

DESAFIOS E OPORTUNIDADES DO ENSINO EM COMPUTAÇÃO: UMA ANÁLISE EXPLORATÓRIA E MULTIVARIADA DOS DADOS DO ENADE

CHALLENGES AND OPPORTUNITIES OF TEACHING IN COMPUTING: AN EXPLORATORY AND MULTIVARIATE ANALYSIS OF ENADE DATA

DESAFÍOS Y OPORTUNIDADES DE LA ENSEÑANZA EN INFORMÁTICA: UN ANÁLISIS EXPLORATORIO Y MULTIVARIADO DE DATOS DEL ENADE

Josicleyton Azevedo dos Santos

Mestrando em Ciência da Computação, Universidade Federal Rural do Semiárido, Mossoró, Rio Grande do Norte, Brasil, <https://orcid.org/0009-0009-6895-2714>
e-mail: santos.josicleyton@gmail.com

Macilene Maria Monteiro Maia

Mestranda em Administração, Universidade Federal Rural do Semiárido, Mossoró, Rio Grande do Norte, Brasil.
<https://orcid.org/0000-0003-2127-6711>
E-mail: macilene.maria@hotmail.com

Marcos Filho Lima Bastos

Mestrando em Administração, Universidade Federal Rural do Semiárido, Mossoró, Rio Grande do Norte, Brasil.
<https://orcid.org/0009-0008-9306-157X>
E-mail: marcosfbastos1995@gmail.com

Mileno Alexandre Barbosa Epifanio

Bacharel em Engenharia de Produção, Universidade Federal Rural do Semiárido, Mossoró, Rio Grande do Norte, Brasil.
<https://orcid.org/0000-0003-4995-6866>
E-mail: mileno.alexandre86@gmail.com

Fábio Francisco da Costa Fontes

Doutor em Engenharia de Computação e Automação pela Université d'Artois, Universidade Federal Rural do Semiárido, Mossoró, Rio Grande do Norte, Brasil.
<https://orcid.org/0000-0002-5070-3578>
E-mail: fabio_fontes@ufersa.edu.br

RESUMO

Este estudo analisa o desempenho dos cursos de Computação e Tecnologias da Informação e Comunicação no Exame Nacional de Desempenho dos Estudantes (ENADE) de 2021 no Brasil. Utilizando uma abordagem baseada no processo de mineração de dados, foram aplicadas diversas técnicas para extrair *insights* sobre as características e padrões destes cursos e instituições de ensino superior. Os resultados revelam uma distribuição heterogênea no desempenho dos cursos, com variações significativas nos níveis médios de conceito ENADE. Além disso, variáveis como a natureza administrativa das instituições, a organização acadêmica, o tipo de modalidade educacional e a localização geográfica, demonstraram influenciar o desempenho dos cursos. A análise multivariada identificou variáveis significativas na explicação da variação do conceito ENADE, enquanto a análise de clusterização revelou grupos distintos de cursos com características semelhantes. Estes resultados oferecem uma base sólida para a melhoria contínua da

qualidade dos cursos de computação no Brasil, fornecendo orientações precisas para as instituições de ensino, empregadores e formuladores de políticas educacionais.

Palavras-chave: ENADE, Ensino Superior, Mineração de Dados Educacionais, Tecnologias da Informação e Comunicação.

ABSTRACT

This study explores the performance of Computer Science and Information and Communication Technology undergraduate programs in the 2021 National Student Performance Examination (ENADE) in Brazil. Employing a methodology grounded in the data mining process, a variety of techniques were utilized to uncover insights regarding the features and trends of these courses and higher education institutions. The results indicate a heterogeneous performance distribution among the programs, marked by substantial variations in the average ENADE ratings. Furthermore, factors such as the administrative nature of the institutions, their academic structure, the educational modality employed, and their geographical locations were found to significantly impact program performance. Through multivariate analysis, key variables contributing to the variations in ENADE scores were identified, while cluster analysis distinguished distinct groups of programs sharing similar attributes. These findings lay a robust foundation for the ongoing enhancement of computer science education in Brazil, providing targeted recommendations for educational institutions, employers, and educational policy makers.

Keywords: ENADE, Higher Education, Educational Data Mining, Information and Communication Technologies.

RESUMEN

Este estudio examina el rendimiento de los programas de grado em Informática y Tecnologías de la Información y Comunicación en el Examen Nacional de Desempeño de Estudiantes (ENADE) de 2021 en Brasil. Mediante un enfoque basado en la minería de datos, se emplearon diversas técnicas para extraer perspectivas acerca de las características y tendencias de estos programas e instituciones de educación superior. Los hallazgos revelan una distribución heterogénea en el rendimiento de los programas, con variaciones significativas en los niveles medios del índice ENADE. Factores como el régimen administrativo de las instituciones, su estructura académica, el tipo de modalidad educativa y la localización geográfica resultaron ser determinantes en el rendimiento de los programas. El análisis multivariante permitió identificar las variables significativas que explican la variación en los índices ENADE, mientras que el análisis de agrupamiento descubrió grupos diferenciados de programas con características afines. Estos resultados proporcionan una base firme para el mejoramiento continuo de la calidad de los programas de informática en Brasil, ofreciendo directrices precisas para las instituciones educativas, los empleadores y los responsables de la formulación de políticas educativas.

Palabras-clave: ENADE, Educación Superior, Minería de Datos Educativos, Tecnologías de la Información y Comunicación.

INTRODUÇÃO

A produção, armazenamento e circulação de informações constituem, historicamente, aspectos centrais das sociedades. A evolução e ascensão da importância e impacto da área de Tecnologia da Informação (TI) é produto do esforço contínuo na promoção de inovações, que possuem a finalidade de transmitir e aperfeiçoar ideias e criações ao longo do tempo e espaço (Rocha; Jucá; Silva, 2019). Essa área científica tem influência em diversos setores econômicos e sociais, incluindo governança, operações

monetárias, comunicação, turismo, logística e questões de desigualdade econômica (Franklin *et al.*, 2022; Kuss; Medaglia, 2023; Luciano; Fantinel; Lubeck, 2021; Merlotto; Moori; Lopes, 2021; Pereira Junior, 2020; Soares *et al.*, 2023).

A TI e seus segmentos fronteiriços e derivados compõem o macrocampo de cursos de ensino superior denominado Computação e tecnologias da informação e comunicação (TIC) (MEC, 2024). Diante da necessidade constante de análise do desempenho da formação dos profissionais da área, em razão do forte movimento de expansão e diversificação do ensino superior nos âmbitos público e privado, o Exame Nacional de Desempenho de Estudantes (ENADE) constitui um importante meio de mensuração da evolução destes cursos e da superação de desigualdades (Crepalde; Silveira, 2016).

Recentemente, muitos estudos têm explorado os dados educacionais do ENADE e seus âmbitos, sendo eles relacionados aos cursos, instituições, alunos e profissionais de computação. A literatura existente revela uma análise variada e ampla desses dados, com diferentes perspectivas e objetivos de pesquisa. Os estudos abordam desde análises detalhadas do perfil e desempenho de alunos até o desenvolvimento de ferramentas práticas para prepará-los para o exame, como comentários sobre questões anteriores do ENADE e sistemas de preparação adaptativa (Figueiro *et al.*, 2018; Hinterholz *et al.*, 2014; Candão; Real, 2017).

Ademais, há um foco significativo na aplicação de técnicas avançadas de análise de dados, como mineração de dados e aprendizado de máquina, para compreender os fatores que influenciam o desempenho dos alunos e das instituições de ensino (Amaro, 2019; Vista *et al.*, 2017). Além disso, com o uso dos dados ENADE, a desigualdade de gênero nesta área de formação também foi alvo de análise (Freitas *et al.*, 2019). Estudos recentes também visaram destacar a importância de explorar estratégias inovadoras, como a gamificação, para melhorar ainda mais a eficácia da preparação dos alunos e a compreensão dos resultados do ENADE na área de computação (Barbosa *et al.*, 2023a).

Logo, evidencia-se que os estudos sobre o desempenho no ENADE na área de computação têm avançado notavelmente, oferecendo *insights* relevantes sobre os fatores que influenciam o desempenho acadêmico. No entanto, a análise abrangente do grupo de cursos TIC, especialmente na edição do exame de 2021, ainda não foi abordada na literatura recente. Isto limita a compreensão das tendências da atualidade, especialmente no

contexto pós-pandemia da COVID-19. Há uma oportunidade para explorar variáveis adicionais, como categorias administrativas, organizações acadêmicas e macrorregiões geográficas, para obter uma visão mais contextualizada.

