

Avaliação da Eficiência de Sistemas Educacionais Aplicando Técnicas de Agrupamento e DEA

Ismayle S. Santos¹, Rafael L. Gomes², William S. Vilaça³

¹Centro de Ciências e Tecnologia – Universidade Estadual do Ceará (UECE)
Caixa Postal 60714-903 – Itaperi, Fortaleza – CE – Brasil

²Centro de Ciências e Tecnologia – Universidade Estadual do Ceará (UECE)
Caixa Postal 60714-903 – Itaperi, Fortaleza – CE – Brasil

³Centro de Ciências e Tecnologia – Universidade Estadual do Ceará (UECE)
Caixa Postal 60714-903 – Itaperi, Fortaleza – CE – Brasil

ismayle.santos@uece.br, rafa.lobes@uece.br, william.vilaca@aluno.uece.br

Abstract. *National and international research in the area of education has shown a dramatic reality, where a large number of students who graduate from public high schools do not master basic knowledge of mathematics and the Portuguese language. In this scenario, it is necessary to develop solutions that support educational environments, as well as help improve the management of the teaching and learning process. Within this context, this article presents a solution that applies Artificial Intelligence (AI) and Data Envelopment Analysis techniques to develop models that, based on educational data, analyze the performance of teaching units, grouping them, in order to indicate actions so that inefficient units can have their teaching and learning process evolved from the characteristics of groups with better performance. The results, using real data, show that the proposed model can approximate schools with similar realities, but that it has achieved different results in external learning assessments.*

Resumo. *Pesquisas nacionais e internacionais na área da educação têm apresentado uma realidade dramática, onde grande parte dos alunos do ensino médio público não têm domínio sobre conhecimentos básicos da matemática e da língua portuguesa. Nesse cenário, faz-se necessário desenvolver soluções que deem suporte aos ambientes educacionais, bem como auxiliem na gestão do processo de ensino e aprendizagem. Dentro desse contexto, este artigo apresenta uma solução que aplica técnicas de Inteligência Artificial (IA) e Análise por Envoltória de Dados para desenvolver modelos que, a partir de dados educacionais, façam uma análise do desempenho das unidades de ensino, as agrupando, a fim de indicar ações para que unidades ineficientes possam ter seu processo de ensino e aprendizagem melhorado a partir das características de grupos com melhor desempenho. Os resultados, usando dados reais, mostram que a solução proposta pode aproximar escolas com realidades similares, mas que têm alcançado resultados distintos nas avaliações externas de aprendizagem.*

1. Introdução

O Sistema de Avaliação da Educação Básica (Saeb) é um processo de avaliação somativa em larga escala, realizado periodicamente pelo Inep¹, que visa oferecer subsídios para a elaboração, o monitoramento e o aprimoramento de políticas educacionais, permitindo a avaliação da qualidade da educação praticada no país. Por meio de testes e questionários, o Saeb reflete os níveis de aprendizagem demonstrados pelo conjunto de estudantes avaliados. Estudos recentes ² indicam que apenas 5% dos alunos da rede pública terminam o ensino médio com conhecimentos adequados de matemática, e vários autores [Soares 2022, Veríssimo 2021, Alves et al. 2019, Malta Campos and Vieira 2021] destacam que o cenário já era alarmante antes da pandemia.

Nesse cenário, diversas políticas e projetos têm sido estabelecidos com foco na melhoria da aprendizagem. No Ceará, por exemplo, o programa Ceará educa mais³ estabelece 25 ações que visam a melhoria contínua da aprendizagem no estado. O projeto Jovem de Futuro do Instituto Unibanco, por sua vez, em parceria com as Secretarias Estaduais de Educação, disponibiliza para as escolas uma metodologia e instrumentos que dão suporte ao trabalho de gestão. No escopo destas ações são realizadas reuniões de boas práticas entre escolas, onde escolas modelos, com resultados exitosos, apresentam suas metodologias para escolas com baixos índices educacionais. Contudo, este processo não considera a realidade contextual das escolas o que pode causar dificuldade na implementação das ações. Nesta perspectiva, o uso de Inteligência Artificial (IA), para extração de informações das escolas, pode ser um meio para melhor contextualizar cada escola e informar como as ações de compartilhamento de metodologias entre as escolas poderiam ser melhor realizadas [Guan et al. 2020, Chang and Wang 2020].

