

## Uso de Agrupamento para Avaliação de Desempenho Educacional e Apoio à Gestão em Áreas de Investimento

Mariana Spanol<sup>1</sup>, Emília Oliveira<sup>2</sup>, Gabriel Alves<sup>2</sup>,  
Ibsen Mateus Bittencourt<sup>4</sup>, Taciana Pontual Falcão<sup>2</sup>, Rafael Ferreira Mello<sup>2,3</sup>

<sup>1</sup> Universidade de São Paulo (USP)  
São Paulo – SP – Brasil

<sup>2</sup> Departamento de Computação - Universidade Federal Rural de Pernambuco (UFRPE)

<sup>3</sup> Centro de Estudos e Sistemas Avançados do Recife (CESAR)

<sup>4</sup> Universidade Federal de Alagoas (UFAL)

marianaspanol@usp.br, ibsen@feac.ufal.br  
{emilia.galdino,gabriel.alves,taciana.pontual,rafael.mello}@ufrpe.br

**Abstract.** *The National Education Plan (PNE) defines goals and indicators that relate to education, and managers of federated entities in Brazil must address them. In the case of the municipal manager, observing the practices of municipalities with similar characteristics that have achieved better results in the PNE indicators can help in the decision on the use of the resources. This work proposes an innovative method with two stages: (1) Clusterization of Brazilian municipalities with unsupervised learning based on the sociodemographic metrics; (2) Comparison of educational metrics between a municipality and the other municipalities of its cluster. This study shows an improvement in the Silhouette Coefficient, confirming that our method created more cohesive clusters. To exemplify this approach, we analyze the Indicator 1A of Goal 1 of the PNE for the city of Salvador. We observed that the good results for some educational metrics did not reflect a significant improvement in the assessed PNE indicator.*

**Resumo.** *O Plano Nacional de Educação (PNE) define metas e indicadores relacionados à educação, que devem ser observados pelos gestores dos entes federados do Brasil. No caso do gestor municipal, observar práticas de municípios com características parecidas e que tenham obtido resultados melhores nos indicadores do PNE, pode auxiliar na decisão sobre a utilização dos recursos. Este trabalho propõe um método inovador que é dividido em duas etapas: (1) Agrupamento das municipalidades brasileiras com aprendizado não-supervisionado com base em métricas sociodemográficas; (2) Comparação de métricas educacionais entre o município e os demais municípios do seu grupo. Este estudo mostrou uma melhora no Coeficiente de Silhueta, atestando que o método proposto criou grupos mais coesos. A fim de exemplificar esta abordagem, foi analisado o Indicador 1A da Meta 1 do PNE, para o município de Salvador. Foi observado que os bons resultados para algumas métricas educacionais não refletiram um aumento significativo no indicador do PNE avaliado.*

## 1. Introdução

O Plano Nacional de Educação – PNE tem o objetivo de organizar um sistema nacional de educação por meio da definição de diretrizes, metas e estratégias voltadas para a qualidade do ensino em seus diversos níveis. Este plano define um conjunto de 20 metas voltadas para a educação no Brasil e que devem ser cumpridas dentro do seu prazo de vigência de 2014 a 2024 [Brasil 2014]. Para tanto, os entes federados (estados, municípios e Distrito Federal) devem elaborar seus planos de forma a atingir os indicadores relacionados às metas do PNE. A elaboração destes planos é uma tarefa minuciosa e especializada que deve otimizar o uso dos recursos da educação com o intuito de cumprir as metas do PNE.

Ao elaborar seus planos de educação e ao monitorar a situação dos indicadores do PNE, gestores da educação brasileiros procuram obter informações não apenas sobre o ente por eles geridos, mas também sobre a situação dos outros entes. Esta comparação permite que se tomem ações a fim de melhorar os seus próprios indicadores. A disponibilização de dados abertos e de visualizações criadas por diferentes entidades [Todos pela Educação 2022, MEC 2022, TCE-SC 2022, MP-SC 2022] auxiliam estes gestores. Contudo, apesar destas iniciativas representarem um grande avanço na gestão baseada em dados, nem todos os gestores conseguem extrair o máximo que estas iniciativas podem oferecer. Ademais, a análise enviesada dos dados e informações disponibilizadas pode levar estes gestores a aplicar seus recursos em áreas que não gerem a melhoria nos indicadores educacionais.

O suporte à tomada de decisão dos gestores da educação deve ser subsidiado por informações inteligíveis e livres de viés. Considerando que o Brasil possui 5.570 municípios (incluindo o Distrito Federal e o Distrito Estadual de Fernando de Noronha) [IBGE 2022], é necessário que se leve em consideração as especificidades destes entes ao se comparar os indicadores das metas do PNE. Comparar indicadores de municípios com características muito discrepantes pode levar gestores a conclusões equivocadas.