Adicionalmente, a aplicação de técnicas avançadas, como regressão e clusterização, especialmente o uso do algoritmo *k-medoids* para lidar com variáveis numéricas e categóricas, é uma necessidade não atendida na literatura atual. Essas abordagens podem oferecer percepções sobre padrões e características que influenciam o desempenho dos alunos, enriquecendo a compreensão das dinâmicas educacionais. Essa abordagem holística pode contribuir significativamente para melhorar as políticas e práticas no ensino superior, promovendo a qualidade e eficácia do processo educacional.

Diante do exposto, definem-se os objetivos norteadores deste estudo: enquanto objetivo geral, a pesquisa pretende analisar o desempenho no ENADE entre os diferentes cursos da área de avaliação “Computação e tecnologias da informação e comunicação (TIC)”. Ainda, enquanto objetivos específicos, evidenciar demandas específicas de cursos da área e identificar variáveis que influenciam o desempenho dos cursos. Nas seções subsequentes, tratar-se-á dos estudos recentes e métricas de avaliação do ENADE.

O ESTADO DA ARTE DOS ESTUDOS SOBRE O DESEMPENHO NO ENADE DE CURSOS DA ÁREA DE COMPUTAÇÃO

O atual estado da arte deste campo de pesquisa revela uma variedade de metodologias adotadas na análise de desempenho dos cursos da área de ciência de computação no ENADE. É possível citar o estudo de Silva *et al.* (2020), que se destaca entre os trabalhos com foco em dados históricos ao examinar as mudanças nos cursos de computação no Brasil, entre 2008 e 2014, usando os dados do ENADE. Os autores evidenciaram a crescente presença de cursos de computação na Região Nordeste do Brasil e o fato de que as instituições públicas superam as privadas em desempenho.

Além disso, Silva *et al.* (2020) forneceram informações sobre a distribuição geográfica e institucional. Enquanto Lima *et al.* (2018) desenvolveram uma metodologia inovadora para categorizar as questões do ENADE em domínios de conhecimento usando um sistema de *software* desenvolvido especificamente para esse propósito. Este algoritmo não apenas separa os resultados dos alunos por descritores gerais das provas, mas também

gera relatórios detalhados que ajudam a compreender melhor a estrutura das provas e os resultados dos alunos por domínios de conhecimento. Como afirmado por Lima *et al.* (2021), esses relatórios demonstram-se capazes de fornecer novos olhares que podem influenciar significativamente nas escolhas estratégicas relacionadas à educação. Como resultado, ter a capacidade de criar métodos para relatórios personalizados permite uma melhor compreensão do desempenho dos alunos, baseada em conhecimentos específicos avaliados no exame.

Em adição, Lima *et al.* (2021) afirmam que, na literatura mais comum, as análises dos dados do ENADE são realizadas com o objetivo de investigar as condições de acesso e permanência no ensino superior, enquanto uma quantidade significativa de trabalhos realizou análises com o objetivo de avaliar o desempenho dos alunos e criar perfis desses. Rosa *et al.* (2021) enfatizam o uso de um método longitudinal para avaliar o impacto de fatores socioeconômicos e administrativos no desempenho acadêmico. O estudo revela que fatores como renda familiar e formação dos pais influenciam o desempenho dos alunos ao longo do tempo, usando técnicas de mineração de dados como árvores de decisão. Isso fornece uma visão de vantagens temporais que complementam as análises transversais mais comuns.

Logo, é possível conceber que os estudos realizados até o momento abrangeram uma ampla diversidade de abordagens e perspectivas, havendo destaque para a importância da consideração de perfis específicos de estudantes nas análises de desempenho educacional, com vista a identificar potenciais fragilidades ou desigualdades entre os estudantes (Hinterholz *et al.*, 2014). Além disso, a utilização de técnicas de mineração de dados para análise de desempenho acadêmico também esteve em destaque (Amaro 2019; Freitas *et al.*, 2019; Vista *et al.*, 2017), o que pode indicar a necessidade de identificação de padrões e tendências, bem como de análises multidimensionais, que abranjam um *corpus* amplo de variáveis e fatores capazes de influenciar no desempenho estudantil e necessidades que podem ser atendidas com a mineração de dados.

Observam-se avanços nas metodologias utilizadas para análise de desempenho no ENADE, conforme mencionado anteriormente, nos estudos de Lima *et al.* (2019), Lima *et al.* (2021) e Araújo (2021), em que foi possível identificar proposições metodológicas inovadoras para a classificação e identificação de disciplinas de impacto e

processamento de dados textuais nas questões do ENADE, visando a uma contribuição mais holística sobre o exame. Ademais, o potencial inovador também é identificado em pesquisas como as desenvolvidas por Cunha *et al.* (2021), Capelari (2022) e Santos (2022), ao empregarem automação e uso de técnicas inovadoras na previsão de desempenho educacional e análise de projetos pedagógicos.

Esta necessidade de inovação também foi identificada na revisão sistemática de literatura, realizada por Barbosa *et al.* (2023a), em que os autores destacam a predominância de abordagens voltadas para a utilização, nas análises de desempenho do ENADE, de técnicas de aprendizado de máquina e estatística, e enfatizam a necessidade de inovação nessas análises de performance educacional.

A presente pesquisa, seguindo uma tendência de estudos recentes, como os mapeamentos sistemáticos de Barbosa *et al.* (2023a), não apenas confirma a utilidade dessas abordagens, mas também examina a forma como as variáveis se relacionam entre si em grupos homogêneos identificados por clusterização. Assim, este estudo fornece uma compreensão mais aprofundada da situação atual do campo estudado, ao mesmo tempo em que tem potencial de promover melhorias nas diretrizes pedagógicas relacionadas à ciência da computação. Esta seção teve como finalidade fornecer uma visão geral e abrangente sobre o atual estado das pesquisas no campo da análise de desempenho acadêmico em ciências da computação, em que foi possível destacar as principais áreas, métodos e tendências emergentes e em ascensão.

MÉTRICAS DO EXAME ENADE

A avaliação ocorre em três pilares da Educação Superior: instituição, curso e estudante, e vem sendo estrategicamente estruturada e supervisionada pela Comissão Nacional de Avaliação da Educação Superior (CONAES), segundo Lima *et al.* (2019). Para isso, segundo BRASIL (2004), a operacionalização foi instruída por meio do ENADE, aplicada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), visando a comparar o conteúdo programático dos cursos de graduação com o desempenho dos acadêmicos, servindo também como parâmetro e medição da evolução das competências dos profissionais oriundos do ensino superior.

Complementando sobre a operacionalização do exame, ele é realizado com base em ciclos periódicos de avaliação, em que são avaliados, a cada três anos pautado em áreas específicas, por meio de uma avaliação de múltipla escolha, os conhecimentos dos estudantes com base nos eixos geral, específicos e tecnológico de cada área avaliada (BRASIL, 2017; Lima *et al.*, 2019). Dentre os benefícios advindos, estão a criação de políticas públicas, junto ao Ministério da Educação (MEC) focadas nos pilares avaliados, assim como a transparência dos indicadores de qualidade do ensino superior (Lima *et al.*, 2021). Os principais indicadores que são utilizados para medir o Ensino Superior no Brasil atualmente, segundo (BRASIL, 2017), são:

- Conceito ENADE;
- Conceito Preliminar de Curso (CPC);
- Indicador de Diferença entre os Desempenhos Observado e Esperado (IDD);
- Índice Geral de Cursos (IGC).

Dentre o mais indicado, e que será foco desta pesquisa, está o Conceito ENADE, que é o desempenho médio obtido pelos estudantes de um curso em comparação com a média de desempenho alcançada pela área de avaliação. Seguindo a Norma Técnica N° 2/2023/CEI/CGGI/DAES, do INEP (2023b), o indicador Conceito ENADE pode ser formulado matematicamente como uma média ponderada 25% do peso na Formação Geral e 75% do peso na Formação Específica, como demonstra a Equação 1:

$$NC_c = 0,25 \cdot NP_{FG_c} + 0.75 \cdot NP_{CE_c} \quad (1)$$

Em que:

- NC_c é a nota dos concluintes no ENADE do curso de graduação c ;
- NP_{FG_c} é a nota padronizada de Formação Geral do exame do curso de graduação c ;
- NP_{CE_c} é a nota padronizada do Componente Específico do exame do curso de graduação c .

Além disso, é crucial destacar que os indicadores são apresentados em uma escala contínua e em cinco níveis ou faixas, indo de 1 (um) a 5 (cinco), em que 3 (três) é visto como

o mínimo satisfatório, e 5 (cinco) representa o excelente, seguindo MEC (2016), a conversão é conforme os dados descritos na **Tabela 1**.

