais entre os diferentes níveis de governo para otimizar o processo de descentralização. Cada ente autônomo, observando suas competências e capacidades de financiamento legalmente estabelecidas, consegue controlar seus procedimentos próprios de contribuição e gestão tributária, transferências fiscais, composição e controle das despesas públicas (Silva, 2005).

O controle de despesas públicas associado ao atendimento das demandas da população em nível local foi objeto de vários trabalhos. Wagner (1890), por meio da Lei dos Dispendios Públicos Crescentes (ou hipótese de Wagner), mostra que o desenvolvimento econômico, de maneira geral, demanda uma crescente participação do governo na oferta de bens públicos, principalmente em casos de mudanças demográficas que podem exigir maiores gastos em saúde, educação, políticas de redistribuição de renda, entre outros (Candido, 2001).

O estudo de Tiebout (1956) mostra que um governo municipal consegue responder de forma mais eficiente e rápida às demandas da população por bens públicos. De forma complementar, Oates (1972) justifica essa afirmativa pelo fato de os governos locais, por estarem mais próximos dos cidadãos, conseguirem entender melhor as necessidades da população e aumentar o bem-estar econômico da localidade. O Teorema da Descentralização de Oates, analisado por Silva (2005), afirma que a perfeita correspondência entre os atos governamentais e as demandas da população significa que, naquele local, a eficiência econômica da ordenação federativa fiscal alocativa está maximizada.

Além disso, Musgrave (1983) constata que despesas públicas descentralizadas geram concorrência entre os governos locais já que a população pode optar, via voto, por governantes que melhor atendam suas necessidades. Portanto, a descentralização fiscal contribui para a eficiência do setor público no que diz respeito ao fornecimento de bens e serviços.

No entanto, é importante ressaltar a “hipótese do Leviatã”, formulada por Brennan e Buchanan (2000). Segundo ela, a falta de uma fiscalização adequada dos dispendios públicos, somada a uma ação parasitária dos grupos de interesse, pode fazer com o que o governo aumente o controle sobre a população utilizando o orçamento público. Isso pode provocar uma elevação dos gastos sem levar em consideração as necessidades e demandas dos cidadãos (Salomão, 2020).



A descentralização fiscal ligada ao federalismo ganhou força no debate político e acadêmico, uma vez que foi considerada como um aspecto importante das relações intergovernamentais, influenciando não só na mobilização e alocação de recursos, mas também na estabilidade da economia e na formulação de políticas públicas de distribuição de renda. É considerada uma área complexa, pois pode tomar diversas formas e dimensões (Litvack, Ahmad, & Bird, 1998).

Nos países em que a política fiscal é realizada de maneira descentralizada, é essencial que a tomada de decisões dos entes federados seja feita em prol da maximização de eficiência na utilização dos recursos. Além disso, Silva (2005) enfatiza a necessidade de o governo federal coordenar as ações fiscais das unidades federadas de forma a minimizar externalidades fiscais como, por exemplo, a geração de *déficits* e o aumento de dívidas governamentais resultantes da ação de um ente sobre o outro. Segundo Rezende, Carneiro e Rezende (2014), no caso do Brasil verificam-se, em vários momentos, desequilíbrios de poder entre os entes subnacionais e a União, com prevalência da esfera central.

Em um sistema de federalismo fiscal ocorrem problemas de desequilíbrios verticais e horizontais no que diz respeito às diferenças entre receitas e despesas dos diferentes níveis de governo. O desequilíbrio vertical ocorre entre os níveis de governo, ou seja, um desbalanceamento de recursos entre municípios, estados e União. Já o desequilíbrio horizontal ocorre entre entes do mesmo nível. As disparidades regionais são um exemplo desse tipo de desequilíbrio, ou seja, enquanto alguns municípios e estados conseguem manter certa sustentabilidade fiscal, outros não conseguem cumprir seus compromissos por falta de recursos. A correção desses desequilíbrios deve ocorrer por meio de transferências intergovernamentais, cujos recursos são repassados da União para estados e municípios e dos estados para os municípios. O objetivo é reduzir problemas relacionados a externalidades e tentar distribuir de forma equitativa os recursos entre regiões com níveis diferentes de renda e de desenvolvimento (Shah, 1991).

### 3. EFICIÊNCIA DAS DESPESAS PÚBLICAS SOB A ÓTICA DOS INDICADORES SOCIOECONÔMICOS

A eficiência das despesas públicas é bastante discutida na literatura internacional e nacional. Esta noção envolve a alocação de recursos dos entes públicos visando a geração de bem-estar para a população. Esta análise é feita tanto no nível macro, que inclui dados de despesa por função, como micro, voltada para áreas específicas dentro de cada função. Busca-se verificar a relação entre o dispêndio de gastos públicos e os produtos e serviços ofertados pelo Estado.

Costa, Ferreira, Braga e Abrantes (2015) propuseram uma visão do processo produtivo no setor público. Segundo eles, a eficiência é considerada a melhor relação possível entre os recursos públicos aplicados e os produtos e serviços entregues, medidos pelos indicadores socioeconômicos. A efetividade, por sua vez, avalia a melhora ou piora dos resultados ao longo do tempo, visando verificar se houve desenvolvimento socioeconômico.

Nos trabalhos nacionais, a alocação dos gastos públicos já vinha sendo discutida desde meados dos anos 2000. Os estudos, de maneira geral, relacionam as despesas governamentais com seus respectivos indicadores socioeconômicos. Estes variam em relação ao objeto de estudo, estados ou municípios, amostra ou ente específico, indicadores utilizados, etc. A metodologia utilizada é, de maneira recorrente, a *Data Envelopment Analysis* (DEA), com poucas exceções. Em relação aos entes municipais, foram identificados trabalhos nas áreas de educação, saúde, indicadores fiscais e desenvolvimento econômico.

Na área de educação, a maior parte dos trabalhos considera, como insumo (input), os recursos aplicados; e como produtos, (output) os resultados na educação. Os estudos são feitos para o Ensino Fundamental (EF), que são de competência e responsabilidade municipal, e usando como indicador o Índice de Desenvolvimento da Educação Básica (IDEB).

Educação

Machado, Irffi e Benegas (2011) utilizaram essa metodologia para analisar as despesas públicas dos municípios cearenses em educação, saúde e assistência social. Concluiu-se que 55% da amostra dos municípios do Ceará poderia ser utilizada como *benchmark* pelos outros que ficaram aquém da fronteira de eficiência técnica estimada. Savian e Bezerra (2013) empregaram DEA para criar um índice de eficiência para a aplicação das despesas públicas nos anos iniciais do EF (1. ao 5. ano) nos municípios do Paraná. Os municípios das mesorregiões Norte Central, Norte Pioneiro e Noroeste foram os que concentraram os maiores escores de eficiência. O estudo mais recente, de Lourenço, Angotti, Nascimento e Sauerbronn (2017), analisou a eficiência dos gastos em educação dos 250 maiores municípios brasileiros por meio do número de matrículas no EF. A partir do cálculo de escores de eficiência técnica, o estudo constatou que apenas 5,2% da amostra teve desempenho considerado eficiente em municípios concentrados nas regiões Norte e Nordeste.