Este trabalho tem como principal contribuição a proposta de um método de suporte à decisão para gestores da educação. O aprendizado não-supervisionado é utilizado para criar grupos de municípios com base em seus indicadores socioeconômicos. Assim, é possível comparar métricas educacionais, como a quantidade de escolas com acesso à internet, entre municípios de um mesmo grupo. Com isso, o gestor pode utilizar estas comparações a fim de priorizar o direcionamento de recursos.

O método proposto pode ser aplicado para quaisquer indicadores do PNE. Neste trabalho, apresenta-se a sua utilização na análise do Indicador 1A da Meta 1 do PNE, para o município de Salvador. O Indicador 1A define que a educação infantil na pré-escola para as crianças de 4 a 5 anos de idade deveria ser universalizada até 2016.

De forma a guiar o estudo a ser realizado, foram desenvolvidas as seguintes perguntas de pesquisa que deverão ser respondidas ao longo deste trabalho.

- PP1: É possível criar um agrupamento entre os municípios que seja coerente e que facilite a comparação entre os indicadores das metas do PNE?
- PP2: Como auxiliar o gestor municipal a comparar a situação do seu município com relação aos demais a fim de otimizar o investimento de recursos para cumprir as metas do PNE?

## 2. Trabalhos Relacionados

O avanço e popularização das técnicas de Aprendizado de Máquina possibilitou que diferentes aplicações pudessem se beneficiar do uso das suas técnicas e métodos. Dentre as possibilidades de aplicações, se destaca a utilização dessas técnicas para auxiliar na criação de sistemas de apoio à decisão [Silva et al. 2021, Zapparolli et al. 2017]. Em especial, a área de *Learning Analytics* aproveita conceitos de áreas como *Big Data* e Ciência de Dados, se utilizando por exemplo de algoritmos de Inteligência Artificial e técnicas de visualizações de dados de tal forma que os gestores educacionais possam tomar decisões baseadas em dados [Clow 2013].

Nesse contexto, diferentes sistemas de apoio à gestão foram propostos na literatura [Herodotou et al. 2019, Gutiérrez et al. 2020, Zapparolli et al. 2017, Silva et al. 2021]. No sentido de acompanhar o desempenho de turmas em um ambiente virtual de aprendizagem, [Zapparolli et al. 2017] propõem uma ferramenta de visualização de interações entre professores e alunos. No trabalho de [Silva et al. 2021] foram utilizados algoritmos e técnicas de Inteligência Artificial para prever a produção de mel a partir de informações fornecidas pelo apicultor.

Outra técnica comum nessa área, é o agrupamento, que permite encontrar similaridades entre os elementos, e assim agrupá-los de forma que os elementos de um grupo sejam semelhantes entre si [Jordan and Mitchell 2015]. Esta técnica é especialmente importante quando é preciso realizar comparações entre elementos, sobretudo quando estes possuem características que podem variar de forma ampla. De forma a evitar comparações enviesadas entre elementos de características discrepantes, o agrupamento se mostra uma alternativa para realizar tais comparações de forma mais adequada.

Na literatura, o trabalho de [Bandeira and Reyes Junior 2021] compara municípios do Brasil em relação à eficácia na realização de programas federais, procurando descrever as características relacionadas com o desempenho dos municípios. Em [Abreu et al. 2020] é proposto um agrupamento de alunos para identificar os diferentes estilos de aprendizagem. Em ambos trabalhos, apesar dos contextos distintos, a formação de grupos é utilizada como ferramenta para proporcionar a análise de fatores que influenciaram a formação de tais grupos.

O presente trabalho se diferencia dos demais por apresentar um método inovador para o apoio à gestão educacional de municípios. A partir da combinação do agrupamento de municípios similares, com a análise multicritério realizada com o auxílio das visualizações de gráficos de caixa, os gestores podem se beneficiar de uma comparação entre municípios semelhantes, além de uma visualização simplificada, mas que também permite a inspeção de informações mais detalhadas sobre a distribuição dos dados.

## 3. Método

Este trabalho propõe um método de suporte à decisão de gestores municipais da educação no Brasil. Este método pode ser dividido em duas grandes etapas: o *agrupamento* [Jordan and Mitchell 2015] de municípios e o *apoio à decisão multicritério* com base nas métricas educacionais [Greco et al. 2016]. A primeira etapa se baseia no agrupamento de municípios com base em seus indicadores socioeconômicos. Técnicas de aprendizado de máquina não-supervisionado foram utilizadas para este fim, permitindo uma análise multivariada para a criação dos referidos grupos.