Tabela 1. Parâmetros de conversão dos indicadores discretos e contínuos do ENADE

Faixa dos indicadores na forma discreta	Faixa dos indicadores na forma contínua
1	$0,000 \leq \text{indicador} \leq 0,945$
2	$0,945 \leq \text{indicador} \leq 1,945$
3	$1,945 \leq \text{indicador} \leq 2,945$
4	$2,945 \leq \text{indicador} \leq 3,945$
5	$3,945 \leq \text{indicador} \leq 5,000$

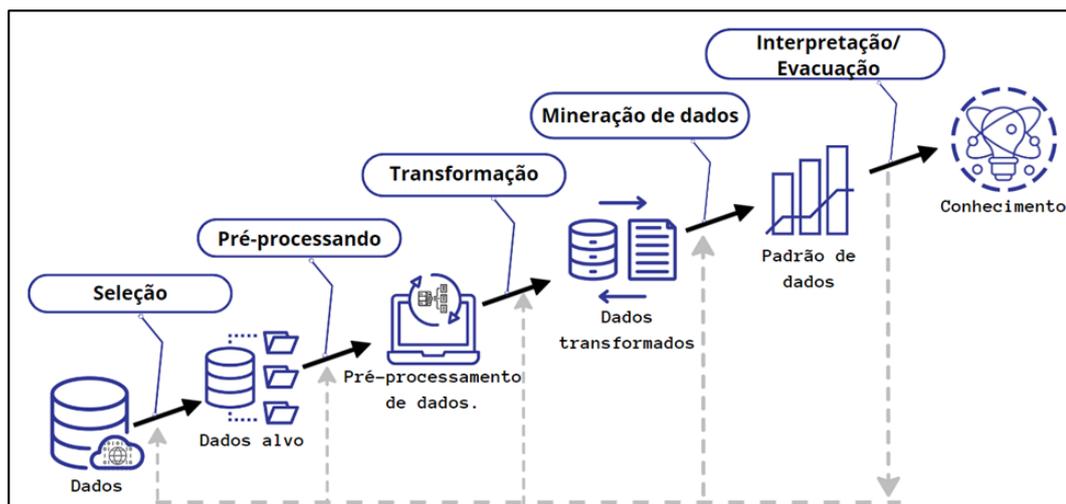
Fonte: adaptada de INEP (2023b, 2023c, 2023d, 2023e)

O núcleo de cursos da área de Computação é formado por Bacharelado em Engenharia de Computação, Licenciatura em Computação, Bacharelado em Ciência da Computação e Bacharelado em Sistemas de Informação, que participam do exame desde 2005 (Lima *et al.*, 2021). Cortelazzo *et al.* (2021) descreve que a partir de 2008, passaram a integrar as avaliações do ENADE cursos técnicos, como é o exemplo de Análise e Desenvolvimento de Sistemas e Redes de Computadores, assim como em 2008 com Licenciatura em Ciências da Computação e 2017, foi a vez de Gestão de Tecnologia da Informação.

METODOLOGIA

O presente estudo adotou uma abordagem metodológica baseada no processo de *Knowledge Discovery in Database* (KDD), visando a compreender as características, relações, condições e padrões dos cursos superiores na área de computação. A pesquisa foi delineada como descritiva quanto aos objetivos, de fonte de dados como documental e de abordagem quantitativa (Prodanov; Freitas, 2013). O processo e suas etapas estão relacionados na **Figura 1**.

Figura 1. Um panorama geral das etapas que compõem o processo KDD



Fonte: Adaptado de Fayyad et al. (1996).

Antes de tudo, é relevante destacar que todas as análises, procedimentos de tratamento, manipulações de dados e resultados foram conduzidos por meio da linguagem de programação *Python*. Essa escolha foi motivada pela sua portabilidade, natureza de código aberto, combinação de simplicidade e robustez, além da facilidade de aprendizado, conforme descrito por Banin (2018).

Seleção

A primeira etapa envolveu a identificação e seleção das fontes de dados relevantes para a análise. Utilizando dados abertos disponibilizados pelo INEP (2023a)¹ por meio da

página de indicadores de qualidade da educação superior¹, foram selecionados quatro conjuntos de dados:

- conceito_enade_2021.xlsx;
- IDD_2021.xlsx;
- resultados_cpc_2021.xlsx;
- IGC_2021.xlsx.

Pré-processamento

Foi realizada uma limpeza nos conjuntos de dados para selecionar informações pertinentes a esta pesquisa aos cursos específicos de interesse. Isso incluiu a identificação e extração dos cursos de:

- Análise e Desenvolvimento de Sistemas (Tecnólogo);
- Ciência da Computação (Bacharelado);
- Ciência da Computação (Licenciatura);
- Gestão da Tecnologia da Informação (Tecnólogo);
- Redes de Computadores (Tecnólogo);
- Sistemas de Informação (Bacharelado).

Transformação

Os dados foram integrados em uma única base, empregando chaves de relacionamento, como o código da área e das instituições de ensino superior. As variáveis foram categorizadas em quantitativas e categóricas. Dentre as quantitativas, encontram-se os indicadores de qualidade, que são classificados em faixas, variando de 1 (um) a 5 (cinco), tanto em sua forma contínua como em faixas, a saber:

- Conceito ENADE;

¹ <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-de-qualidade-da-educacao-superior>

- CPC;
- IDD;
- IGC.

Contudo, é importante observar que cursos desprovidos de um indicador específico são identificados mediante a sigla "SC", denotando a ausência de conceito correspondente. Além disso, outras variáveis quantitativas foram selecionadas, sendo estas:

- Nota Padronizada - Conhecimento Específico;
- Nota Padronizada - Doutores;
- Nota Padronizada - Formação Geral;
- Nota Padronizada - Infraestrutura e Instalações Físicas;
- Nota Padronizada - Mestres;
- Nota Padronizada - Organização Didático-Pedagógica;
- Nota Padronizada - Oportunidade de Ampliação da Formação;
- Nota Padronizada - Regime de Trabalho;
- Proporção de Concluintes Participantes com nota no Enem.

Já as variáveis qualitativas ou categóricas foram agrupadas em:

- Área de Avaliação: Análise e Desenvolvimento de Sistemas, Ciência da Computação (Bacharelado), Ciência da Computação (Bacharelado), Gestão da Tecnologia da Informação, Redes de Computadores e Sistemas de Informação.
- Categoria administrativa: especiais (ESP), privadas com fins lucrativos (PCFL), públicas estaduais (PE), públicas federais (PF), públicas municipais (PM) e privadas sem fins lucrativos (PSFL);
- Organização acadêmica: Centros Federais de Educação Tecnológica ou apenas Centros Federais (CEFET), Centros Universitários (CU), Faculdades (FA), Institutos Federais de Educação Tecnológica ou apenas Institutos Federais (IF) e Universidades (UN);
- Modalidade de Ensino: presencial ou à distância;
- Macrorregião: Centro-Oeste (CO), Nordeste (NE), Norte (NO), Sudeste (SE) e Sul (SU).

Mineração de dados

Nesta fase, foi realizada a aplicação de técnicas e ferramentas de mineração de dados (Costa *et al.*, 2012) nos conjuntos já processados. Esta abordagem compreendeu a utilização de técnicas de análise exploratória (Medri, 2011), como gráficos, tabelas e medidas descritivas, bem como técnicas multivariadas (Barroso; Artes, 2003), incluindo regressão (Coelho; Cunha, 2007) e clusterização (Park; Jun, 2009). Essa combinação permitiu extrair *insights* significativos sobre as características, relações, condições e padrões relacionados aos cursos e instituições de ensino superior na área de computação, oferecendo uma visão abrangente do panorama educacional no contexto brasileiro.

Destaque-se que, na análise multivariada, foram considerados exclusivamente os cursos que apresentaram todos os indicadores simultaneamente, descartando aqueles que possuíam informações ausentes. Dessa forma, na análise exploratória, foram examinados um total de 1.650 cursos, enquanto na análise multivariada esse número foi reduzido para 1.505. Além disso, no processo de análise multivariada, as variáveis quantitativas passaram por normalização, utilizando-se a técnica Min-Max, enquanto as variáveis qualitativas foram transformadas em variáveis *dummy*, garantindo uma representação adequada e comparabilidade entre os diferentes tipos de dados (Oliveira *et al.*, 2020).

Avaliação

Por fim, a etapa de avaliação envolveu a interpretação e análise crítica dos resultados derivados da mineração de dados. A avaliação metódica dos cursos e instituições possibilitou a extração de informações relevantes, anteriormente implícitas nos dados. Essa abordagem visou a contribuir para o alcance dos objetivos da pesquisa, fornecendo percepções relevantes e promovendo uma compreensão mais aprofundada do atual cenário educacional na área de computação.