Saúde

Na área da Saúde, Politelo, Rigo e Hein (2014) utilizam a DEA para avaliar a eficiência das despesas das cidades de Santa Catarina levando em consideração a qualidade do atendimento do Sistema Único de Saúde (SUS) a partir de dados e indicadores do DATASUS, o departamento de informática do sistema. As conclusões da pesquisa mostraram que a eficiência média das cidades analisadas foi de 86,26%, considerada satisfatória pelos autores.

Almeida, Vasconcelos, Miranda e Feitosa (2017) utilizaram o Índice de Desempenho do Sistema Único de Saúde (IDSUS) na DEA para analisar as despesas públicas em saúde. Como amostra, consideraram 128 municípios pernambucanos cujos dados de despesa em saúde estavam disponíveis nas Finanças do Brasil (FINBRA) e que foram classificados no Grupo Homogêneo (GH) 6 (com piores condições de saúde) pelo Ministério da Saúde em 2010 e 2011. Foram construídos três modelos de regressão linear: Efeitos Aleatórios, Efeitos Fixos e Mínimos Quadrados Ordinários Agrupados. Os autores consideraram o modelo de efeitos fixos o mais adequado para análise das variáveis e suas consequências no IDSUS. Os resultados mostraram que apenas os gastos com prevenção influenciaram, de maneira significativa, o índice analisado.

Portulhak, Raffaeli e Scarpin (2018), por sua vez, utilizaram a DEA como forma de avaliar a eficiência dos municípios na destinação de recursos públicos a serviços de saúde prestados pelo SUS a partir do IDSUS. Busca-se, assim como em outros estudos, descobrir *benchmarks* para outros municípios do ponto de vista da eficiência de alocação das despesas em saúde. Os autores utilizaram uma amostra com 4.598 municípios. A amostra foi agrupada conforme os GHs definidos pelo IDSUS. Os resultados mostraram que 281 municípios puderam ser considerados como *benchmarks*.

Fiscal

Cruz e Afonso (2018) investigaram a relação entre o cumprimento das metas e limites fiscais, trazidos com a promulgação da Lei de Responsabilidade Fiscal (LRF), e as características da gestão fiscal. Para isso foi escolhida uma amostra com 282 municípios considerados de grande porte. O período de análise foi 2010 a 2013 devido à publicação da Lei Complementar n. 131/2009 (Lei da Transparência). O estudo focou na folga fiscal dos municípios em relação aos limites máximos da LRF e das Resoluções 40 e 43 de 2001 do Senado Federal. Foi elaborado um modelo econométrico para verificar a relação entre um indicador de cumprimento de metas e limites fiscais e as variáveis que representam outros pilares da gestão fiscal: a ação planejada, a transparência e o controle. Por fim, os resultados confirmaram a hipótese dos autores de que existe uma falta de integração entre os pilares da gestão fiscal e o cumprimento das metas.

Em relação a política fiscal dos municípios, destaca-se também o estudo de Salomão e Santos (2020), que, por meio de estimações de dados em painel a partir dos Modelos Logit/Probit, constataram que quanto maior a participação dos municípios nas transferências correntes, maior a probabilidade de elevação das despesas desses com pessoal. Isso faz com que o município deixe de cumprir os limites de alerta da LRF e comprometa o atendimento às demandas dos cidadãos.

Em relação ao desenvolvimento econômico, Cândido (2001) analisa a relação entre gastos públicos e crescimento econômico no Brasil entre 1947 e 1995. O autor utilizou duas metodologias: a primeira que estimou o efeito externalidade dos gastos e a diferença de produtividade em relação ao setor privado e a

segunda que buscou os efeitos da relação gasto público versus produto. O trabalho encontrou indícios de baixa produtividade do gasto público.

Além desses trabalhos, vale ressaltar o estudo de Matos e Ponczek (2013), que analisou os efeitos da divisão municipal sobre o fornecimento de bens públicos e indicadores sociais. Foram utilizados dados de coleta de lixo, rede de esgoto, iluminação, renda per capita, Índice de Desenvolvimento Humano (IDH), escolaridade média, taxa de analfabetismo e índice de Gini de desigualdade de renda para analisar os efeitos nos municípios que se dividiram entre 1991 e 2000. Os resultados encontrados mostram que houve ganhos de escala no fornecimento de bens públicos entre os municípios. Segundo os autores, os municípios menores, mesmo com o maior recebimento per capita nas transferências intergovernamentais, possuem dificuldade para transformar esses recursos em bens públicos.

Em relação ao número de municípios, o estudo de Salomão e Siani (2019) teve como objetivo verificar a forma como uma maior descentralização política gera consequências no comportamento fiscal dos municípios brasileiros. Como *proxy*, utilizou-se a criação de municípios nos anos 1990 e 2000. Foi constatado que a descentralização influenciou o desempenho fiscal dos municípios de forma geral, no que diz respeito a receitas e despesas.

#### 4. METODOLOGIA

Considerando os objetivos desta pesquisa, a técnica utilizada é quantitativa, pois a formação do conhecimento é feita por meio do levantamento, coleta e análise de dados secundários, coletados de fontes públicas de informação. Além de quantitativa, a pesquisa é descritiva, uma vez que descreve como se manifesta um fenômeno (Sampieri, Collado, & Lucio, 2006).

Foram utilizados dois softwares, livres e gratuitos: o R, que fornece uma ampla variedade de técnicas estatísticas e gráficas (<https://www.r-project.org/>); e o Weka, uma coleção de algoritmos que contém ferramentas para preparação de dados, classificação, regressão, agrupamento, mineração de regras de associação e visualização (<https://www.cs.waikato.ac.nz/ml/weka/>).

→ A população desta pesquisa é formada pelos 5.570 municípios brasileiros. Considerando que as fontes de dados são heterogêneas, ao longo do trabalho foi necessário excluir alguns municípios com dados faltantes conforme o tipo de análise e o indicador de referência considerado para cada área.

A Mineração de Dados é um processo de extração de conhecimento de bases de dados extensas e complexas cujo objetivo principal é descobrir padrões frequentes e correlações contidas nos dados, com especial interesse em padrões inesperados. Ela envolve um conjunto de métodos e técnicas, derivados de áreas como bancos de dados, estatística, aprendizado de máquina, visualização e ciência da informação (Han & Kamber, 2006).

→ Para aplicar essas técnicas, foi feita uma coleta dos dados municipais disponíveis nas principais fontes de informação pública do governo, nos anos mais atuais de cada fonte (2010, 2011, 2015, 2016 e 2017). Em seguida, eles foram preparados para limpeza de informações inconsistentes. Os dados, então, passaram por nova seleção e transformação de modo a produzir a integração necessária, feita a partir dos códigos municipais do Instituto Brasileiro de Geografia e Estatística (IBGE). Em seguida, transformou-se os dados em indicadores para, depois, eliminar as redundâncias. Depois, foi feito um processo de discretização para, então, iniciar a mineração dos dados.