O aprendizado de máquina, subcategoria da IA por exemplo, acelera o processo de tomada de decisão, melhorando as previsões com base nos padrões apresentados nos dados. Tais padrões que, na maior parte das vezes, são invisíveis à análise cognitiva humana devido a incapacidade de processamento de grande massa de dados ou, em alguns casos, devido a dificuldade em perceber os padrões existentes [Guan et al. 2020].

Dentro desse contexto, este artigo apresenta uma metodologia para aperfeiçoar a política de parceria entre escolas públicas de ensino médio do estado do Ceará, promovendo a aproximação entre escolas com realidades contextuais semelhantes, mas que não alcançam os mesmos índices educacionais. Para definir a metodologia, foi utilizado a base de dados do Saeb de 2019 focando nas proficiências de matemática, a área mais deficitária entre as 3 áreas avaliadas no Saeb.

A utilização da técnica não supervisionada de clusterização para gerar grupos de escolas com características similares, seguida da Análise por Envoltória de Dados (DEA) aplicada a cada grupo de escola, estabelecendo uma fronteira de eficiência no grupo e indicando que parâmetros cada escola deve alterar para chegar a fronteira de eficiência, fazendo uso da proficiência, das características estruturais, metodológicas e pedagógicas, podem estabelecer um caminho de melhoria contínua para o sistema educacional Cearense e por conseguinte para outros sistemas similares. O desafio nesta proposta foi então identificar as escolas com menor eficiência no ensino de matemática e possibilitar aos

¹<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb>

²<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa/resultados>

³<https://www.seduc.ce.gov.br/ceara-educa-mais>

gestores uma análise dos dados direcionada, definindo parcerias de escolas ineficientes com escolas eficientes para o crescimento da aprendizagem em matemática.

O restante desse artigo está organizado da seguinte forma. A Seção 2 detalha as soluções existentes e estado da arte no que se refere a soluções de IA em ambientes educacionais. A Seção 3 descreve a solução proposta, enquanto a Seção 4 discute os resultados obtidos. Finalmente, a Seção 5 conclui o artigo e apresenta trabalhos futuros.

2. Estado da Arte

2.1. Trabalhos Relacionados

Esta seção irá destacar trabalhos acadêmicos e pesquisas relevantes que exploraram a otimização dos sistemas educacionais públicos fazendo uso da inteligência artificial ou análise envoltória de dados. Além disso, será feita uma análise das abordagens existentes, identificando suas vantagens e limitações.

Em [Rassouli-Currier 2007] foi testado a eficiência dos 354 distritos escolares de Oklahoma usando duas especificações com a DEA. Neste estudo, os autores utilizaram variáveis que podem ser controladas, como exemplo o salário dos professores e variáveis para as quais não se tem controle, tais como alunos em educação especial e taxa de pobreza no distrito. Os autores também realizaram a comparação dos modelos produzidos no trabalho e concluíram que os principais fatores que afetam as medidas de eficiência para estes distritos são principalmente as características dos alunos e o ambiente familiar.

Na referência [Miranda and Miranda 2018], os autores utilizam a análise envoltória de dados (DEA) visando apresentar e validar um indicador derivado de variáveis que compõem o fluxo de alunos dentro da escola, índice LOED. A partir da idealização de pesos ótimos que capturam a contribuição dessas variáveis para o indicador, a análise com DEA pode aferir as medidas de eficiência das escolas da rede pública de ensino da cidade de Campinas, próximos ao indicador analisado, produzindo parâmetros para avaliação da rede.

[Pereira et al. 2020] investigaram a aprendizagem em Matemática e Língua Portuguesa dos alunos do 2º, 5º e 9º ano nas escolas do estado do Ceará. A partir dos dados do Spaece (Sistema Permanente de Avaliação da Educação Básica do Ceará)⁴ de 2014, os autores realizaram uma análise comparativa entre o método proposto no trabalho, agrupamento através do algoritmo de aprendizagem não supervisionada Kmeans e o método utilizado no Spaece que faz uso da localização geográfica. Os autores tinham como objetivo aumentar a fidelidade dos resultados às reais taxas de desempenho estudantil, considerando suas fragilidades e potencialidades e eles concluíram que o método de agrupamento alcançou maior precisão e portanto deve ser considerado para fomento de políticas públicas.