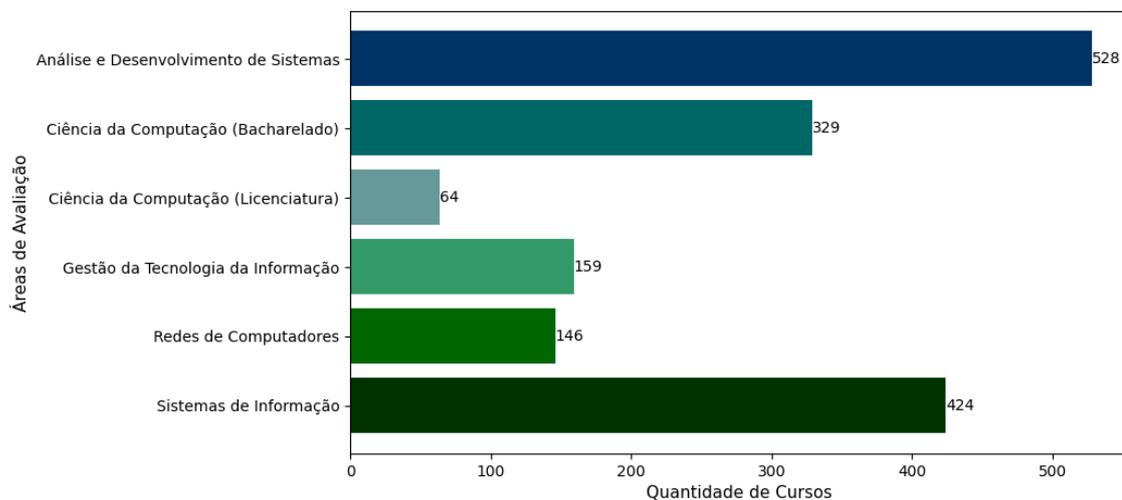
RESULTADOS

A seguir serão apresentados os resultados e discussões utilizando as técnicas de análise exploratória e multivariada de dados dos cursos e instituições de computação na edição de 2021 do ENADE.

Análise exploratória de dados dos cursos e instituições da área de computação no ENADE de 2021

No que tange à distribuição de cursos de computação em diferentes áreas de avaliação, a Análise e Desenvolvimento de Sistemas se destaca com a maior quantidade de cursos, possuindo 528, o que corresponde a 32% do total de 1.650 avaliados, como mostra a **Figura 2**.

Figura 2. Número total de cursos por área de avaliação



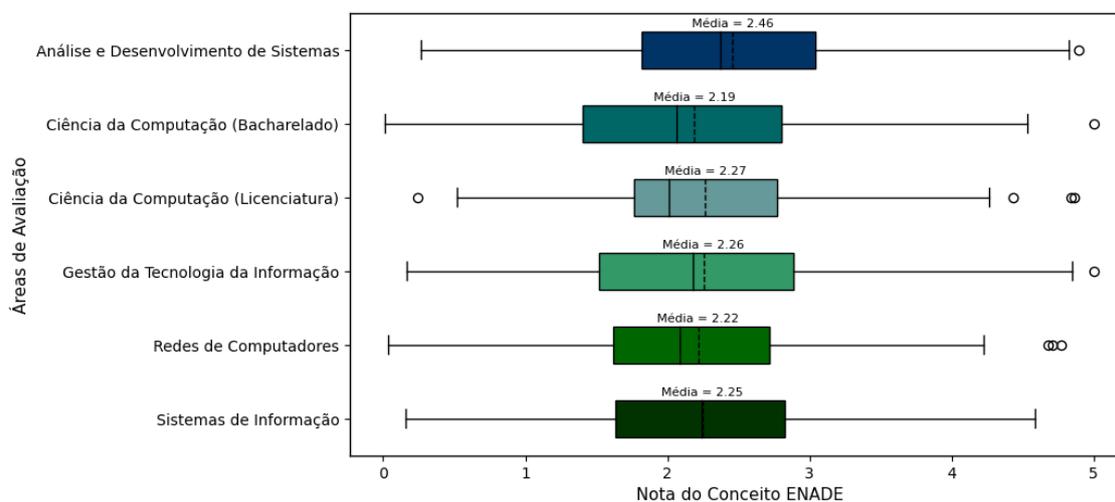
Fonte: dados da pesquisa (2024)

Isso pode indicar uma crescente demanda por profissionais com habilidades práticas e voltadas para essa área no mercado. Por outro lado, a presença mais modesta de cursos de Licenciatura em Ciência da Computação, com apenas 64 cursos, cerca de

3,88% do total, poderia sugerir uma área de oportunidade de expansão e de maior análise, por exemplo, aumentando a quantidade de cursos nas próximas edições do ENADE.

Paralelamente, sabe-se que a nota do Conceito ENADE é crucial para avaliar a qualidade no desempenho de cada área de avaliação. Para tanto, a **Figura 3** apresenta o *boxplot*, com a finalidade de representar a variação de dados do Conceito ENADE em cada área de avaliação.

Figura 3. Variação do Conceito ENADE por área de avaliação

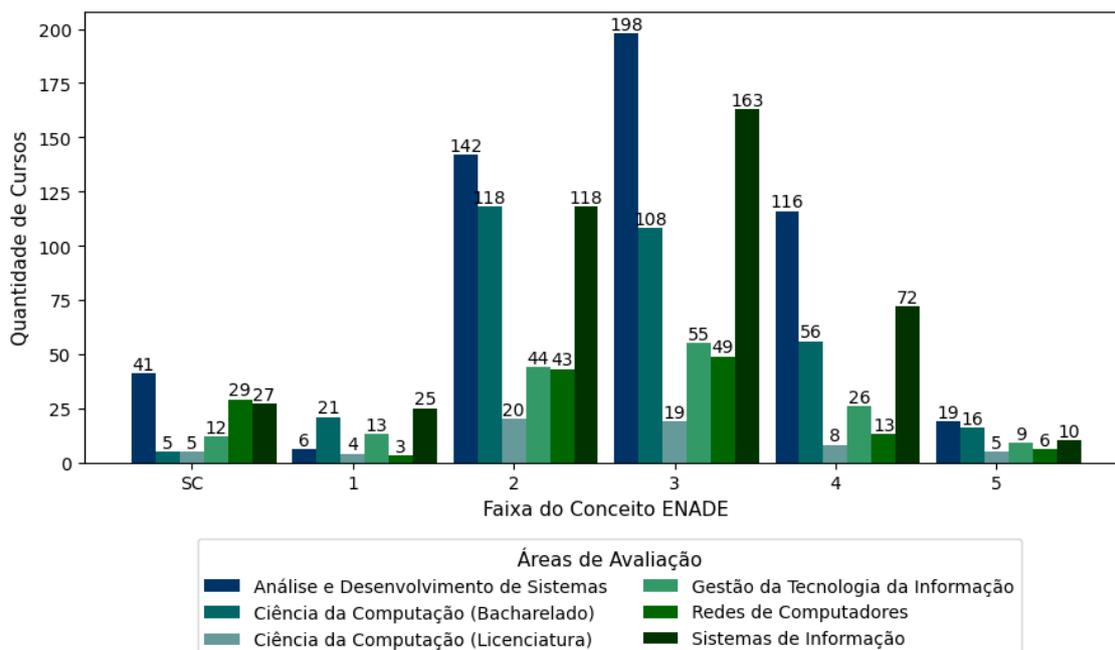


Fonte: dados da pesquisa (2024)

Dessa forma, foi possível identificar desempenho variado entre as áreas de ensino. A área de Análise e Desenvolvimento de Sistemas lidera com uma média de aproximadamente 2,46 no Conceito ENADE, indicando um desempenho positivo. Por outro lado, Ciência da Computação (Bacharelado) apresenta a menor média, cerca de 2,19, sugerindo necessidade de melhorias. Apesar dessas discrepâncias, as médias estão dentro de uma faixa estreita, indicando desempenhos relativamente similares entre as áreas avaliadas.

Sabe-se ainda que o Conceito ENADE categoriza as notas em faixas. A visualização da distribuição de cursos por área de avaliação e faixa do Conceito ENADE oferece uma visão mais detalhada do atual cenário dos cursos de computação, como é possível observar na **Figura 4**.