Na segunda etapa do método proposto, espera-se que uma vez que os municípios pertencentes a um mesmo grupo possuem indicadores socioeconômicos equivalentes, também possuam indicadores educacionais equivalentes. Assim, caso um município esteja em um grupo com um alto desenvolvimento socioeconômico, mas seus indicadores do PNE estiverem aquém dos indicadores dos demais municípios de seu grupo, há indícios que o gestor deste município deveria priorizar a destinação de recursos para a melhoria destes indicadores. Os dados utilizados para o presente estudo foram provenientes dos resultados do PNE a partir das estimativas da *Pesquisa Nacional de Amostra por Domicílio Contínua* (PNAD Contínua) [PNAD Contínua 2022].

### 3.1. Agrupamento

Dada a enorme extensão territorial do Brasil e as grandes diferenças geopolíticas e socioeconômicas de seus municípios, comparar seus indicadores em quaisquer domínios pode levar a conclusões enviesadas ou mesmo erradas. Por isso, o método aqui proposto cria grupos de municípios com base em seus indicadores socioeconômicos a fim de reduzir possíveis vieses nas análises realizadas pelos gestores públicos.

Assim, os municípios são agrupados considerando a *Renda per Capita*, o *Índice de Desenvolvimento Humano* (IDH) municipal e a *População*. Estas variáveis foram selecionadas para a criação dos grupos uma vez que são facilmente obtidas a partir de publicações oficiais e confiáveis [IBGE 2022] e são amplamente utilizadas e aceitas na literatura para caracterização de territórios e perfil populacional [Gordia 2010].

Este agrupamento tem o objetivo de responder à primeira pergunta de pesquisa. Para tanto, será utilizado um algoritmo de agrupamento (*clustering*) que pertence à classe de técnicas de aprendizado não-supervisionado [Jordan and Mitchell 2015]. Este tipo de algoritmo procura agrupar os elementos da população em grupos com base em suas variáveis, quando não há uma variável que indique a classe a que cada elemento pertence.

Os algoritmos de agrupamento normalmente são sensíveis aos *outliers*, fazendo com que os elementos representados por estes *outliers* fiquem isolados em grupos muitas vezes compostos apenas por eles mesmos [Gan and Ng 2017]. Dada a natureza do problema aqui apresentado, de comparar um município com outros municípios a fim de auxiliar os gestores da educação, é importante evitar a ocorrência de grupos desse tipo. Neste sentido, antes de aplicar o agrupamento, os *outliers* superiores devem ser substituídos pelos limites superiores dos dados, enquanto os *outliers* inferiores, devem ser substituídos pelos limites inferiores dos dados. Para este fim, foi utilizado o valor de  $1,5 \times IQR$ , onde *IQR* é o *Intervalo Inter Quartil* para o cálculo dos limites [Montgomery and Runger 2018].

Ao aplicar algoritmos de agrupamento, pode-se definir uma quantidade arbitrária de grupos ou utilizar técnicas para auxiliar nesta decisão. O método aqui proposto aplica a análise de silhueta [Rousseeuw 1987] que calcula o coeficiente para grupos formados a partir de diferentes iterações do mesmo algoritmo de agrupamento com uma quantidade de grupos distinta em cada iteração. Quanto mais próximo de 1 for o Coeficiente de Silhueta mais bem definidos estão os grupos, enquanto um coeficiente próximo de 0 (zero) indica que os grupos não possuem uma separação clara entre si.

Assim, considera-se que a primeira pergunta de pesquisa será respondida através do Coeficiente de Silhueta. Quanto mais próximo a 1 seja este coeficiente, maior é o indício

de que as variáveis utilizadas no agrupamento são válidas e que os grupos estão coesos, permitindo a comparação dos indicadores do PNE entre os seus municípios.

### 3.2. Apoio à Decisão Multicritério

Após a criação dos grupos, o gestor da educação pode comparar de forma equânime a situação dos municípios com relação às metas do PNE. Uma vez selecionado um indicador de uma meta do PNE é possível não apenas observar a situação dos municípios do Brasil, de uma Região ou de um Estado, mas também a situação dos municípios que pertencem a um mesmo grupo.

A partir do indicador selecionado, este método propõe a criação de um novo grupo, doravante denominado *Grupo+*, composto pelos municípios do grupo que obtiveram resultados superiores à mediana do grupo. Com isso, o gestor pode analisar como os municípios deste subgrupo investem seus recursos da educação e usar esta informação ao tomar decisões para seu próprio município. Por exemplo, se o gestor observar que a maior parte dos municípios do *Grupo+* possuem mais escolas com acesso a internet por estudante que o seu município, este gestor pode decidir investir em um projeto de acesso universal à internet para as escolas.