Assim como no trabalho de [Pereira et al. 2020], este artigo tinha como objetivo explorar um melhor agrupamento para as escolas do ensino médio com base em número maior de características. Contudo, diferentemente dos trabalhos existentes, na proposta desse artigo foi realizado o agrupamento das escolas utilizando as variáveis qualitativas presentes no questionário do Saeb. Estas variáveis esboçam a estrutura, os atores envolvidos e as metodologias empregadas nos ambientes educacionais. A partir disso, para

⁴<https://www.seduc.ce.gov.br/spaece/>

Tabela 1. TRABALHOS RELACIONADOS / TRABALHO PROPOSTO

	Método	Tipo De Dados	Objetivos/Resultados
Rassouli-currier, 2007	DEA	Quantitativos	Definição das principais características que impactam na eficiência
Miranda; Miranda, 2018	DEA	Quantitativos	Índice LOED validado/ definição de parâmetros para alcance da eficiência
Pereira et al, 2020	Kmeans	Quantitativos	Método com maior precisão na separação pela proficiência
Trabalho Proposto	Kmeans+DEA	Categoricos/ Quantitativos	Agrupamento de unidades similares e parâmetros para melhoria da eficiência

cada um dos grupos de escolas formados, a análise DEA é empregada para mensurar a distribuição das eficiências de cada unidade tomadora de decisão, (i.e., uma escola), definindo parcerias entre elas para subsidiar a melhoria dos indicadores de aprendizagem. Assim, esta proposta inova no que se refere à capacidade analisar e correlacionar os ambientes educacionais a partir de modelos de IA.

2.2. Análise por Envoltória de Dados

Para avaliar a eficiência em cada grupo de escolas, estabelecendo ações para melhoria de escolas ineficientes, neste artigo é utilizado a Análise por Envoltória de Dados (DEA), ferramenta que deve ser usada para avaliar o desempenho de unidades tomadoras de decisão [Miranda and Rodrigues 2010]. A partir dessa análise é possível aumentar a eficiência das unidades através da diminuição de recursos (inputs) mantendo-se o resultado produtivo (outputs) ou o aumento do outputs mantendo-se os inputs inalterados. Este método determina, ainda, a melhor combinação de pesos que maximize a combinação linear dos inputs e outputs, sujeita às restrições convenientes. Desse modo, este estudo propõe um modelo que possibilita o aumento da eficiência das escolas da Seduc-CE (Secretaria de Educação do Estado do Ceará) no papel de unidades tomadoras de decisão (DMUs).

A partir do método de Análise por Envoltória de Dados é possível comparar a eficiência em grupo de unidades tomadoras de decisão (DMUs), analisando os insumos necessários no processo e os produtos gerados, revelando a eficiência de cada unidade analisada e estabelecendo um indicador de avaliação [Chang and Wang 2020]. O método também pode ser modelado para diminuição dos insumos ou aumento da produção, sendo possível a aplicação em qualquer segmento que atenda a estas características.

Duas abordagens podem ser utilizadas nesta análise, o modelo CCR, desenvolvido por [Charnes et al. 1978], desenhado para uma análise com retornos constantes de escalas, posteriormente, o modelo BCC, proposto por [Banker et al. 1984], que não exige aumento proporcional entre recursos e resultados, mantendo rendimentos crescentes ou decrescentes na fronteira de eficiência. O modelo CCR será usado neste artigo em cada grupo definido da etapa anterior, tendo como meta aumentar os resultados (output), sem alterar os recursos (input).

3. Proposta

Este trabalho propõe a utilização de um modelo de aprendizado não supervisionado, K-Means, para gerar uma organização conveniente e válida das escolas de ensino médio do Ceará a fim de projetar grupos que se tornem unidades gerenciáveis. Além disso, a abordagem inclui a aplicação de DEA nestes grupos para propor parceria entre as escolas com foco no aumento da eficiência das unidades tomadoras de decisão. As etapas do algoritmo implementado, que tem como entrada de dados a base de dados do Saeb, e como saída as escolas semelhantes agrupadas são apresentadas na Figura 1.