Figura 4. Número total de cursos por área de avaliação e faixa do Conceito ENADE

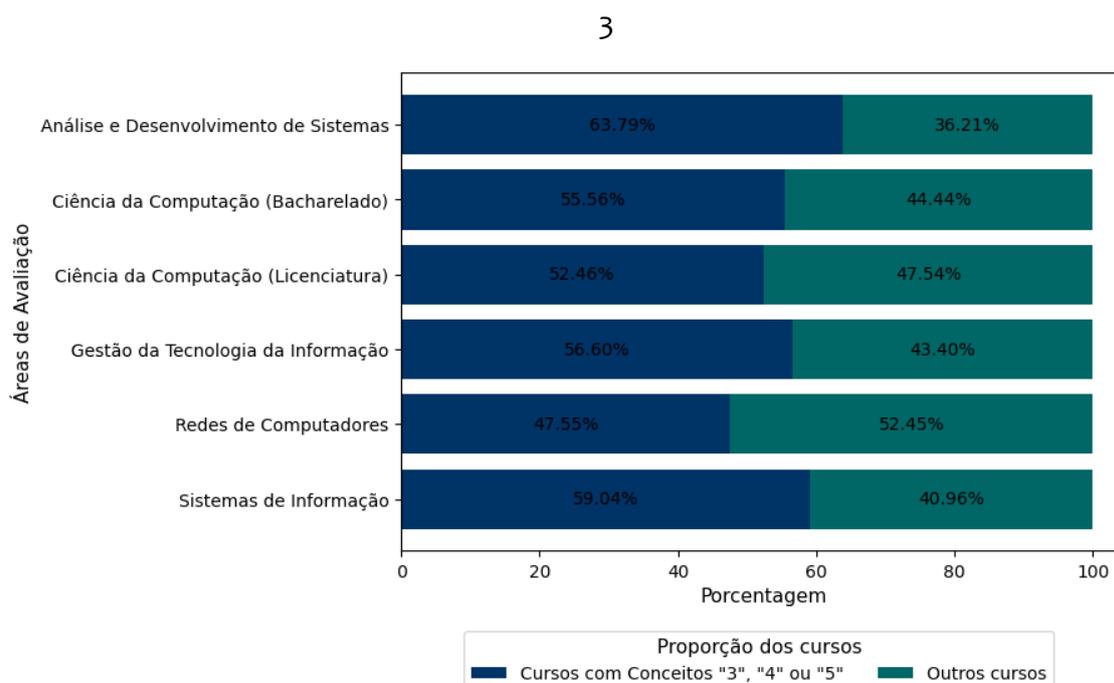


Fonte: dados da pesquisa (2024)

Embora existam variações, é interessante notar que a maioria dos cursos, independentemente das áreas, tende a se concentrar nas faixas de Conceito ENADE 2 (dois) e 3 (três). Essa concentração em níveis médios poderia indicar desafios comuns enfrentados pelos cursos em termos de qualidade acadêmica ou eficácia dos métodos de ensino. Uma investigação mais profunda dessas faixas poderia revelar os motivos subjacentes e identificar estratégias para elevar o desempenho geral.

De qualquer forma, é interessante se ter uma noção sobre desempenho mínimo dos cursos, considerando a classificação 3 (três) como um mínimo satisfatório (MEC, 2016), como mostra a **Figura 5**.

Figura 5. Distribuição dos cursos das áreas de avaliação com conceitos ENADE superior a



Fonte: dados da pesquisa (2024)

A área de Análise e Desenvolvimento de Sistemas se destaca novamente, com a maior porcentagem de cursos, cerca de 63,79%, alcançando um alto nível de desempenho. Isso sugere que esta área está conseguindo manter um padrão de qualidade em uma faixa mais alta de avaliação. No entanto, Redes de Computadores apresenta uma porcentagem menor, aproximadamente 47,55%, indicando que há espaço para melhorias neste aspecto.

Ao considerar como as áreas de avaliação se comportam em função das demais características relevantes, é necessário se aprofundar na média do Conceito ENADE em relação à categoria administrativa das instituições, informações disponíveis na **Tabela 2**.

Tabela 2. Média do Conceito ENADE por área de avaliação e categoria administrativa

Área de Avaliação	Categoria Administrativa					
	ESP	PCFL	PSFL	PE	PF	PM
Análise e Desenvolvimento de Sistemas	1,85	2,13	2,47	2,88	3,03	2,84
Ciência da Computação (Bacharelado)	--	1,60	2,09	2,46	3,05	1,40
Ciência da Computação (Licenciatura)	--	1,51	2,71	2,02	2,33	--
Gestão da Tecnologia da Informação	1,14	2,12	2,34	2,92	2,70	2,33
Redes de Computadores	--	2,02	2,30	1,69	3,00	--
Sistemas de Informação	2,52	2,00	2,27	2,26	2,68	2,69
Média por Categoria Administrativa	1,98	1,99	2,30	2,52	2,86	2,21

Fonte: dados da pesquisa (2024)

Ao analisar a **Tabela 2**, observamos que as instituições públicas federais têm a média mais alta, cerca de 2,86, sugerindo que políticas e recursos disponíveis nelas contribuem para esse desempenho. Já as instituições especiais apresentam a menor média, aproximadamente 1,98. Embora a ausência de dados em algumas categorias, como públicas municipais, possa afetar a análise, fica evidente um desempenho abaixo do esperado nas instituições privadas com fins lucrativos, em torno de 1,99. Isso destaca a necessidade de identificar e abordar os fatores que prejudicam essas instituições, para que elas possam ter uma maior qualidade de ensino.

Vale ressaltar que a **Tabela 2** demonstra as médias de cada área de avaliação por cada categoria, bem como as médias das variáveis categóricas. Isso não significa que a média de uma categoria é igual à média das médias das áreas de avaliação em relação à categoria em questão. O mesmo pressuposto se aplica para as Tabelas 3, 4 e 5. Simultaneamente, a **Tabela 3** explora a média do Conceito ENADE por área de avaliação e organização acadêmica, o que revela a influência da estrutura institucional no desempenho dos cursos.

Tabela 3. Média do Conceito ENADE por área de avaliação e organização acadêmica

Área de Avaliação	Organização Acadêmica				
	CEFET	CU	FA	IF	UN
Análise e Desenvolvimento de Sistemas	--	2,35	2,38	3,00	2,25
Ciência da Computação (Bacharelado)	2,86	1,79	1,38	2,46	2,47
Ciência da Computação (Licenciatura)	--	2,27	0,24	2,08	2,47
Gestão da Tecnologia da Informação	--	2,36	2,35	2,70	2,05
Redes de Computadores	--	2,13	2,08	2,92	2,11
Sistemas de Informação	4,39	2,21	1,96	2,47	2,36
Média por Organização Acadêmica	3,63	2,20	2,10	2,72	2,34

Fonte: dados da pesquisa (2024)

Institutos federais de educação se destacam com a média mais alta, cerca de 2,72, indicando uma abordagem educacional eficaz com base nos critérios adotados durante a avaliação. Em contraste, as faculdades, por apresentarem média de 2,10, possivelmente enfrentam desafios associados à qualidade de ensino e padronização dos projetos pedagógicos. No entanto, uma análise mais coesa ainda levaria em conta a presença de dados ausentes nos centros federais, podendo influenciar a média dessa categoria. Paralelamente, a **Tabela 4** aprofunda compreensão do desempenho dos cursos ao explorar a média do Conceito ENADE de acordo com a modalidade de ensino.

Tabela 4. Média do Conceito ENADE por área de avaliação e modalidade de ensino

Área de Avaliação	Modalidade de Ensino	
	Educação presencial	Educação a distância
Análise e Desenvolvimento de Sistemas	2,48	2,36
Ciência da Computação (Bacharelado)	2,19	1,81
Ciência da Computação (Licenciatura)	2,36	2,09
Gestão da Tecnologia da Informação	2,08	2,63
Redes de Computadores	2,22	2,21
Sistemas de Informação	2,27	1,87
Média por Modalidade de Ensino	2,30	2,33

Fonte: dados da pesquisa (2024)

A educação a distância tem uma média ligeiramente superior, cerca de 2,33, comparada ao ensino presencial, aproximadamente 2,30, sugerindo eficácia apesar da limitação física. No entanto, análises por área revelam diferenças consideráveis. Por exemplo, em Gestão da Tecnologia da Informação, o ensino a distância apresenta uma média de 2,63, enquanto o presencial fica em 2,08. Isso levanta questões sobre as metodologias e recursos tecnológicos utilizados, que podem estar contribuindo para esse desempenho superior. Por fim, a análise macrorregional do desempenho acadêmico, conforme apresentado na **Tabela 5**, revela padrões interessantes nas diferentes macrorregiões.