Este trabalho propõe o uso de diagramas de caixa [Montgomery and Runger 2018] para a visualização dos dados referentes às métricas educacionais e indicadores do PNE. A escolha deste tipo de diagrama se deu pela sua capacidade de denotar não apenas os valores médios da amostra, mas os seus *outliers* e intervalos em que os dados se concentram. Os gráficos são criados com um diagrama diferente para os dados do *Estado*, *Estado+*, *Grupo* e *Grupo+*. Uma linha de referência para o município é incluída para que o gestor possa observar e comparar seu município com relação aos demais municípios do seu estado e do seu grupo.

A escolha das métricas educacionais utilizadas nos estudos a seguir se deu pela disponibilidade e confiabilidade dos dados, obtidos a partir da PNAD Contínua [PNAD Contínua 2022], e pela relação das variáveis com o Indicador 1A da Meta 1 do PNE, que é o foco do estudo deste trabalho. Assim, foram selecionadas as métricas: *número de docentes de pré-escola e educação infantil*, *acesso à internet*, *acesso à internet para uso no ensino e aprendizagem*, *espaços de atividades de integração da escola e comunidade* e outras métricas relacionadas de maneira mais indireta à meta.

## 4. Agrupamento dos Municípios Brasileiros

Esta seção apresenta as análises realizadas a fim de responder à primeira pergunta de pesquisa que trata da criação de grupos para os municípios. Neste sentido, serão apresentados os resultados obtidos sem e com o tratamento dos *outliers* proposto.

A Tabela 1 apresenta os resultados do cálculo do coeficiente de silhueta sem e com o tratamento dos *outliers*, considerando de 2 a 9 grupos (*clusters*). Observa-se que o tratamento dos *outliers* melhorou o resultado para todos os experimentos realizados, ressaltando a sua eficácia. Ressalta-se ainda que o melhor resultado foi obtido para a criação de 3 grupos, obtendo-se um coeficiente de silhueta de 0,56.

A criação dos grupos utilizou os dados de *Renda per Capita*, do *Índice de Desenvolvimento Humano* (IDH) municipal e da *População* provenientes do último Censo nacional realizado pelo IBGE no ano de 2010. Esta etapa deve ser realizada sempre que for

Grupos	Coeficiente de Silhueta	
	SEM Método	COM Método
2	0,44	0,55
3	0,45	0,56
4	0,41	0,49
5	0,38	0,48
6	0,33	0,44
7	0,34	0,43
8	0,33	0,40
9	0,31	0,41

**Tabela 1. Coeficiente de Silhueta SEM e COM o tratamento de *outliers*.**

realizado um novo Censo ou estudo equivalente, proveniente de órgão oficial e confiável. Uma vez que a população dos municípios do Rio de Janeiro e de São Paulo é bem superior que a dos demais municípios brasileiros, estes tendiam a ficar em um grupo isolado, quando não se realizava o tratamento dos *outliers* proposto. Este tratamento atribui o limite superior da população nestes casos. Com isso, o município tende a ser colocado em grupos com maior população. Esta decisão do método se mostrou acertada pois, caso fosse atribuído um valor médio, por exemplo, os municípios poderiam ser colocados em grupo com municípios características diferentes das suas.

Por fim, considerando as métricas indicadas, os municípios brasileiros foram agrupados em 3 grupos utilizando o algoritmo de aprendizado não-supervisionado *K-Means* que avalia e agrupa as informações de acordo com suas características. Esse algoritmo consiste na avaliação dos grupos de dados de acordo com a análise da inércia, que busca formar grupos com menor distância entre os elementos do mesmo grupo [Arthur and Vassilvitskii 2006].

Uma vez que os municípios de um mesmo grupo possuem características socioeconômicas e geopolíticas similares, é esperado que seus indicadores educacionais também sigam o mesmo comportamento. Portanto, como os grupos criados foram coerentes, conforme mostra o Coeficiente de Silhueta, considera-se que o método criado responde de forma satisfatória à primeira pergunta de pesquisa.

## 5. Apoio à Decisão Multicritério com Agrupamentos

Esta seção se propõe a responder à segunda pergunta de pesquisa, na qual se procura investigar formas de dar apoio à decisão de gestores municipais da educação, ao analisar diferentes critérios e municípios com o objetivo de priorizar a destinação de recursos. A fim de exemplificar esta aplicação do método, será apresentado um estudo de caso considerando o município de Salvador. Estudos equivalentes podem ser conduzidos para quaisquer municípios brasileiros.