## 5. CARACTERIZAÇÃO DOS DADOS

### 5.1 Coleta

A coleta dos dados foi feita nas fontes de informação públicas do governo, conforme a disponibilidade dos dados para os municípios nos anos mais atuais de cada fonte (2010, 2011, 2015, 2016 e 2017). Como este trabalho identificou padrões, não foi necessário que todos os dados fossem coletados para um mesmo ano, não só devido a variabilidade das fontes, o que implica periodicidades diferentes de publicações, mas também porque os dados, independente do ano, informam a realidade municipal nas principais áreas de atuação do Estado. Foram coletados dados de oito áreas (Quadro 1).

Educação

→ Na área de educação considerou-se apenas o Ensino Fundamental (EF). Isso significa que foram utilizadas apenas as despesas públicas desta subfunção. Além disso, para os dados coletados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), foram considerados os anos iniciais e finais do EF, que são publicados anualmente, com exceção do IDEB, cuja divulgação é feita a cada dois anos.

Quadro 1– Dados coletados

Classificação	Fonte	Dados Coletados	Ano
Dados Gerais	Instituto Brasileiro de Geografia e Estatística (IBGE)	Unidade da Federação	2017
		Município	2017
		Estimativa de população	2017
		Área	2017
		Partido do prefeito	2013/2017
Despesas Públicas	Secretaria do Tesouro Nacional - Finanças Municipais (FINBRA)	Existência de Plano Diretor	2015
		Despesas Pagas por Função: Segurança Pública, Assistência Social, Saúde, Educação (Ensino Fundamental) e Saneamento	2017
Indicadores Fiscais	Secretaria do Tesouro Nacional - Finanças Municipais (FINBRA)	Despesa Consolidada Líquida (DCL)	2017
		Despesa Total com Pessoal (DTP)	
		Receita Corrente Líquida (RCL)	
Educação	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP)	IDEB	2017
		Número de matrículas	
		Número de docentes	
		Número de estabelecimentos	
		Número de turmas	
		Média de alunos por turma	
		Média de horas-aula diária	
		Taxa de distorção idade-série	
		Percentual de docentes com curso superior	
Saúde	Ministério da Saúde	IDSUS	2011
	DATASUS	Número de internações hospitalares pelo SUS	2017
		Média de permanência hospitalar	
		Número de ambulatórios de atenção básica	
		Número de óbitos por ocorrência	
		Óbitos por causas evitáveis - 0 a 4 anos e 5 a 74 anos	
Segurança Pública	Atlas da Violência	Doses aplicadas de vacinação	2017
		Quantidade de homicídios	
		Índice de Desenvolvimento Humano Municipal (IDHm)	
Desenvolvimento	Atlas do Desenvolvimento Humano	Índice Firjan de Desenvolvimento Municipal (IFDM)	2010
	Federação das Indústrias do Rio de Janeiro (FIRJAN)	Índice Firjan de Desenvolvimento Municipal (IFDM)	2016
Saneamento	Sistema Nacional de Informações em Saneamento (SNIS)	População total atendida com abastecimento de água	2017
		População total atendida com esgotamento saneamento	

Fonte: Elaborado pelos autores.

Uma vez coletados, esses dados foram integrados em uma base de dados única.

## 5.2 Integração

A integração dos dados foi feita por meio dos códigos municipais feitos pelo IBGE, presentes nas bases que foram coletadas. No entanto, mesmo com essa padronização, verificou-se algumas fontes que alteram esses

códigos. Exemplos incluem DATASUS e o SNIS, cujos códigos excluem alguns dígitos considerados pelo IBGE. Em seguida, os dados passaram por um processo de transformação.

### 5.3 Transformação

Como foi observado no Quadro 1, alguns dados coletados já vieram no formato de indicadores, enquanto outros não. Sendo assim, foi necessário realizar um processo de transformação, para que todos os dados assumissem a forma de indicadores, normalizando seus valores para eliminar o efeito da dimensão populacional do município. O Quadro 2 resume esses indicadores e mostra suas referências.

Quadro 2 – Indicadores considerados para análise.

Indicador	Referência
Despesa por aluno no EF	Savian e Bezerra (2013) Machado, Irffi e Benegas (2011)
IDEB	INEP /IDEB
Alunos por turma	INEP /Dicionário de Indicadores Educacionais
Horas-aula diária	
Distorção idade-série	
Docentes com curso superior	
Taxa de aprovação	
Taxa de reprovação	
Taxa de abandono	Machado, Irffi e Benegas (2011)
Despesa per capita em saúde	
IDSUS	Ministério da Saúde (2014) Almeida <i>et al.</i> (2017) Portulhak, Raffaeli e Scarpin (2018)
Internações por 100.000 habitantes	Adaptado de Politelo, Rigo e Hein (2014)
Média permanência hospitalar	Sugerido pelos autores
Ambulatórios por 100.000 habitantes	
Óbitos por causas evitáveis	
Doses aplicadas de vacinação	Adaptado do Atlas da Violência
Despesa per capita com Segurança	
Homicídios por 100.000 habitantes	LRP, artigo 2º, § 2º
Despesa Total com Pessoal (DTP)	Resolução n. 40/2001 do Senado Federal
Dívida Consolidada Líquida (DCL)	Atlas do Desenvolvimento Humano
IDHm	FIRJAN
IFDM	Machado, Irffi e Benegas (2011)
Despesa per capita com Saneamento	Sugerido pelos autores
População atendida com o abastecimento de água	
População atendida com esgotamento sanitário	Machado, Irffi e Benegas (2011)
Despesa per capita com Assistência Social	

Fonte: Elaborado pelos autores.

Uma vez definidos os indicadores, foi realizada uma verificação, para retirar informações redundantes que poderiam prejudicar a análise.

### 5.4 Limpeza

A etapa de limpeza inclui não só a verificação de dados faltantes, mas também a detecção de inconsistências, que eram problemas esperados considerando a heterogeneidade das fontes de dados. No caso de dados ausentes, a literatura recomenda a estimação do valor mais provável por meio de regressão ou inferências. Já no caso de valores com ruídos, estes podem ser eliminados por meio de interpolação, agrupamento, inspeção humana e computacional combinadas ou regressão (Han & Kamber, 2006). Alternativamente, dados ausentes ou inconsistentes podem ser retirados do banco. Neste trabalho, optou-se por retirar os dados faltantes ou inconsistentes, pois, considerando a heterogeneidade dos municípios brasileiros, não seria adequada a realização de regressões e inferências.

Para identificar os dados inconsistentes (*outliers*), utilizou-se o software R. Os indicadores foram avaliados em termos de dispersão por meio de *scatterplots* e histogramas. Assim, foi possível identificar valores discrepantes em relação ao restante da amostra e excluí-los. Em seguida, foram feitas análises de correlação, usando o coeficiente de correlação de Pearson, para verificar a relação entre os indicadores.

→ Para os pares de indicadores que apresentaram forte correlação, foi escolhido apenas um deles. No caso da educação, optou-se por manter o IDEB em detrimento das taxas de aprovação, reprovação e abandono, por ser mais utilizado nos trabalhos anteriores, voltado à educação básica e com uma meta nacional pré-estabelecida para os municípios. Para a área de desenvolvimento, a escolha do indicador se deu pelo número de municípios com dados disponíveis nas bases de dados. Optou-se por incluir na análise apenas o IDHm, que não está disponível apenas para 5 municípios. Em relação às outras áreas, não foram identificadas correlações fortes entre os indicadores que justificassem a retirada de qualquer um deles.