Tabela 5 – Média do Conceito ENADE por área de avaliação e macrorregião

Área de Avaliação	Macrorregião				
	CO	NE	NO	SE	SU
Análise e Desenvolvimento de Sistemas	--	2,35	2,38	3,00	2,25
Ciência da Computação (Bacharelado)	2,86	1,79	1,38	2,46	2,47
Ciência da Computação (Licenciatura)	--	2,27	0,24	2,08	2,47
Gestão da Tecnologia da Informação	--	2,36	2,35	2,70	2,05
Redes de Computadores	--	2,13	2,08	2,92	2,11
Sistemas de Informação	4,39	2,21	1,96	2,47	2,36
Média por Macrorregião	2,14	2,39	2,10	2,18	2,62

Fonte: dados da pesquisa (2024)

Essa análise regional ressalta a importância de adaptar abordagens educacionais às particularidades de cada macrorregião, garantindo uma qualidade uniforme em todo o país. Enquanto as macrorregiões Sul e Nordeste se destacam com médias mais altas, cerca de 2,62 e 2,39, respectivamente, sugerindo uma abordagem eficaz para o ensino de computação, as macrorregiões Norte e Centro-Oeste apresentam médias mais baixas, aproximadamente 2,10 e 2,13, respectivamente, indicando possíveis desafios nestes contextos.

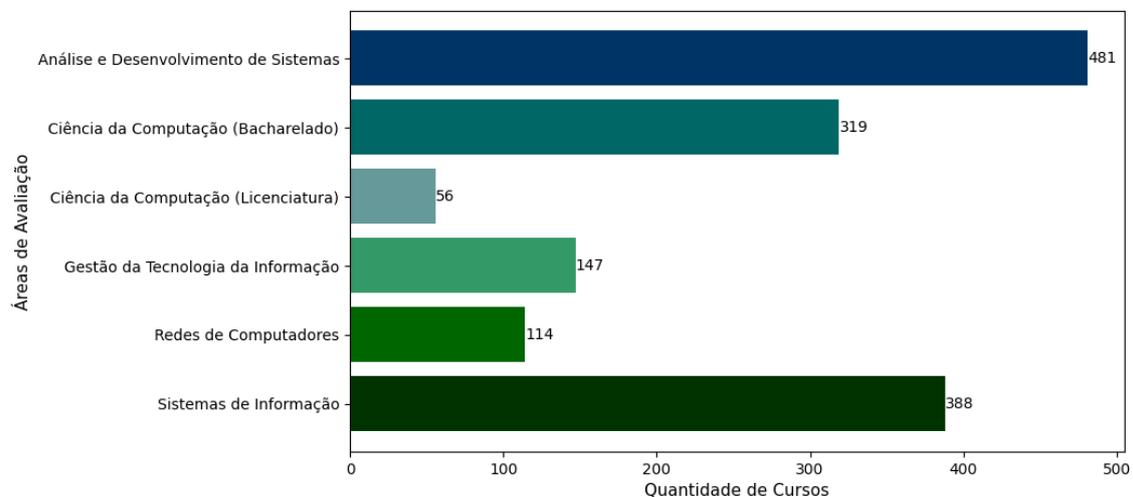
Ressalte-se, portanto, a partir das análises das Tabelas 2, 3, 4 e 5, que o tipo de categoria administrativa, organização acadêmica, modalidade de ensino ou macrorregião

podem constituir fatores determinantes no desempenho dos cursos na área de computação.

Análise multivariada de dados dos cursos da área de computação no ENADE de 2021

Conforme destacado na seção 2.4 deste trabalho, ao realizar a análise multivariada dos dados, optou-se por excluir os dados do tipo “SC” ou campos vazios (“null”). O resultado da criação desta nova base de dados, em relação à quantidade total de cursos por área de avaliação, após este processo, é apresentado na **Figura 6**.

Figura 6. Número total de cursos por área de avaliação após a triagem para a análise multivariada



Fonte: dados da pesquisa (2024)

Em comparação à **Figura 2**, na **Figura 6** é possível observar uma pequena redução da parcela total de cursos por cada área de avaliação, de 1.650 para 1.505 cursos. Posteriormente, procedeu-se à normalização das variáveis quantitativas e à conversão das variáveis categóricas em variáveis *dummy*.

Regressão múltipla

Na sequência, a análise prosseguiu com a aplicação da regressão múltipla, com o intuito de selecionar as variáveis significativas na explicação da variação do Conceito ENADE. Portanto, o resultado da regressão múltipla, em que o Conceito ENADE é tratado

como variável dependente e as demais variáveis são consideradas como independentes, é apresentado na **Figura 7**, com um nível de significância de 2,5%.

Figura 7. Resultado da regressão múltipla

OLS Regression Results						
Dep. Variable:	ConceitoEnadeContínuo	R-squared:	0.581			
Model:	OLS	Adj. R-squared:	0.579			
Method:	Least Squares	F-statistic:	239.8			
Date:	Fri, 01 Sep 2023	Prob (F-statistic):	5.90e-255			
Time:	13:54:05	Log-Likelihood:	1041.6			
No. Observations:	1392	AIC:	-2065.			
Df Residuals:	1383	BIC:	-2018.			
Df Model:	8					
Covariance Type:	nonrobust					
		coef	std err	t	P> t	[0.025 0.975]
Intercept		-0.0874	0.022	-3.982	0.000	-0.131 -0.044
CategoriaAdministrativaPúblicaEstadual		0.0896	0.011	8.100	0.000	0.068 0.111
ProporçãoDeConcluintesParticipantesComNotaNoEnem		-0.0507	0.016	-3.156	0.002	-0.082 -0.019
NotaPadronizadaRegimeDeTrabalho		-0.0745	0.014	-5.423	0.000	-0.101 -0.048
NotaPadronizadaMestres		0.0555	0.015	3.811	0.000	0.027 0.084
NotaPadronizadaOportunidadeDeAmpliaçãoDaFormação		0.0918	0.017	5.532	0.000	0.059 0.124
CategoriaAdministrativaPúblicaFederal		0.1151	0.009	12.480	0.000	0.097 0.133
IGCContínuo		0.3994	0.036	11.223	0.000	0.330 0.469
IDDContínuo		0.5423	0.019	28.362	0.000	0.505 0.580
Omnibus:	5.941	Durbin-Watson:	1.671			
Prob(Omnibus):	0.051	Jarque-Bera (JB):	6.132			
Skew:	0.117	Prob(JB):	0.0466			
Kurtosis:	3.226	Cond. No.	24.7			

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Fonte: dados da pesquisa (2024)

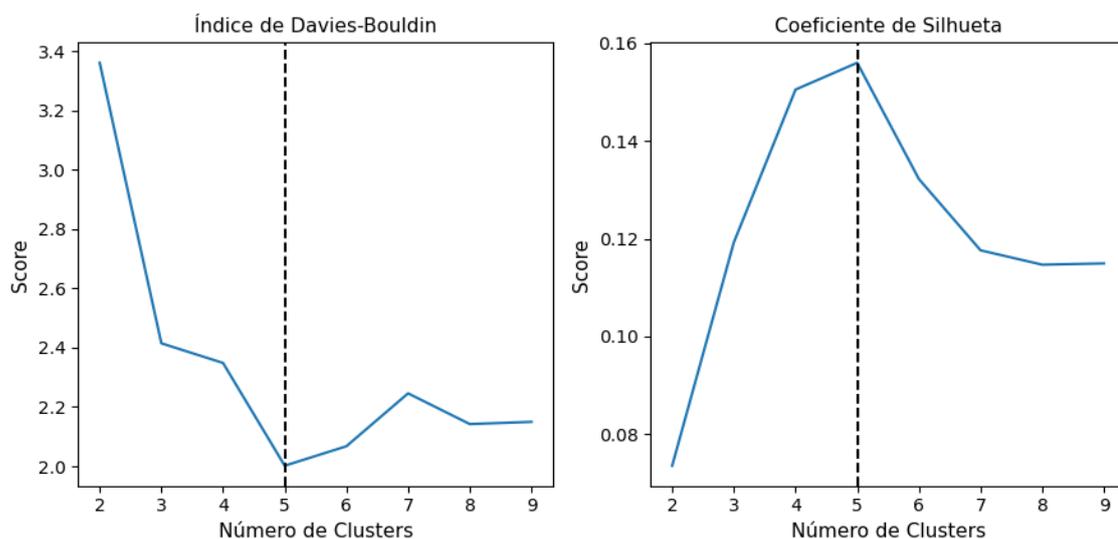
As informações contidas na **Figura 7** confirmam a significância do modelo e de cada uma das variáveis consideradas. O R-quadrado ajustado corresponde a 0,579 e valor-p associado à f é cerca de $5,90 \cdot 10^{-255}$, valor este próximo a zero. E os valores-p dos testes individuais de cada coeficiente são todos abaixo do nível de significância.

O modelo de regressão utilizado neste estudo desempenha um papel crucial na seleção criteriosa das variáveis que compõem a base de dados, já que permite identificar e manter apenas as variáveis mais significativas e relevantes para a análise em questão, representando um recurso fundamental para a próxima fase de clusterização. Essa abordagem meticulosa visa a maximizar a qualidade e a pertinência da análise, permitindo extrair conclusões mais precisas e significativas sobre a diversidade atual intrínseca dos cursos de computação, que transcende a mera utilização da base de dados inicial em seu estado bruto.