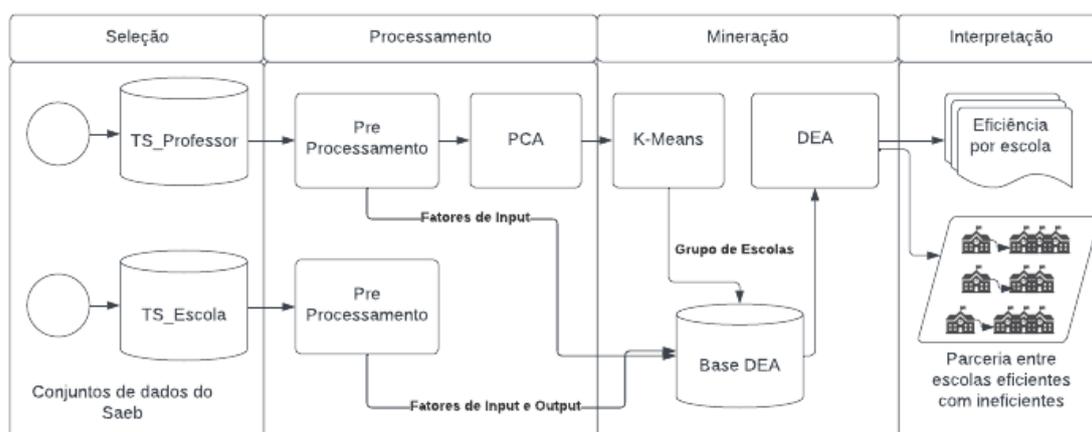


Figura 1. Modelo proposto

Para a formação dos grupos foi utilizada, na etapa de Seleção, a parte da base de dados do Saeb 2019 que é composta por atributos qualitativos/categóricos representando as respostas dos professores nas escolas cearenses às questões da avaliação. Estas questões tratam, por exemplo, da experiência do professor, da estrutura da sala de aula, dos recursos pedagógicos utilizados, das características das turmas avaliadas e da metodologia utilizada pelo professor. Na etapa de Processamento foi necessário transformar estas variáveis categóricas em quantitativas para implementação no algoritmo Kmeans utilizando as técnicas: (i) *Substituição simples*, substituindo o carácter por um decimal adequado a informação contida no dado, (ii) *Ordinal Enconding*, relacionando a cada classe um número de 0 até $N_{Classe} - 1$ e (iii) *One Hot Enconding*, transformando uma coluna com N rótulos em N colunas com o valor 0 ou 1, onde 1 representa a ocorrência do rótulo.

Na etapa de Mineração, o algoritmo do modelo K-Means utilizado difere-se da aplicação convencional na escolha dos centróides, pois seleciona aleatoriamente o primeiro centróide dos pontos de dados, mas seleciona os próximos centróide dos pontos de dados de forma que a probabilidade de escolher um ponto como centróide seja diretamente proporcional à sua distância do centróide mais próximo, previamente escolhido.

Para avaliação da qualidade do agrupamento utilizamos dois métodos, que são utilizados amplamente no contexto de clusterização, o índice Calinski Harabasz (CH)[Wang and Xu 2019] e o método do cotovelo [Gustriansyah et al. 2020]. O primeiro considera a maximização da variação entre os grupos e a minimização da variação dentro dos grupos, quanto maior a pontuação CH, melhor será o resultado do agrupamento. O

segundo busca o equilíbrio em que as observações que formam cada agrupamento sejam o mais homogêneas possível e que os agrupamentos formados sejam o mais diferentes um dos outros, uma vez que aumentando a quantidade de clusters no K-Means as diferenças entre clusters se tornam muito pequenas, e as diferenças das observações intra-clusters vão aumentando.

Após a construção da base com os grupos de escolas semelhantes, ainda na etapa de Mineração, foi selecionado um grupo aleatório para realizar a Análise por Envoltória de Dados, a fim de determinar as eficiências relativas das escolas, identificar as metas para melhoria da eficiência e encontrar a relação de escolas parceiras que poderão contribuir para o alcance das metas estabelecidas. A base de dados para a análise com DEA foi composta pelo grupo de escolas definidos na etapa anterior, com os seguintes fatores para cada DMU (i.e., escola do grupo selecionado):

- Quantidade de professores por vínculo trabalhista: concursado, contrato temporário, contrato clt e etc;
- Indicador percentual de adequação de formação docente: apresenta o percentual de disciplinas que são ministradas, por professores com formação superior de Licenciatura na mesma disciplina que leciona;
- Um dos dois tipos possíveis de escola: ensino médio regular e ensino médio integrado.
- Nível socioeconômico: calculado a partir do nível de escolaridade dos pais e da posse de bens e contratação de serviços pela família dos alunos.
- Número de alunos matriculados na escola;
- Número de presentes na avaliação;
- Percentual de participação;
- Média dos alunos em Matemática.
- Percentual de alunos no grupo de níveis iniciais (0 ao 4);
- Percentual de alunos no grupo de níveis finais (5 ao 10);

Desses fatores, os dois *outputs* são o Percentual de alunos no grupo de níveis iniciais de proficiência e o Percentual de alunos no grupo de níveis finais. Para o primeiro foi utilizado, na base de dados, o inverso do percentual real, assim será garantida a diminuição desse grupo à medida que o valor apresentado aumenta.