## 5.5 Discretização

Segundo Zaki e Meira (2014), a discretização (*binning*), tem como objetivo converter atributos numéricos em categóricos. Trata-se de um processo que ajuda na redução do número de valores distintos para um atributo, principalmente quando há ruídos nos dados, pois permite ignorar diferenças pequenas e irrelevantes nos valores. Para este trabalho, os dados foram categorizados considerando os desvios-padrão dos dados municipais na maior parte das áreas estudadas.

Cada indicador foi analisado conforme suas estatísticas descritivas. Foram feitos testes com os algoritmos disponíveis no Weka para verificar a melhor maneira de discretizá-los. Padronizaram-se algumas formas de discretização para indicadores similares (Quadro 3).

Quadro 3 – Discretizações padronizadas

Tipo de indicador	Forma de discretização
Despesas por função	Os municípios foram ordenados de maneira crescente e divididos em três faixas.
Taxas por 100.000 habitantes	Considerou-se a média de todos os municípios com dados disponíveis (média nacional); Tirou-se a média dos municípios abaixo da média e a média dos municípios acima da média.
Indicadores percentuais	Considerou-se a média dos municípios menos o desvio padrão como uma faixa. Em seguida, tirou-se uma nova média para formar outras duas faixas.

Fonte: Elaborado pelos autores.

Além desses, cabe ressaltar os indicadores já coletados com alguma forma de discretização já definida, como foi o caso do IDHm e do IDSUS, cujas fontes estabelecem a classificação dos municípios, e da mudança de partido do prefeito e existência de Plano Diretor, com classificações binárias (SIM/NÃO). Para a DTP e a DCL, também foi adotado esse tipo de classificação, considerando a superação ou não dos limites legais (ACIMA/ABAIXO).

Os outros indicadores foram discretizados como se segue. As faixas do IDEB foram formuladas a partir da meta nacional de 6,0 estabelecida pelo Inep para 2022. Para o número de alunos por turma, os municípios foram ordenados e divididos, assim como foi feito para as despesas por função. O indicador de horas-aula diária foi dividido a partir da padronização do EF: um turno somente ou integral. Por fim, para a média de

permanência hospitalar, utilizou-se a mesma classificação das taxas por 100.000 habitantes, já que para esse indicador, foi constatado que seria melhor a divisão dos municípios em quatro grupos homogêneos.

A Tabela 1 mostra os valores discretizados para as despesas públicas, objetos deste trabalho. É importante ressaltar que se optou pela utilização do orçamento por função, conforme previsto na Lei 4.320, de 17 de março de 1964.

No caso da educação, para que a discretização ficasse fidedigna com a realidade dos municípios brasileiros, considerou-se o valor mínimo anual por aluno de R\$ 2.875,03, conforme Portaria Interministerial nº 08 de 26 de dezembro de 2016, artigo 2º. O valor máximo, por sua vez, foi obtido pela média da amostra somada ao seu desvio padrão – R\$ 12.721,00. Os valores não incluídos nessa faixa foram considerados *outliers* e excluídos da análise. Após extrair os valores máximo e mínimo, dividiram-se os municípios em três grupos.

Tabela 1 – Grupos de municípios por faixas de valores de despesas públicas

Despesa	Faixas de valores	Número de municípios
Educação (EF)	Até R\$ 6.000,00*	1.749
	R\$ 6.000,00 –  R\$ 9.000,00	2.071
	R\$ 9.000,00 e acima	950
Saúde	Até R\$ 500,00	1.674
	R\$ 500,00 –  R\$ 800,00	2.306
	R\$ 800,00 e acima	1.499
Saneamento	Até R\$ 15,00	1.297
	R\$ 15,00 –  R\$ 65,00	781
	R\$ 65,00 e acima	1.016
Segurança Pública	Até R\$ 3,00	799
	R\$ 3,00 –  R\$ 12,00	749
	R\$ 12,00 e acima	723
Assistência Social	Até R\$ 70,00	1.916
	R\$ 70,00 –  R\$ 120,00	1.817
	R\$ 120,00 e acima	1.755

Fonte: Elaborada pelos autores

\* Considerou-se como valor mínimo R 287500 conforme Portaria Interministerial nº 08 de 26 de dezembro de 2016 artigo 2º

Em relação aos indicadores com metodologia já definida – IDSUS e IDHm – cabem algumas observações. O IDSUS adota uma metodologia que divide os municípios em seis grupos a partir do Índice de Desenvolvimento Socioeconômico (IDSE), Índice de Condições de Saúde (ICS) e Índice de Estrutura do Sistema de Saúde do Município (IESSM) – média e alta complexidade (Ministério da Saúde, 2014).

Para reduzir a possibilidade de problemas de desbalanceamento no processamento das classes, optou-se por unir os grupos do IDSUS. Isso foi feito a partir do IDSE e do ICS. O IESSM foi excluído, considerando a participação dos estados no atendimento à média e alta complexidade hospitalar. A Tabela 2 mostra o resultado dessa junção. O grupo com os melhores resultados em termos do desempenho do SUS é o IDSUS 3.



Tabela 2 – Grupos IDSUS adaptados

Grupo IDSUS	IDSE	ICS	Grupo IDSUS adaptado	Quantidade de Municípios
6 e 4	Baixo	Baixo	IDSUS 1	2.770
5 e 3	Médio	Médio	IDSUS 2	2.671
1 e 2	Alto	Médio	IDSUS 3	123

Fonte: Elaborado pelos autores.

Em relação ao IDHm, os valores variam entre 0 e 1. Os municípios são divididos em cinco grupos de resultados de IDHm: muito baixo, entre 0,000 e 0,499; baixo, entre 0,500 e 0,5999; médio, entre 0,600 e 0,699; alto, entre 0,700 e 0,799; e muito alto, entre 0,800 e 1,000. Quanto mais próximo de 1, maior o desenvolvimento humano. (PNUD, 2013).

Como os grupos de IDHm intitulados muito baixo e muito alto apresentavam apenas 32 e 44 municípios respectivamente, foi necessário, para evitar desbalanceamento, juntá-los com outros grupos. Assim, os municípios de IDHm muito baixo foram realocados para o grupo baixo; e os de IDHm muito alto para o grupo alto. A discretização adaptada do IDHm é detalhada no Tabela 3.

Tabela 3 – Grupos IDHm adaptados

Indicador	Grupos	Faixas de Valores	Número de municípios por faixa
<i>IDHm</i>	IDHm baixo	0,000 – 0,599	1.399
	IDHm médio	0,599 – 0,699	2.233
	IDHm alto	0,699 – 1,000	1.933

Fonte: Elaborado pelos autores.

Por fim, cabe ressaltar que foi acrescentada, como atributo, a região geográfica dos municípios, com cinco valores, com o objetivo de verificar se a região se relaciona com os resultados dos indicadores. Ao final, foi formado um banco de dados com 24 atributos (*features*) por município.