Clusterização

Uma vez feita a regressão, a base de dados resultante, composta exclusivamente por variáveis que apresentaram relevância estatística, serviu para a identificação de grupos distintos de cursos da área de computação, por meio da clusterização, explorando padrões e relações, antes implícitas na base bruta de dados. Para tanto, antes de tudo, faz-se necessário a determinação do número ideal de *clusters* que, neste estudo, foi feito com o auxílio gráfico do Índice de Davies-Bouldin e o Coeficiente de Silhueta, como mostra a **Figura 8**.

Figura 8. Número de cluster pelo Índice de Davies-Bouldin e Coeficiente de Silhueta

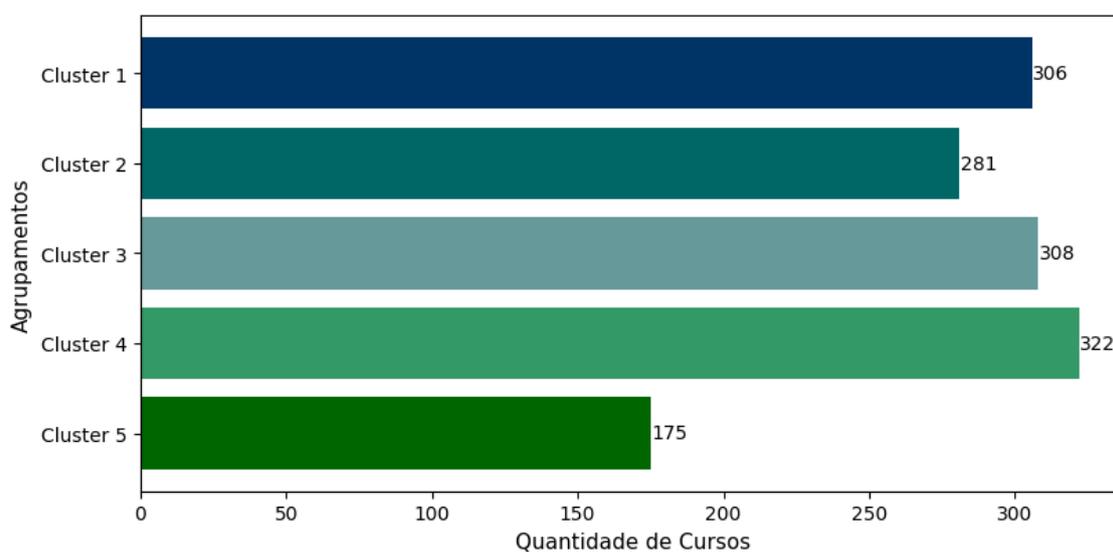


Fonte: dados da pesquisa (2024)

Conforme evidenciado na **Figura 8**, a determinação do número ideal de *clusters* para a análise em questão apontou consistentemente para a escolha de um número ótimo de *clusters* igual a 5 (cinco). A partir desse número, foi possível realizar a técnica de clusterização *K-Medoids* na base de dados, resultante da clusterização, a fim de determinar grupos distintos nos cursos de computação, cada um com suas próprias características e padrões. Antes de propor a análise e discussão para cada grupo, na **Figura 9**, é apresentada

a distribuição da quantidade de cursos por grupo, resultado da aplicação da técnica de clusterização.

Figura 9. Número total de cursos por cluster.

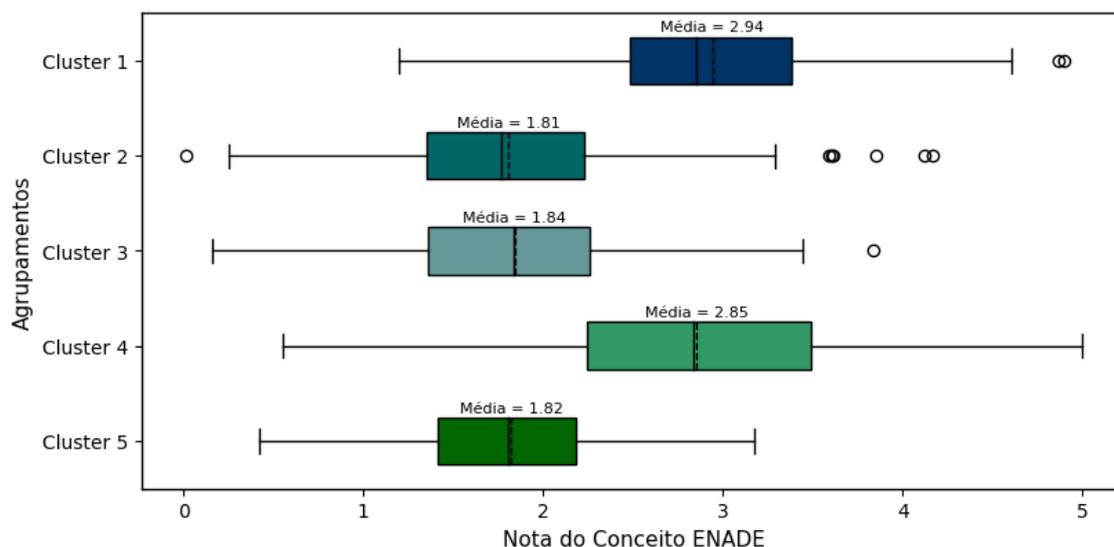


Fonte: dados da pesquisa (2024)

A **Figura 9** fornece uma representação visual do agrupamento dos cursos de computação com base em suas características intrínsecas. Ela nos permite examinar como estes cursos estão distribuídos em cada grupo. Destaque-se que o quarto *cluster* possui a maior quantidade de cursos, um total de 322; enquanto o quinto *cluster* apresenta uma quantidade relativamente menor, com 175 cursos.

No entanto, embora estes agrupamentos sejam notáveis em sua distribuição, é essencial conduzir uma análise mais aprofundada de cada um deles, a fim de compreender suas características específicas relacionadas à qualidade, composição, padrões e outros aspectos relevantes. Para tanto, a **Figura 10** exhibe o *boxplot* que ilustra a distribuição do Conceito ENADE dentro de cada *cluster*.

Figura 10. Variação do Conceito ENADE por cluster.

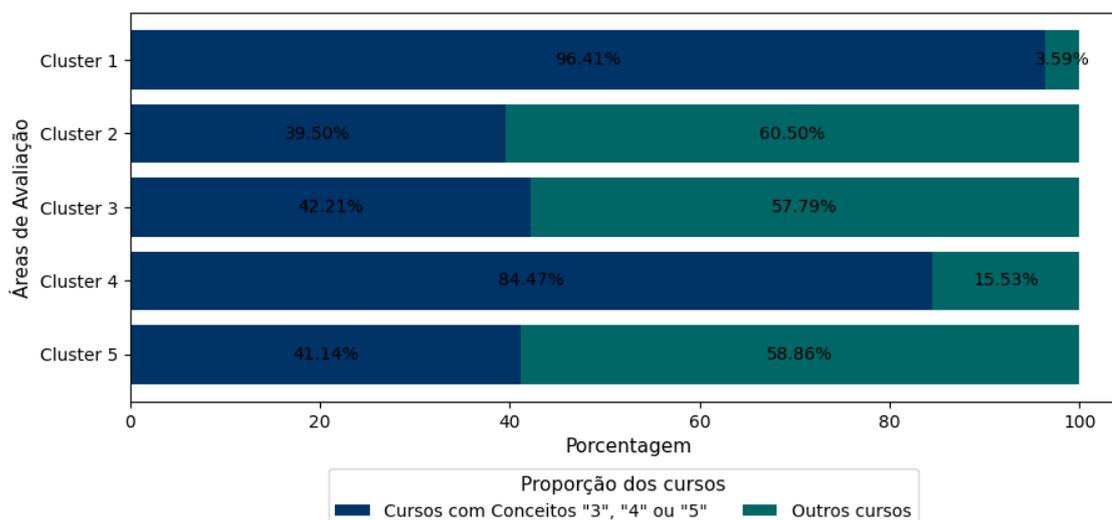


Fonte: dados da pesquisa (2024)

Ao analisar a **Figura 10**, nota-se que os grupos exibem uma dispersão nos resultados obtidos, com destaque para os *clusters* 1 (um) e 4 (quatro), que apresentam uma média do Conceito ENADE de aproximadamente 2,94 e 2,85, respectivamente; em contrapartida, os *clusters* 2 (dois) e 5 (cinco), registram a menor média, em torno de 1,81 e 1,82, respectivamente. Esta considerável variação sugere que cursos que compartilham o mesmo grupo também compartilham características semelhantes no que diz respeito ao desempenho acadêmico ou qualidade de ensino. Essa observação pode ser considerada crucial para identificar áreas passíveis de melhoria em cursos específicos e para a implementação de estratégias de aprimoramento direcionadas de maneira eficaz.