A Figura 1 mostra os indicadores utilizados para o agrupamento (Seção 4) dos municípios com o foco no grupo do município de Salvador. São mostrados os diagramas de caixa para todos os municípios do Brasil, da Região, do Estado e do Grupo do qual Salvador faz parte. Vale frisar que o grupo pode possuir municípios de diferentes estados e regiões do país. Os diagramas desta figura mostram que Salvador ocupa uma posição de

destaque em relação a todos os pares analisados: Brasil, Região, Estado e Grupo. No caso do IDH (Figura 1a), ele está no terceiro quartil do seu grupo e bem próximo do quarto quartil. Ou seja, entre os maiores IDH do seu grupo. Analisando a região e o estado, ele chega a ser considerado um *outlier* positivo, estando bem acima de seus pares.

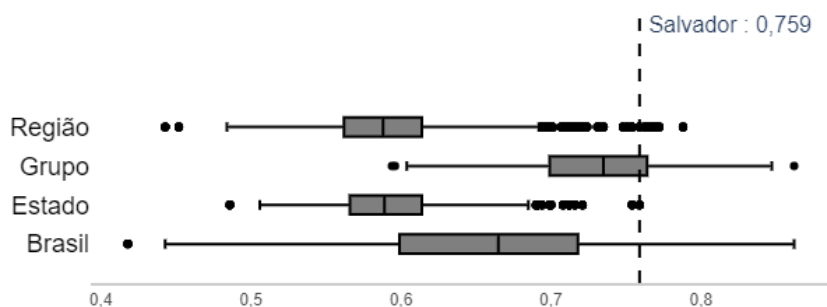
Observando a Renda per Capita (Figura 1b), Salvador também ocupa lugar de destaque. No caso do seu grupo e do Brasil, está entre os 25% dos municípios com maior Renda per Capita. Observando-se o estado e a região, Salvador é considerado um *outlier* positivo estando bem acima dos demais municípios. Quanto à população (Figura 1c), Salvador está bem acima dos demais municípios em todas as comparações. No caso específico da população, é comum a ocorrência de *outliers* dada a maior densidade demográfica em capitais. Considerando-se as variáveis apresentadas, seria esperado que este município ocupasse posição equivalente também em relação aos indicadores educacionais. Contudo, como será apresentado a seguir, este comportamento não ocorre, sendo facilmente detectado pelo método proposto.

A Figura 2 apresenta o resultado atual do Indicador 1A da Meta 1 do PNE para o município de Salvador, comparando este resultado com municípios do próprio estado e do seu grupo. Diferentemente do que ocorre com relação às variáveis utilizadas para realizar os agrupamentos, Salvador ocupa uma posição de destaque negativo com relação ao Indicador 1A. Tanto com relação ao estado quanto ao seu grupo, o município está entre os 25% dos municípios com a menor proporção de crianças (60,89%) de 4 a 5 anos de idade matriculadas na pré-escola. Vale salientar que a mediana da proporção de crianças nesta faixa etária matriculadas na pré-escola é de 93% e 88%, respectivamente, em seu estado e em seu grupo.

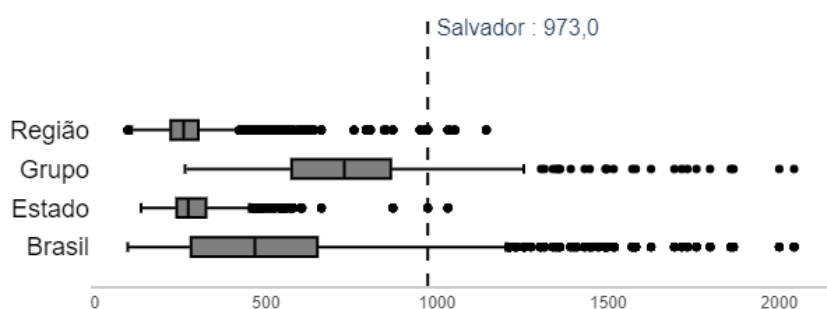
Dado o resultado abaixo do desejado com relação ao Indicador 1A, é importante comparar a situação de Salvador com relação às métricas relacionadas à educação, especialmente ao Indicador 1A. Para isso, a Figura 3 denota estas métricas, apresentando um diagrama de caixa com uma série para *Grupo*, *Grupo+*, *Estado* e *Estado+*. Além de permitir a comparação com o estado e seu grupo, a comparação com os municípios que obtiveram os melhores resultados com relação ao Indicador 1A em seu grupo (*Grupo+*) e em seu estado (*Estado+*) auxilia o gestor na tomada de decisão.