## 6. APRESENTAÇÃO, ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS

Para a identificação de padrões, tendências e relações entre os indicadores municipais formulados, foram utilizados dois algoritmos de mineração de dados: InfoGainAttributeEval e J48.

Primeiramente, aplicou-se o algoritmo InfoGainAttributeEval para identificar os atributos que melhor explicam as variações das classes do indicador de referência por meio do ganho de informação. Segundo Vural e Subasi (2015), o ganho de informação é uma propriedade estatística que mede o quanto um atributo consegue explicar as variações dos valores de outro atributo. Para tal, o valor da entropia é utilizado. Para o processamento desse algoritmo, os atributos são considerados como independentes e os valores do ganho variam de 0 (nenhuma informação) a 1 (nível máximo de informação). Além disso, selecionou-se o método de análise ranker, que classifica os atributos, em ordem crescente, conforme o ganho de informação. O modo de seleção dos atributos, por fim, se deu a partir de validação cruzada (*cross-validation*) considerando a divisão da base em 5 partes.

Uma vez verificados os atributos que melhor distinguem os valores do atributo de referência, esses foram separados para gerar árvores de decisão por meio do algoritmo J48. As árvores de decisão são estruturas de dados que registram o caminho, em ordem, dos atributos mais relevantes para explicar um padrão. No caso

deste trabalho, as folhas são os conjuntos de municípios e os caminhos da raiz até as folhas são os atributos que esses municípios têm em comum.

A geração dessas árvores segue a abordagem top-down e utiliza a medida de ganho de informação para escolher os atributos em cada ramificação da árvore (Han & Kamber, 2006). Os nós internos de uma árvore de decisão indicam os diferentes atributos. Os ramos entre eles fornecem os prováveis valores que esses atributos podem ter nas amostras observadas. O atributo que deve ser previsto é reconhecido como a variável dependente, uma vez que a sua classificação é escolhida a partir dos valores de todos os outros atributos. Para a raiz da árvore, o algoritmo seleciona o atributo que melhor discrimina os valores da variável dependente (Baitharu & Pani, 2016). O processo continua recursivamente para cada subgrupo de municípios até que uma medida de uniformidade no grupo seja alcançada, ou até que o subgrupo fique pequeno demais. A medida de uniformidade e o menor tamanho do subgrupo são parâmetros do algoritmo. Cabe ressaltar que, nesse algoritmo, também foi utilizada a validação cruzada, que divide a base de dados em 5 partes para a geração das árvores, sendo 4 usadas para gerar a árvore e uma para avaliar a qualidade do resultado.

## 6.1 Educação

Para a área de educação (EF), escolheu-se, como variável de referência, o IDEB, para o qual o governo federal estabeleceu uma meta nacional de nota 6,0 para 2022, com o objetivo de tornar a educação brasileira comparável à dos países desenvolvidos. Foram considerados 5.231 municípios com dados disponíveis.

A Tabela 4 mostra os dez indicadores com o maior ganho de informação em relação ao IDEB, separados para aplicação do algoritmo J48.

Tabela 4 – Ranking dos indicadores com maior ganho de informação em relação ao IDEB

<b>Indicador</b>	<b>InfoGain</b>
Região	0,407
IDHm	0,322
IDSUS	0,247
Percentual de docentes com curso superior	0,178
Despesa per capita em saúde	0,135
Despesa por aluno do EF	0,078
Despesa per capita em assistência social	0,048
Número de alunos por turma	0,044
DTP	0,037
Número de ambulatórios por 100.000 habitantes	0,028

Fonte: Elaborado pelos autores.

A Figura 1 mostra a árvore de decisão formada. O percentual de instâncias classificadas corretamente foi de 63,03%. Como parâmetros, utilizou-se a validação cruzada e o número mínimo de 100 objetos por folha. O objetivo foi identificar os caminhos da raiz até as folhas da árvore identificadas com o melhor resultado do IDEB. Na árvore, os números indicados entre parênteses correspondem, respectivamente, ao número total de municípios classificados e o número de municípios classificados incorretamente. A existência de um grande número de classificações incorretas enseja a subdivisão do grupo.

Os atributos que melhor explicam o IDEB em termos de ganho da informação estão mais próximos da raiz da árvore: IDSUS, região dos municípios e IDHm. Os municípios com melhores performances no IDEB,

ou seja, com nota igual ou maior do que a meta nacional, estão agrupados em folhas da árvore acessíveis por meio de cinco caminhos diferentes, com as seguintes características em comum:

- (1) 131 municípios com IDSUS de IDSE e ICS baixos (IDSUS 1), localizados na região Sudeste, com turmas de até 19 alunos e com mais de 87% dos docentes com formação superior;
- (2) 129 municípios com IDSUS de IDSE e ICS médios (IDSUS 2), IDHm médio, da região Sudeste, com turmas de até 19 alunos;
- (3) 772 municípios com IDSUS de IDSE e ICS médios (IDSUS 2), IDHm alto, da região Sudeste;
- (4) 292 municípios com IDSUS de IDSE e ICS médios (IDSUS 2), IDHm alto, da região Sul, com despesa em saúde per capita entre R\$ 500,00 e R\$ 800,00;
- (5) 306 municípios com IDSUS de IDSE e ICS médios (IDSUS 2), IDHm alto, da região Sul, com despesa em saúde per capita acima de R\$ 800,00.

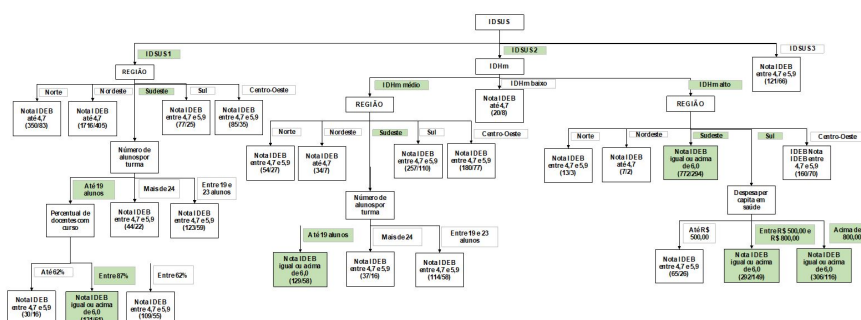


Figura 1 – Árvore de decisão IDEB

Fonte: Elaborado pelos autores.

Nota-se que a despesa paga por aluno no EF não se mostrou relevante para explicar a nota do IDEB. Isso implica que o volume de recursos dispendido por aluno não é determinante para a obtenção dos melhores resultados. As despesas devem ser mais bem alocadas.

## 6.2 Saúde

Na área da saúde, o IDSUS foi escolhido como indicador de referência, uma vez que foi criado para avaliar o desempenho do SUS. Analisou-se 5.564 municípios com dados disponíveis. A Tabela 5 mostra os indicadores que apresentaram o maior ganho de informação em relação ao IDSUS. Entre os indicadores específicos de saúde, destacou-se a despesa per capita com saúde.