4. Resultados e Discussões

A proposta desse artigo foi implementada utilizando a linguagem Python em um Sistema Operacional Linux Mint, com processador Intel Core i5 2.3GHZ, e memória RAM de 8GB. A principal plataforma de software em ambiente virtual é o Jupyter Notebook versão 6.4.12. O conjunto de dados utilizado registra todo o contexto de aplicação do Sistema de Avaliação da Educação Básica - Saeb. Foram utilizados 2 conjuntos de dados dos 9 disponíveis, TS_Professor e o TS_Escola. O primeiro descreve majoritariamente a visão do professor sobre o ambiente escolar no qual ele está inserido; o segundo contempla o resultado das proficiências alcançadas e as características da aplicação para cada uma das escolas participantes.

Cada registro no primeiro conjunto de dados contém 140 atributos, dos quais, 75 foram utilizados neste trabalho, já no segundo conjunto cada registro contém 136 atributos, dos quais, 20 foram utilizados. Atributos irrelevantes para este estudo como: ID_Saeb,

ID_Região, ID_Turma, Co_Professor, e atributos sem dados registrados, foram descartados. As primeiras perguntas do questionário que gera a base de dados TS_Escola são expostas na Tabela 2. No campo 'Número' está o número da questão aplicada. No campo 'Questionamento - Opções de resposta', consta o enunciado da questão e as respostas possíveis.

Tabela 2. Exemplo de questão aplicada aos professores no questionário do Saeb

Nº	Enunciado	Opções
	Avalie as condições da sala de aula que você utiliza nesta escola com relação aos seguintes elementos:	(A)Inadequado; (B)Pouco Adequado; (C) Razoavelmente Adequado; (D) Adequado.
23	Tamanho da sala com relação ao número de alunos	
24	Acústica	
25	Iluminação Natural	
26	Ventilação Natural	
27	Temperatura	
28	Limpeza	
29	Acessibilidade	
30	Mobiliário (Mesas, Cadeiras)	

Devido ao excesso de features, com foco em acelerar o aprendizado do K-means, foi utilizado a PCA na etapa de Processamento para redução da dimensionalidade do conjunto de dados, fazendo uso da quantidade de componentes que explicam 90% da variância total dos dados . Na Figura 2 é possível observar que a avaliação CH é sempre melhor com o uso da PCA para qualquer quantidade inteira n de clusters com $2 \leq n \leq 20$.