Os gráficos da Figura 3 mostram que Salvador está entre os melhores do seu grupo apenas com relação à quantidade de escolas com espaços para atividades de integração, com aproximadamente 33 escolas para cada 1000 crianças entre 4 a 5 anos (Figura 3a). Salvador também está apenas um pouco abaixo da mediana do seu grupo e do seu estado, com suas quase 11 escolas com material disponível para prática esportiva para cada 1000 crianças (Figura 3d). Observando as séries referentes ao *Grupo+*, aproximadamente 50% dos municípios que obtiveram os melhores resultados no Indicador 1A tiveram um resultado inferior a Salvador nestas duas métricas. Logo, seria possível afirmar que não é nestas duas áreas que o gestor do município deveria focar novos investimentos, se seu objetivo for melhorar os resultados do Indicador 1A.

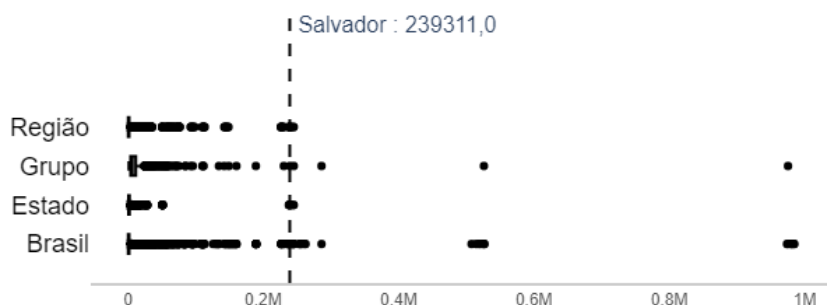
A Figura 3b mostra que a situação de Salvador com relação à quantidade de escolas com biblioteca está abaixo do primeiro quartil para todos os casos. No caso dos municípios do *Estado+*, Salvador sequer chega a estar dentro do limite inferior deste grupo. Já o *Grupo+* possui uma mediana de 18 escolas com biblioteca para cada 1000



(a) IDH.



(b) Renda per Capita em Reais.



(c) População.

**Figura 1. Diagramas de Caixa com as variáveis utilizadas para o agrupamento, apresentando o grupo e a linha de referência para o município de Salvador.**

crianças, o que é mais de 100 vezes as 0,136 escolas com biblioteca de Salvador. Apesar desta métrica não estar relacionada diretamente à educação de crianças de 4 a 5 anos, esta pode ter uma relação indireta. Por exemplo, pode ser que municípios que possuem mais escolas com biblioteca também façam ações voltadas para a matrícula de crianças de 4 a 5 anos. Ou mesmo que indique um investimento maior na educação como um todo.

Por fim, foi avaliado o indicador de escolas com espaços de atividade de integração (Figura 3). Neste caso, diferentemente dos outros, os grupos que melhor desempenharam foram *Estado+* e *Estado*, com mediana de 50 escolas a cada mil habitantes. Para este indicador, Salvador apresentou 33 escolas a cada mil habitantes. Os indicadores educa-



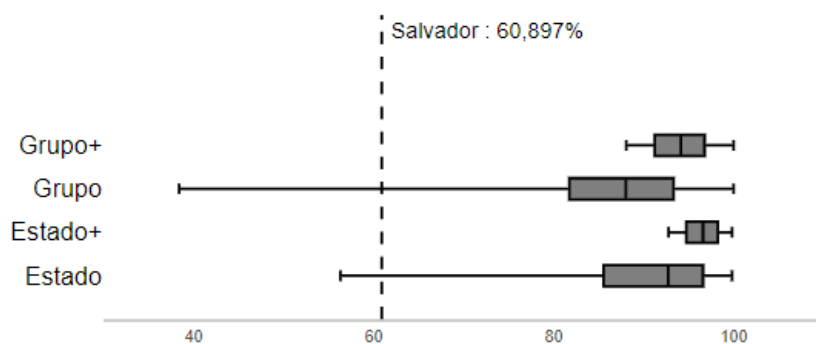
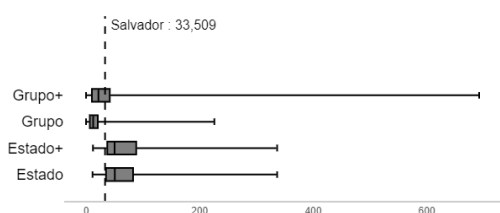
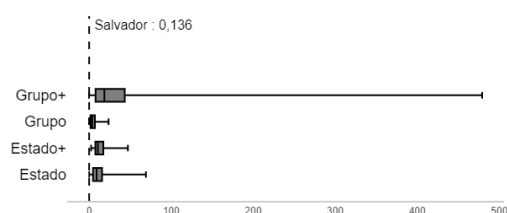


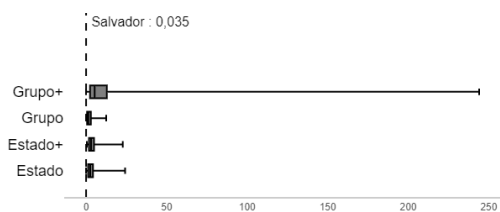
Figura 2. Situação de Salvador em relação ao Indicador 1A da Meta 1 do PNE.