Para a formação da árvore de decisão, foram utilizados os cinco indicadores melhor colocados no ranking. Foram escolhidos menos indicadores do que na área de educação, pois o grupo IDSUS 3 é formado apenas por 123 municípios. Por isso, para que esse conseguisse ser visualizado como folha na árvore, foi necessário adotar um menor número de atributos. Os municípios que estão neste grupo são os que apresentaram os melhores resultados do indicador e, portanto, é essencial que esses apareçam para que suas características sejam verificadas.







## 6.5 Saneamento

Para a área de saneamento, foi escolhido como indicador de referência o percentual da população atendida com esgotamento sanitário. Foram analisados 2.437 municípios com informações disponíveis no SNIS. Apenas o percentual da população atendida com abastecimento de água obteve resultado significativo, com um InfoGain de 0,158. Os resultados dos outros indicadores foram abaixo de 0,035. Uma pesquisa mais aprofundada revelou que os municípios possuem uma participação pouco definida no saneamento básico, podendo ser realizado de forma indireta – delegação para autarquia municipal –, companhia estadual, concessionária privada ou de forma direta – sem prestador institucionalizado (Agência Nacional de Águas [ANA], 2017).

## 6.6 Indicadores Fiscais

Para a gestão fiscal, escolheu-se os indicadores DTP e DCL, que compõem os Relatórios de Gestão Fiscal (RGF) divulgados quadrimestralmente pelos municípios. Essa obrigação veio com a publicação da LRF em 2000.

Primeiramente, foram analisados os municípios com dados disponíveis sobre a DTP. Trata-se do percentual da Receita Corrente Líquida (RCL) do ente destinado ao pagamento de pessoal. Foram considerados 2.742 municípios. A Tabela 7 mostra o ranking dos dez indicadores que apresentam o maior ganho de informação em relação à DTP, considerados para formação da árvore com o algoritmo J48.

Tabela 7 – Ranking dos indicadores com maior ganho de informação em relação à DTP

Indicador	InfoGain
Região	0,186
IDHm	0,153
IDSUS	0,151
IDEB	0,139
Despesa per capita em saúde	0,081
Percentual de docentes com curso superior	0,074
Despesa por aluno do EF	0,025
Despesa per capita em assistência social	0,022
Internações por 100.000 habitantes	0,022
Doses aplicadas de vacinação	0,021

Fonte: Elaborado pelos autores.

O percentual de instâncias classificadas corretamente foi de 75,71% e o número mínimo de instâncias por folha foi de 100. Como a classificação da DTP é binária, os municípios foram divididos em dois grupos – aqueles que ultrapassaram o limite e aqueles que não. Na Figura 4, observa-se que os municípios com os melhores resultados na saúde (IDSUS 2 e 3) se encontram com o gasto com pessoal abaixo do limite legal. Os outros apresentam resultados que variam conforme a localização do município. Na região Nordeste, nota-se que os municípios com melhores resultados do IDEB também possuem despesas abaixo do limite. Os outros se encontram com gastos acima do limite. Isso demonstra que o volume de despesas com pessoal compromete a qualidade dos serviços públicos.

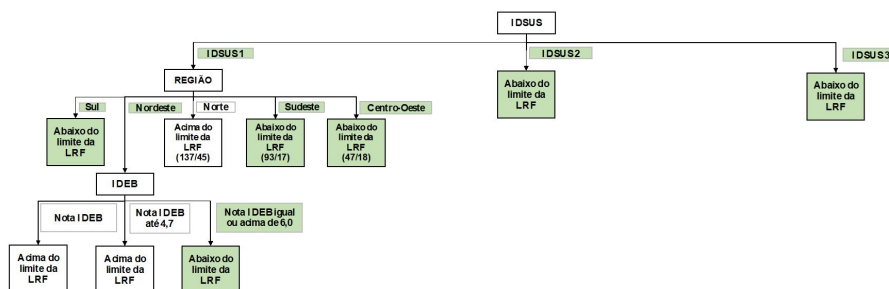


Figura 4 – Árvore de decisão DTP

Fonte: Elaborado pelos autores.

Em seguida, foram analisados os municípios com dados disponíveis de DCL. Essa, por sua vez, refere-se ao total de obrigações financeiras do ente, seja por lei, contrato, convênio ou tratado, o que inclui as operações de crédito com prazo superior a um ano. O acompanhamento desses tipos de gastos é importante para os governos manterem o equilíbrio fiscal (Davis & Biondini, 2018). Foram considerados 1.499 municípios. Os valores de InfoGain foram abaixo de 0,013, ou seja, os atributos não são capazes de explicar as variações nas classes da DCL.

A falta de ganho de informação com os atributos considerados na análise pode ser explicada pela composição da DCL. Ela corresponde ao montante da Dívida Consolidada (DC) desconsideradas as disponibilidades de caixa, as aplicações financeiras e outros haveres financeiros. Não foi possível verificar, portanto, se as áreas abordadas neste trabalho são consideradas dentro da composição da DCL.

## 7. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Este trabalho analisou comparativamente a eficiência na alocação de recursos pelos municípios a partir de dados financeiros e operacionais. No cenário atual de orçamentos engessados e falta de disponibilidade financeira, é importante que as prefeituras saibam alocar seus recursos de maneira estratégica, para que esses consigam gerar os melhores resultados para a sociedade.

A partir da análise dos trabalhos anteriores, nota-se a repetição da metodologia de DEA, na abordagem insumo-produto, diferenciando na amostra e nos outputs escolhidos. Neste trabalho, as técnicas de mineração de dados possibilitaram realizar a pesquisa com o maior número de municípios possível. Foram detectados padrões por meio dos algoritmos InfoGainAttributeEval e J48, que possibilitaram a formação de árvores de decisão. Para aplicação desses algoritmos, foram coletados dados disponíveis das fontes de informação pública que, em seguida, passaram por processos de integração, transformação e limpeza, baseada em critérios estatísticos de regressão.

A base de dados deste trabalho incluiu indicadores gerais e referentes às áreas de educação, saúde, segurança pública, desenvolvimento, saneamento e situação fiscal. Para todas essas áreas, com exceção do desenvolvimento, foram formulados indicadores de despesa per capita ou, no caso da educação, por aluno. A discretização foi feita de maneira comparativa, observando todas as variações de valores nos atributos.

Nas árvores de decisão, as áreas de educação, saúde, desenvolvimento e situação fiscal (DTP) apresentaram resultados significantes. Para segurança pública e saneamento a inexistência de significância se justifica pela indefinição do papel dos municípios. A variação de atuação é tão expressiva que não foi possível estabelecer padrões.

Para a educação, a árvore de decisão mostrou que o atributo que melhor explica as variações de resultados no IDEB foi o IDSUS, seguido da região e do IDHm. Isso demonstra a relação existente entre a educação e a saúde, o que permite concluir que municípios organizados e eficientes em educação tendem a ter um

bom resultado na saúde também. Além disso, percebe-se que o percentual de docentes com curso superior é determinante para os bons resultados no IDEB.

Para a saúde, a região foi o atributo que mais explica as variações de resultados no IDSUS, seguido da despesa em saúde e do IDEB. Em relação às despesas, caberia um estudo mais aprofundado sobre os processos internos de saúde nos municípios, de maneira a verificar se o maior volume de gastos é justificado pela infraestrutura do local ou pela necessidade de atendimento da demanda proveniente de outros municípios. Isso porque as capitais costumam ser os centros de referência para atendimentos mais complexos, o que, em alguns dos casos, pode justificar um maior volume de despesas.