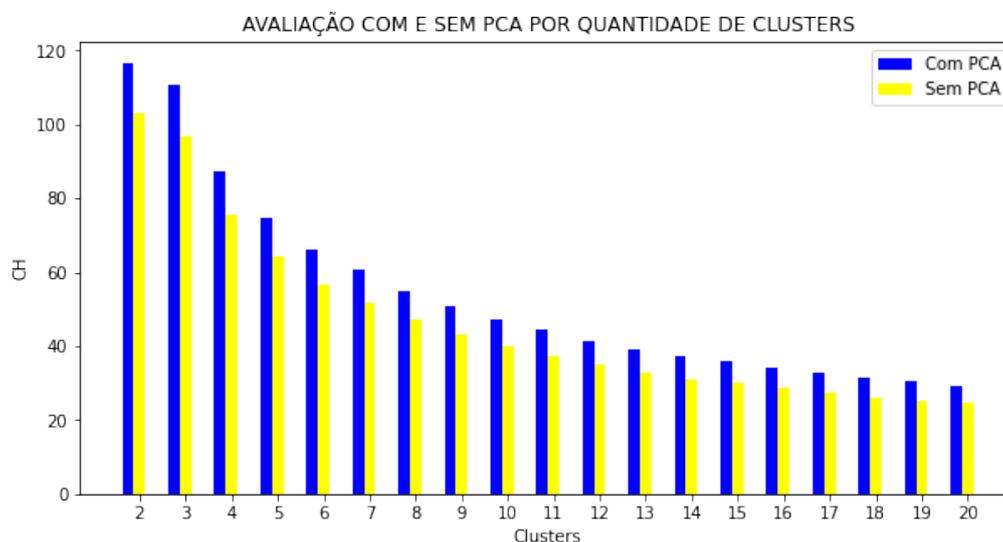


Figura 2. Avaliação CH com e sem PCA

Foi executado o modelo K-Means++, na etapa de Mineração, com diferentes quantidades de clusters, começando com 2 clusters e incrementando de 1 em 1 até atingir 20 clusters. Em seguida foi feita uma avaliação usando as métricas CH e o método do cotovelo. Estes métodos indicaram que sete grupos era a quantidade ideal. Foi então adicionado o rótulo do grupo aos exemplos da base de dados. A Figura 3 apresenta uma simplificação em 2 dimensões dos clusters com os centróides destacados em vermelho formados.

Para a base que foi submetida a DEA, foi utilizado um dos grupos formados acrescido de parte dos atributos do conjunto TS_Escola. Apenas atributos que fazem referência às escolas do ensino médio regular e integral foram utilizados (Tabela 3).

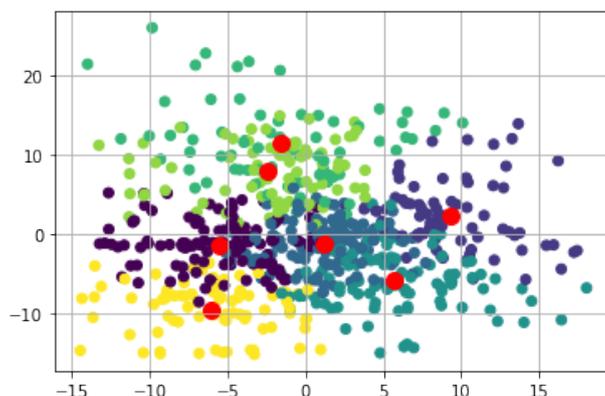


Figura 3. Clusters formados com os centróides em destaque

Tabela 3. Exemplo de dados da base de dados submetida a DEA

ID_Escola	Efet	Temp	Terc	%Form	Nível	Matric	Presen	%Partic	Média
61251475	0	1	4	74	3	149	145	0,97	309,07
61251482	3	2	6	33,9	2	183	184	1,01	264,59
61251810	3	0	3	47,6	3	84	84	1,00	287,03
61251818	3	3	0	39,7	2	119	117	0,98	291,7
61251819	3	1	6	52,3	2	180	178	0,99	282,27
61251828	1	5	6	72,3	3	349	339	0,97	284,19
61254116	1	6	2	81,7	3	270	248	0,92	272,11
61254427	0	3	5	41,8	3	167	164	0,98	261,98
61254428	0	1	2	58,6	3	91	90	0,99	283,85

Na Tabela 4 os atributos nomeados com prefixo 'Nível' exibem o nível de proficiência dos alunos do ensino médio na edição de 2019 do Saeb. Agrupados formam os fatores de Outputs na DEA, destacando que para o NIVEL_INICIAL, nível 0 ao 4, foi utilizado o inverso da quantidade real, para que quando o valor for aumentando na DEA a quantidade real diminua. Foi utilizada a análise CCR orientada a output com objetivo de conseguir migrar os alunos do NIVEL_INICIAL para o NIVEL_FINAL, nível 5 ao 10. Os demais atributos junto com o atributo 'tipo de vínculo' do conjunto de dados TS_Professor formam os fatores de Input.

Na Tabela 5, é possível observar parte das 131 escolas pertencentes ao grupo selecionado para a etapa de Mineração com DEA. Dentre elas, na etapa de Interpretação, vemos que 18 escolas figuraram com 100% de eficiência após a análise, e serão escolas 'modelo' para as escolas ineficientes poderem melhorar seus indicadores, direcionando as ações a serem tomadas. A comunicação entre gestores e professores com foco no compartilhamento de boas práticas irá abrir espaço para o alcance das metas estabelecidas pela DEA. Também são destacadas na Tabela 5 as 18 DMUs mais ineficientes (abaixo dos 60%), seguido dos seus valores, reais e projeções em percentual, para os Outputs Nível_Inicial e Nível_Final. Na imagem foi feito o ajuste do output Nível_inicial, conforme mencionado anteriormente, para a DEA, foi utilizado o inverso do valor real.