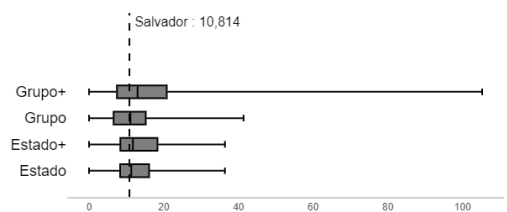
(a) Escolas com espaço para atividade de integração.



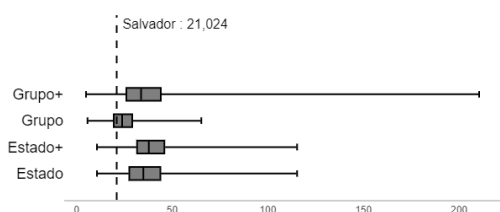
(b) Escolas com biblioteca.



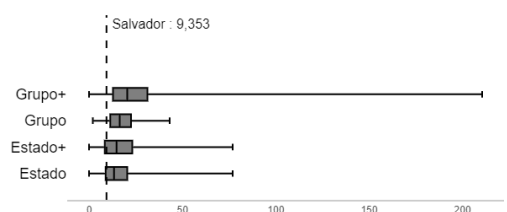
(c) Docentes na educação infantil.



(d) Escolas com material para prática esportiva.



(e) Escolas com internet.



(f) Escolas com internet para uso no ensino.

Figura 3. Situação de Salvador em relação às métricas educacionais de outros municípios do estado e do grupo. As métricas são descritas em quantidade por 1000 estudantes de 4 e 5 anos.

cionais avaliados foram relevantes para agregar pontos de melhoria na capital de análise em relação ao universo educacional e desempenho na meta alvo.

Quando se observa a quantidade de escolas com acesso à internet (Figura 3e) e com o uso de internet para o ensino (Figura 3f), Salvador também se encontra abaixo do primeiro quartil em todos os casos. Por exemplo, enquanto o *Grupo+* e o *Estado+* possuem uma mediana de 20 e 14 escolas com uso da internet para ensino por 1000 estudantes, respectivamente, Salvador possui apenas 9 escolas nesta situação. Assim como foi descrito para o caso de escolas com biblioteca, apesar deste indicador não possuir uma relação direta ou mesmo de causalidade com a taxa de matrícula de crianças na escola,

caberia observar ações dos municípios que se encontram melhor nestas métricas.

Uma métrica diretamente relacionada ao Indicador 1A é a quantidade de docentes atuando na educação infantil (Figura 3c). Salvador também se encontra abaixo dos demais municípios tanto para o estado quanto para o seu grupo, com 0,035 professores para cada 1000 estudantes de 4 a 5 anos, contra a mediana de 3 professores para o estado da Bahia. A diferença é ainda maior ao observar o *Grupo+*, que possui uma mediana de 5 docentes, valor maior que o terceiro quartil das demais séries do gráfico. Logo, pode-se especular que o gestor da educação de Salvador deveria observar com cuidado especial a quantidade de docentes atuando na educação infantil de 4 a 5 anos, a fim de melhorar o Indicador 1A.

O estudo de caso apresentado nesta seção mostra como a criação dos grupos fornece mais uma visão importante para o gestor municipal da educação. Além dos municípios do próprio estado, o gestor passa a poder comparar seu município com outros municípios com características similares. Portanto, é possível afirmar que o método proposto responde à segunda pergunta de pesquisa ao fornecer suporte à decisão multicritério do gestor da educação municipal, uma vez que permite que tal gestor avalie diferentes variáveis antes de priorizar a destinação de recursos.

## 6. Considerações Finais

O presente estudo abordou o desenvolvimento de um método para auxiliar no processo decisório de destinação de recursos para gestores municipais, visando otimizar o desempenho dos municípios nas metas do PNE. A abordagem proposta partiu do agrupamento de municípios, posteriormente refinado para o tratamento de valores discrepantes. Considerando os testes realizados a fim de calcular o número ideal de grupos, o Coeficiente de Silhueta melhorou independentemente da quantidade de grupos utilizada. Considerando o melhor caso, para a criação de 3 grupos, o coeficiente passou de 0,45 para 0,56, implicando em uma melhoria de 0,11, ou aproximadamente 25%. Isto mostra que o método proposto foi capaz de criar grupos mais coesos, reduzindo possíveis vieses decorrentes da comparação entre municípios muito diferentes.