Para o desenvolvimento, o IDSUS é o indicador que melhor explica as variações do IDHm, seguido da localização regional dos municípios e do percentual de docentes com curso superior. Ao retornar na base de dados, verificou-se que na composição do IDHm dos municípios destacados, a longevidade foi a dimensão que mais influenciou na nota final do indicador. Nesse sentido, cabe também a reflexão sobre a distribuição dos serviços de saúde no território brasileiro, principalmente aqueles que envolvem procedimentos de média e alta complexidade.

Para os indicadores fiscais, apenas a DTP apresentou resultados significativos. Os grupos com a despesa abaixo do limite estabelecido pela LRF apresentaram os melhores resultados no IDEB e no IDSUS. Isso mostra que os gastos excessivos com pessoal podem prejudicar o desempenho do município nas áreas de educação e saúde. Além disso, esse resultado reforça a importância da regulação para determinados tipos de despesas. A STN precisa realizar um controle efetivo desse tipo de gasto, aplicando as sanções previstas na LRF. Não foi possível verificar a existência desse controle. Em relação à DCL, os resultados não se mostraram significativos pois não foi possível verificar, explicitamente, nenhuma das áreas analisadas em sua composição.

Em relação às despesas públicas, na área de educação, verificam-se municípios capazes de apresentar uma boa nota no IDEB com despesa por aluno menor que R\$ 6.000,00. A maior parte dos municípios com desempenho destacado em saúde apresentaram despesas na faixa de R\$ 500,00 e R\$ 800,00. Isso permitiu a conclusão de que a eficiência alocativa também pode ser considerada para os municípios que gastam esses montantes em saúde. Já no caso do desenvolvimento, a combinação das despesas de educação e saúde permitiu identificar municípios eficientes tanto em apenas uma delas quanto nas duas. Já as despesas das outras áreas variaram de maneira significativa em todos os grupos analisados, o que não permitiu gerar conclusões.

Por fim, foi possível identificar os municípios com desempenhos destacados e considerados eficientes do ponto de vista da alocação de recursos, o que permitiu conclusões que corroboraram os resultados das árvores.

O investimento na formação dos professores se mostrou altamente determinante para o desempenho municipal. Os municípios com os melhores resultados apresentaram, em comum, metas de formação superior e de pós-graduação dos docentes. Na saúde, as ações variaram. Foram identificadas ações de ampliação do horário de atendimento, de diálogo entre os atores para melhora de processos e de humanização dos atendimentos. O município que pode ser considerado o mais eficiente no conjunto das áreas analisadas, dentro dos parâmetros e comparações estabelecidas neste trabalho, é o de São José/SC.

A principal limitação deste trabalho é a qualidade dos dados públicos, que são disponibilizados das mais variadas formas nos portais dos órgãos governamentais. Os dados de despesa poderiam ser mais detalhados. Na educação, por exemplo, seria interessante distinguir os gastos com transporte escolar, merenda, infraestrutura, salário e formação dos professores, etc.

Além disso, seria importante o aumento da periodicidade de atualização de algumas bases de dados como, por exemplo, a que se refere ao IDSUS. Outras bases, como a do SNIS, poderiam apresentar melhor detalhamento sobre o significado dos dados, mas metadados deficientes são uma característica de todas as bases utilizadas. Verifica-se também a necessidade de continuidade dessas coletas de dados, de forma que ao longo do tempo se constituam séries históricas para a realização de análises preditivas.

Mesmo considerando as limitações dos dados, é possível propor outros tipos de análises, como estudos comparativos entre estados/regiões, avaliação dos municípios ineficientes (gastos elevados e indicadores

ruins) e dos municípios presos na armadilha dos recursos limitados (gastos baixos e indicadores ruins). Além disso, o uso da mineração de dados mostrou que o enriquecimento dos dados é um insumo importante para a descoberta de novos padrões, por isso, um trabalho futuro consistiria em expandir o conjunto de fontes de dados utilizadas neste estudo.

## REFERÊNCIAS

- Agência Nacional de Águas (2017). *Atlas Esgotos: Despoluição de Bacias Hidrográficas*. Brasília: Autor.
- Almeida, J.E.G., Vasconcelos, M.T.C., Miranda, L.C., Feitosa, M.O. (2017, novembro) Gasto público e sua influência no Índice de Desempenho do Sistema Único de Saúde dos municípios pernambucanos. In *Anais do XXIV Congresso Brasileiro De Custos*, Florianópolis, SC. Recuperado de <https://anaiscbc.emnuvens.com.br/anais/article/view/4322>
- Baitharu, T.R. Pani, S.K. (2016). Analysis of Data Mining Techniques For Healthcare Decision Support System Using Liver Disorder Dataset. *International Conference on Computational Modeling and Security*, 85(1), p. 862-870. Recuperado de <https://core.ac.uk/download/pdf/82422223.pdf>
- Brennan, G., Buchanan, J. (2000). The Power to Tax, Analytical Foundations of a Fiscal Constitution. In: *Collected Works of James Buchanan*. Indianapolis: Liberty Fund.
- Cândido, J. O. Júnior (2001). Os gastos públicos no Brasil são produtivos? *Planejamento e Políticas Públicas*, 23, 233-260. Recuperado de <https://www.ipea.gov.br/ppp/index.php/PPP/article/view/77/88>
- Costa, C.C.M., Ferreira, M.A.M., Braga, M.J., Abrantes L.A. (2015). Fatores associados à eficiência na alocação de recursos públicos à luz do modelo de regressão quantílica. *Revista de Administração Pública*, 5(49). 1319-1347. Recuperado de [www.scielo.br/pdf/rap/v49n5/0034-7612-49-05-01319.pdf](http://www.scielo.br/pdf/rap/v49n5/0034-7612-49-05-01319.pdf)
- Costa, F.L., Castanhar, J.C. (2003). Avaliação de programas públicos: desafios conceituais e metodológicos. *Revista de Administração Pública*, 5(37), 969-992. Recuperado de <http://bibliotecadigital.fgv.br/ojs/index.php/rap/article/view/6509/5093>
- Cruz, C. F., Afonso, L.E. (2018). Gestão fiscal e pilares da Lei de Responsabilidade Fiscal: evidências em grandes municípios. *Revista de Administração Pública*, 52(1), 126-148. Recuperado de [http://www.scielo.br/scielo.php?script=sci\\_abstract&pid=S0034-76122018000100126&lng=en&nrm=iso&tlng=pt](http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0034-76122018000100126&lng=en&nrm=iso&tlng=pt)
- Davis, P.G., Biondini, I.V.F. (2018). Endividamento Público de Minas Gerais: Histórico e Recentes Propostas de Recuperação. *Cadernos Gestão Pública e Cidadania*, 23(76), 413-435. Recuperado de <http://bibliotecadigital.fgv.br/ojs/index.php/cgpc/article/view/75373>
- Di Pietro, M.S.Z. (2017). *Direito Administrativo*. 30 ed. Rio de Janeiro: Editora Forense Ltda.
- Giacomoni, J. (2017). *Orçamento Público* (17a ed.). São Paulo, SP: Atlas.
- Han, J. & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2a ed.). San Francisco, CA: Elsevier.
- Litvack, J., Ahmad, J., Bird, R. (1998). Rethinking Decentralization in Developing Countries. *The World Bank: Sector Studies Series* (p. 1-48). Washington DC.
- Lourenço, R. L., Angotti, M., Nascimento, J.C.H.B., Sauerbronn, F. F. (2017). Eficiência do gasto público com ensino fundamental: uma análise dos 250 maiores municípios brasileiros. *Revista Contabilidade Vista & Revista*. 28(1), 89-116. Recuperado de <http://revistas.face.ufmg.br/index.php/contabilidadevistaerevista/article/view/3406>
- Machado, S.P. Jr., Irffi, G., Benegas, M. (2011). Análise da eficiência técnica dos gastos com educação, saúde e assistência social dos municípios cearenses. *Planejamento e Políticas Públicas (PPP)*, 1(36), 87-113. Recuperado de <http://www.ipea.gov.br/ppp/index.php/PPP/article/view/223/204>
- Matos, E., Ponczek, V. (2013). Efeitos da divisão municipal na oferta de bens públicos e indicadores sociais. *Revista Brasileira de Economia*, 67(3), 315-336. Recuperado de [https://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0034-71402013000300004](https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-71402013000300004)