Tabela 4. Parte dos atributos do conjunto de dados TS_Escola

Variável	Tipo	Tamanho	Descrição
FORMACAO DOCENTE	Num	5	Indicador de Adequação da Formação Docente (Informação referente ao Grupo 1, para o Ensino Médio)
NIVEL SOCIO ECONOMICO	Char	9	Indicador de Nível Socioeconômico (Inse): Nível_I Nível_II Nível_III Nível_IV Nível_V Nível_VI Nível_VII
MATRICULADOS	Num	3	Número de alunos matriculados na 3ª/4ª série do ensino médio tradicional no censo 2019
PRESENTES	Num	3	Número de alunos presentes na aplicação
PARTICIPACAO	Num	5	Razão entre o total de alunos presentes no Saeb e o total de alunos cadastrados no Censo Escolar que são público alvo do Saeb
Nível 0	Num	5	(-00;225)
Nível 1	Num	5	[225;250)
Nível 2	Num	5	[250;275)
Nível 3	Num	5	[275;300)
Nível 4	Num	5	[300;325)
Nível 5	Num	5	[325;350)
Nível 6	Num	5	[350;375)
Nível 7	Num	5	[375;400)
Nível 8	Num	5	[400;425)
Nível 9	Num	5	[425;450)
Nível 10	Num	5	[450;+00)

Tabela 5. Eficiência de parte das Escolas gerado na DEA

DMUs Eficientes	%	DMUs Ineficientes	%
61320723	100,00%	61288931	59,72%
61251818	100,00%	61307764	59,39%
61251819	100,00%	61280084	59,21%
61276884	100,00%	61306401	58,97%
61322982	100,00%	61297796	58,95%
61317697	100,00%	61280224	58,85%
61272732	100,00%	61320047	58,71%
61286562	100,00%	61255598	58,40%
61251810	100,00%	61287445	58,38%
61321868	100,00%	61276885	58,34%
61303448	100,00%	61291168	58,17%
61271146	100,00%	61287443	58,11%
61314257	100,00%	61257028	57,74%
61301897	100,00%	61301907	57,32%
61320036	100,00%	61307997	57,29%
61322988	100,00%	61254116	57,12%
61312874	100,00%	61258837	56,54%
61307765	100,00%	61308002	54,93%

Através da colaboração das escolas parcerias apresentadas na tabela 6, espera-se diminuir o grupo de alunos no NIVEL_INICIAL em 42% aproximadamente, esses alunos devem ser transportados para o grupo de alunos no NIVEL_FINAL acarretando em um aumento de 5 vezes, em média, a quantidade atual de alunos no grupo. A Escola de código 61288931 por exemplo, por meio das parcerias com as escolas 61276884, 61303448 e 61321868 conseguirá reduzir o percentual de alunos no grupo NIVEL_INICIAL de 93,75% para 55,99%, e aumentará o percentual de alunos no grupo NIVEL_FINAL de

Tabela 6. Resultados das 18 DMUs Ineficientes com as escolas parceiras.

	DMU	NIVEL INICIAL		NIVEL FINAL		ESCOLAS PARCEIRAS
		REAL	PROJEÇÃO	REAL	PROJEÇÃO	
1	61288931	93,75	55,99	6,24	43,73	61276884, 61303448, 61321868.
2	61307764	95,52	56,73	4,48	43,45	61303448, 61320723.
3	61280084	85,71	50,75	14,29	48,58	61276884, 61303448, 61321868.
4	61306401	95,94	56,57	4,07	43,59	61303448, 61320723.
5	61297796	87,46	51,56	12,54	47,39	61276884, 61303448, 61321868.
6	61280224	87,06	51,23	12,94	47,11	61276884, 61321868.
7	61320047	93,96	55,16	6,05	43,64	61276884, 61303448, 61321868.
8	61255598	93,81	54,78	6,2	43,86	61276884, 61303448, 61321868.
9	61287445	87,53	51,10	12,47	47,92	61276884, 61303448, 61321868.
10	61276885	92,58	54,02	7,42	44,77	61276884, 61303448, 61321868.
11	61291168	95,53	55,57	4,47	43,14	61276884, 61321868.
12	61287443	91,93	53,42	8,06	45,38	61276884, 61303448, 61321868.
13	61257028	90,60	52,31	9,41	46,06	61276884, 61321868.
14	61301907	91,20	52,27	8,81	46,57	61276884, 61303448, 61321868.
15	61307997	93,70	53,68	6,31	45,12	61276884, 61303448, 61321868.
16	61254116	90,45	51,67	9,55	47,25	61276884, 61303448, 61321868.
17	61258837	91,24	51,59	8,77	46,77	61276884, 61321868.
18	61308002	89,33	49,07	10,67	48,44	61276884, 61303448, 61321868.