Em seguida, o agrupamento foi utilizado como base para o desenvolvimento de uma visualização para comparar a situação do município com outros grupos de municípios de interesse, com relação a várias métricas educacionais. A análise multicritério foi realizada com variáveis provenientes do censo da educação, que auxiliam na caracterização dos municípios quanto a aspectos relacionados ao cumprimento da meta analisada.

Como resultado da aplicação do método no município de Salvador, foi observado que apesar do município possuir características socioeconômicas que se destacam positivamente em relação aos demais grupos, os seu desempenho relacionado às variáveis ligadas à meta não mostrou o mesmo. Essa tendência pode ser aproveitada como incentivo ao investimento nas áreas prioritárias, como quantidade de docentes na educação infantil, visando à melhoria do desempenho educacional para a meta em questão.

Como trabalhos futuros, pode-se avaliar a utilização de outros indicadores socioeconômicos para a criação dos grupos. Também é possível desenvolver um sistema que permita a visualização das informações e gráficos apresentados, relacionados ao PNE e às métricas educacionais.

## Referências

- Abreu, R., Pitanguí, C., Andrade, A., Assis, L., and Silva, C. (2020). Detecção automática de estilos de aprendizagem por meio de técnicas de clusterização e classificação. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 1022–1031, Porto Alegre, RS, Brasil. SBC.
- Arthur, D. and Vassilvitskii, S. (2006). k-means++: The advantages of careful seeding. Technical Report 2006-13, Stanford InfoLab.
- Bandeira, L. F. and Reyes Junior, E. (2021). Programa minha casa minha vida (2012-2016): análise da eficiência relativa dos municípios brasileiros na execução de programas federais. *Revista de Administração Pública*, 55:309–330.
- Brasil (2014). Lei nº 13.005, de 25 de junho de 2014. aprova o plano nacional de educação - pne e dá outras providências. *Diário Oficial [da] República Federativa do Brasil*.
- Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, 18(6):683–695.
- Gan, G. and Ng, M. K.-P. (2017). k-means clustering with outlier removal. *Pattern Recognition Letters*, 90:8–14.
- Gordia, A. P. e. a. (2010). Variáveis comportamentais e sociodemográficas estão associadas ao domínio psicológico da qualidade de vida de adolescentes. *Chinese Science Bulletin*, 28(1):29–35.
- Greco, S., Figueira, J., and Ehr Gott, M. (2016). *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer.
- Gutiérrez, F., Seipp, K., Ochoa, X., Chiluíza, K., De Laet, T., and Verbert, K. (2020). Lada: A learning analytics dashboard for academic advising. *Computers in Human Behavior*, 107:105826.
- Herodotou, C., Rienties, B., Borowka, A., Zdrahal, Z., and Hlosta, M. (2019). A large-scale implementation of predictive learning analytics in higher education: the teachers' role and perspective. *Educational Technology Research and Development*, 67(5):1273–1306.
- IBGE (2022). Áreas territoriais. url<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html?=&t=o-que-e>.
- Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260.
- MEC (2022). Pne em movimento. url[http://simec.mec.gov.br/pde/grafico\\_pne.php](http://simec.mec.gov.br/pde/grafico_pne.php).
- Montgomery, D. C. and Runger, G. C. (2018). *Applied Statistics and Probability for Engineers*. John Wiley & Sons.
- MP-SC (2022). Relatório de monitoramento do plano municipal de educação. url<https://www.mpsc.mp.br/programas/relatorios-planos-municipais-educacao>.
- PNAD Contínua, I. (2022). Pnad contínua - pesquisa nacional por amostra de domicílios contínua.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Silva, W. D., Pereira, S. S. L., Pereira, D. S., and COSTA, M. (2021). Um sistema baseado em machine learning para apoio à decisão no gerenciamento de produção apícola. *Embrapa Amazônia Oriental-Artigo em periódico indexado (ALICE)*.

TCE-SC (2022). Espaço tce educação. [ur-  
lhttps://servicos.tce.sc.gov.br/tceeducacao/#acompanhamento](https://servicos.tce.sc.gov.br/tceeducacao/#acompanhamento).

Todos pela Educação (2022). Observatório do plano nacional de educação. [ur-  
lhttps://www.observatoriodopne.org.br](https://www.observatoriodopne.org.br).

Zapparolli, L., Stiubiener, I., Braga, J., and Pimentel, E. (2017). Aplicando técnicas de business intelligence e learning analytics em ambientes virtuais de aprendizagem. *Simpósio Brasileiro de Informática na Educação*, 28(1):536–545.