- Musgrave, R. A. (1983). Who should tax, where, and what. In: MCLURE, Charles (Ed.). *Tax assignment in federal countries* (p. 2-19). Canberra: Centre for Research on Federal Financial Relations, Australian National University.
- Oates, W. (1972). *Fiscal federalism*. Nova York: Harcourt Brace Jovanovich.
- Oliveira, F. A. (2007). Teorias da Federação e do Federalismo Fiscal: o caso brasileiro. Escola de Governo da Fundação João Pinheiro, Belo Horizonte Texto para Discussão n. 43. Recuperado de <http://www.eg.fjp.mg.gov.br/index.php/docman/publicacoes-2007/84-teorias-da-federacao-e-do-federalismo-fiscal-o-caso-brasileiro>
- Oliveira, F.A. (2012). *Economia e política das finanças públicas no Brasil*. Campinas, SP: Hucitec.
- Politelo, L., Rigo, V.P., Hein, N. (2014). Eficiência da aplicação de recursos no atendimento do Sistema Único de Saúde (SUS) nas cidades de Santa Catarina. *Revista de Gestão em Sistemas de Saúde*, 3(2), 45-60. Recuperado de <http://www.revistargss.org.br/ojs/index.php/rgss/article/view/105/141>
- Portullhak, H., Raffaelli, S.C.D., Scarpin, J.E. (2018). A Eficiência da Aplicação de Recursos Voltada à Saúde Pública nos Municípios Brasileiros. *Revista Contabilidade, Gestão e Governança*, 21(1), 21-39. Recuperado de <https://cg-amg.unb.br/index.php/contabil/article/view/1154/pdf>
- Programa das Nações Unidas para o Desenvolvimento (2013). *O Índice de Desenvolvimento Humano Municipal Brasileiro: Atlas do Desenvolvimento Humano do Brasil*. Brasília: Portal Print Gráfica e Editora LTDA-ME, 2013.
- Rezende, A. M. R. T., Carneiro, R., Rezende, J. V. S. (2014). Federalismo Fiscal e Crise das Finanças Públicas no Brasil: uma análise de indicadores fiscais nos Estados de 2007 a 2012. In *Anais do XXXVIII Encontro da ANPAD*, Rio de Janeiro, RJ. Recuperado de [http://www.anpad.org.br/admin/pdf/2014\\_EnANPAD\\_APB2208.pdf](http://www.anpad.org.br/admin/pdf/2014_EnANPAD_APB2208.pdf)
- Ricardo, C.M., Caruso, H.G.C. (2007). Segurança pública: um desafio para os municípios brasileiros. *Revista Brasileira de Segurança Pública*, 1(1), 102-119. Recuperado de <http://revista.forumseguranca.org.br/index.php/rbsp/article/download/10/7>
- Salomão, B.A., Saiani, C.S. (2019). Descentralização, Esforço Fiscal e Captura de Recursos Públicos: Efeitos da Criação de Municípios no Brasil. *47º Encontro Nacional de Economia*, ANPEC. Recuperado de [https://www.anpec.org.br/encontro/2019/submissao/files\\_I/i5-f23f15cdbf19827ed56f4b01deafe793.pdf](https://www.anpec.org.br/encontro/2019/submissao/files_I/i5-f23f15cdbf19827ed56f4b01deafe793.pdf)
- Salomão, B. A. Neto (2020). Problemas fiscais, efeito Flypaper e hipótese do Leviatã em Uberlândia: diagnóstico e proposição de soluções. *Revista de Desenvolvimento e Políticas Públicas*, 4(1), 21-39. Recuperado de <https://redapp.ufv.br/REDEPP/article/view/83/36>
- Salomão, B. A. Neto, Santos, J.F.C. (2020). Dependência Fiscal, Ilusão Fiscal e Dificuldades Orçamentárias nos Municípios Brasileiros: Diagnóstico a partir de Modelos de Painel Logit/Probit. *53ª Jornadas Internacionales de Finanzas*. Recuperado de <http://jifp.eco.unc.edu.ar/ediciones-antiores/anales-publicaciones-2020>
- Sampieri, R.H., Collado, C.F., Lucio, P.B. (2006). *Metodologia de Pesquisa* (3a ed.). São Paulo, SP: McGraw-Hill Interamericana.
- Samuelson, P. (1954). The pure theory of public expenditure. *Review of Economics and Statistics*, 36(4), 387-389.
- Savian, M.P.G., Bezerra, F.M. (2013). Análise de eficiência dos gastos públicos com educação no ensino fundamental no estado do Paraná. *Economia & Região*, 1(1), 26-47. Recuperado de <http://www.uel.br/revistas/uel/index.php/ecoreg/article/view/12963/12458>
- Shah, A. (1991). The New Fiscal Federalism in Brazil. *World Bank Discussion Papers*, 124, 1-130.
- Silva, M. S. (2005). Teoria do federalismo fiscal: notas sobre as contribuições de Oates, Musgrave, Shah e Ter-Minassian. *Nova Economia*, 15(1), 117-137. Recuperado de <https://revistas.face.ufmg.br/index.php/novaeconomia/article/view/447>
- Silva, M.R.C.S., Costa, E.P.A., Castro, M.A.R.C., Silva, L.P. (2016). O Efeito dos Gastos Públicos no Nível de Desenvolvimento de Municípios no Brasil: os Gastos Municipais Importam? In *Anais VII Encontro de Administração Pública e Governança*, São Paulo, SP.
- Tiebout, C. (1956). A pure theory of local expenditures. *The Journal of Political Economic*, 64(5), 416-424.
- Wagner, A. (1890). *Finanzwissenschaft*. Leipzig: C.F. Winter'sche Verlagshandlung.



Zaki, M.J.; Meira, W.Jr. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. New York: Cambridge University Press.