6,24% para 43,73%.

As trocas de experiências, de didáticas, os comparativos de estrutura física, de metodologias e de ações de gestão são essenciais para o alcance do objetivo apresentado para os outputs na DEA. A Tabela 6 apresenta pelo menos duas escolas parceiras para todas as escolas com eficiência abaixo de 60%. No grupo de 113 escolas ineficientes, algumas contaram com apenas uma escola parceira. A escola de código 61307764, por exemplo, terá como escolas parceiras as escolas eficientes de código: 61303448 e 61320723.

5. Conclusão

A partir dos resultados apresentados, percebe-se que o agrupamento através do modelo não supervisionado K-Mean, utilizando os dados contextuais do Saeb pode aproximar escolas com realidades similares, mas que têm alcançado resultados distintos nas avaliações externas de aprendizagem. A Análise por envoltória de dados pode ajudar as escolas ineficientes em parceria com escolas eficientes a criar um plano de ação para melhoria da eficiência e dos resultados educacionais nas diversas avaliações realizadas. Espera-se, portanto, que o modelo proposto neste artigo possa otimizar a definição de parcerias entre as escolas de ensino médio do estado do Ceará.

Nesse contexto, no qual os estudantes não têm desenvolvido a aprendizagem adequada ao final da etapa de ensino básico, é crucial desenvolver soluções que ofereçam apoio aos ambientes educacionais e auxiliem no aprimoramento da gestão do processo de ensino e aprendizagem. Neste contexto, este trabalho apresenta uma solução inteligente que visa iluminar a realidade, fornecendo elementos de juízo para interpretá-la, trazendo dados que favoreçam a reflexão de todos os atores da comunidade escolar, ajudando a compreender a complexidade das instituições e dos processos educacionais, criando condições para as intervenções necessárias. Como trabalhos futuros, pretende-se expandir os dados utilizados no modelo de IA desenvolvido.

Referências

- Alves, A. C., Fischer, B., Schaeffer, P. R., and Queiroz, S. (2019). Determinants of student entrepreneurship: An assessment on higher education institutions in brazil. *Innovation & Management Review*, 16(2):96–117.
- Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimation technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9):1078–1092.
- Chang, X. and Wang, X. (2020). Research performance evaluation of university based on super dea model. In *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, volume 9, pages 1252–1255. IEEE.
- Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring efficiency of decision making units. *European Journal of Operational Research*, 1:429–444.
- Guan, C., Mou, J., and Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies*, 4(4):134–147.
- Gustriansyah, R., Suhandi, N., and Antony, F. (2020). Clustering optimization in rfm analysis based on k-means. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(1):470–477.
- Malta Campos, M. and Vieira, L. F. (2021). Covid-19 and early childhood in brazil: Impacts on children's well-being, education and care. *European Early Childhood Education Research Journal*, 29(1):125–140.
- Miranda, A. C. and Miranda, E. C. M. (2018). Alternative methodology in the elaboration of indicators to evaluate schools. *Pro-Posições*, 29(3):207.
- Miranda, A. C. and Rodrigues, S. C. (2010). O uso da dea como ferramenta alternativa da gestão escolar na avaliação institucional. *Educação: Teoria e Prática*, 20(35):163–180.
- Pereira, V. R. F., de Paula, A. D., and Araújo, C. O. (2020). Método de agrupamento aplicado à avaliação escolar: um estudo de caso para avaliações de larga escala. *EDUCA-Revista Multidisciplinar em Educação*, 7(17):901–919.
- Rassouli-Currier, S. (2007). Assessing the efficiency of oklahoma public schools: a data envelopment analysis. *Southwestern Economic Review*, 34:131–144.
- Soares, T. S. S. (2022). O sistema de avaliação da educação básica (saeb) em tempos de pandemia: ensino de matemática e as tecnologias digitais. *Com a Palavra, o Professor*, 7(19):95–106.
- Veríssimo, T. E. d. O. (2021). O sistema de avaliação da educação básica (saeb) e a qualidade do ensino de matemática. B.S. thesis.
- Wang, X. and Xu, Y. (2019). An improved index for clustering validation based on silhouette index and calinski-harabasz index. In *IOP Conference Series: Materials Science and Engineering*, volume 569, page 052024. IOP Publishing.