Ainda, é possível avaliar a qualidade de cada *cluster* examinando a proporção de cursos com conceitos 3 (três), 4 (quatro) ou 5 (cinco) em relação ao total de cursos em cada grupo. Essa abordagem visa, mais uma vez, a identificar os cursos que demonstram um desempenho minimamente satisfatório. Os resultados dessa análise são apresentados na **Figura 11**.

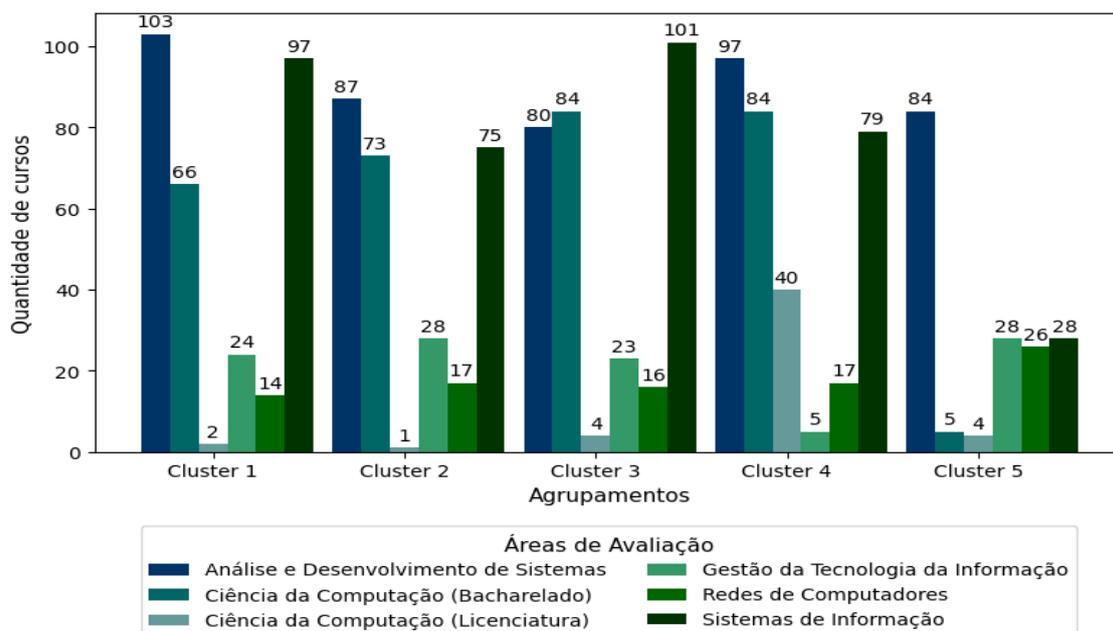
Figura 11. Distribuição dos cursos dos clusters com Conceito ENADE 3, 4 ou 5



Fonte: dados da pesquisa (2024)

A **Figura 11** reforça a análise apresentada na **Figura 10**, indicando novamente que os *clusters* 1 (um) e 4 (quatro) se destacam em termos de qualidade dos cursos, com uma alta proporção de 96,41% e 84,47%, respectivamente, dos cursos obtendo conceitos 3 (três), 4 (quatro) ou 5 (cinco). Por outro lado, os *clusters* 2 (dois) e 5 (cinco) demonstram um desempenho inferior, com apenas 39,50% e 42,21%, respectivamente, dos cursos alcançando a mesma alta proporção de conceitos. Quanto à composição e heterogeneidade de cada grupo, a **Figura 12** mostra o número de cursos de acordo com as áreas de avaliação em cada agrupamento.

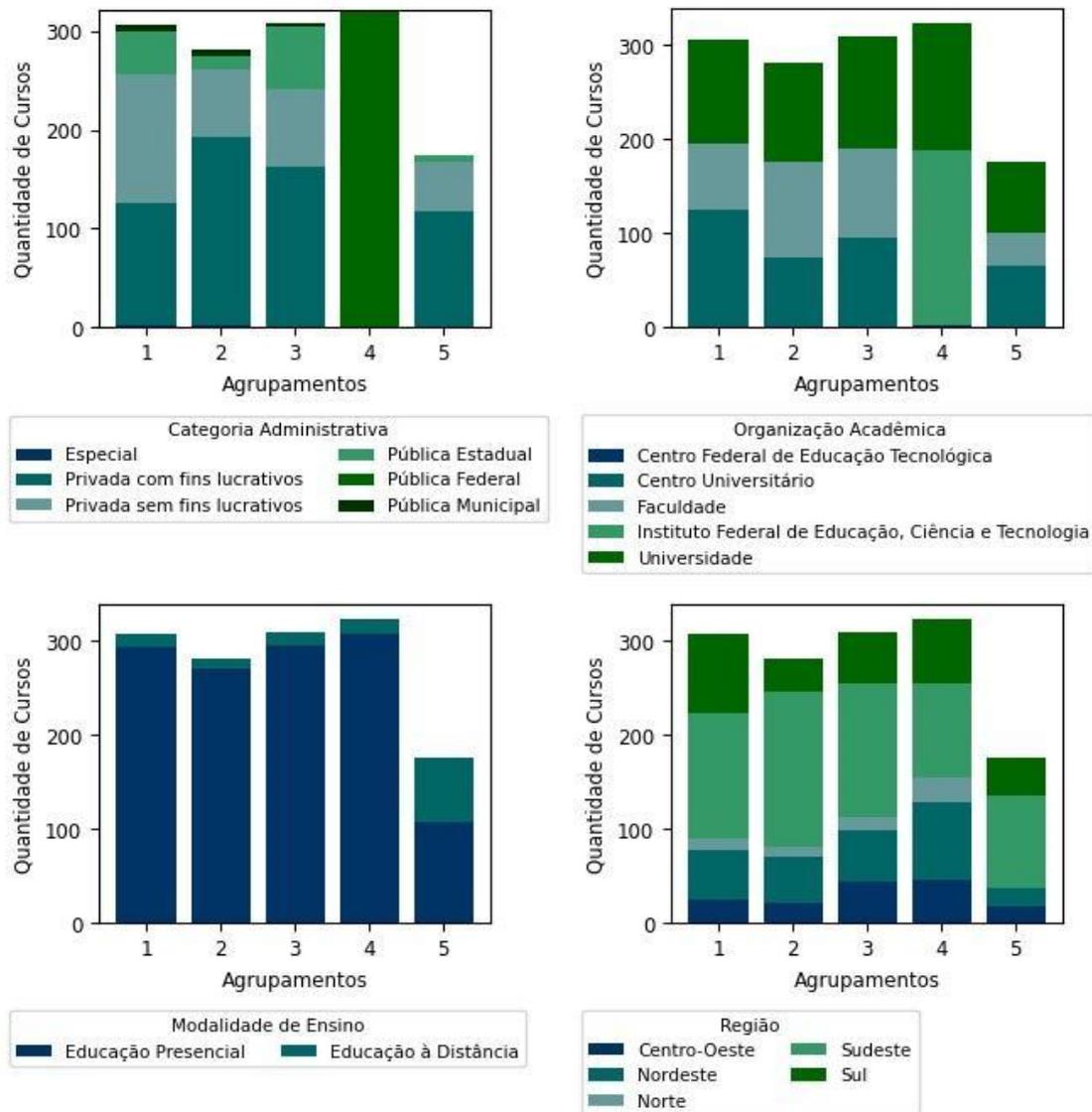
Figura 12. Número total de cursos por área de avaliação e cluster



Fonte: dados da pesquisa (2024)

Observa-se que a **Figura 12** segue, relativamente, a mesma proporção de cursos por áreas de avaliação em cada agrupamento, com exceção dos *clusters* 3 (três) e 5 (cinco), em que se destacam as áreas Sistemas de Informação e Análise e Desenvolvimento de Sistemas, respectivamente. Por fim, no que concerne às demais características que compõem os grupos, tais como as proporções relacionadas à categoria administrativa, organização acadêmica, modalidade de ensino e região, estas informações podem ser encontradas na **Figura 13**.

Figura 13. Número total de cursos por área de avaliação e cluster



Fonte: dados da pesquisa (2024)

A análise da **Figura 13** revela que a maioria dos cursos é oferecida por instituições de ensino privadas, com ou sem fins lucrativos. Além disso, esses cursos são predominantemente ministrados em universidades, com foco na educação presencial e estão concentrados na macrorregião Sudeste do país. Notavelmente, o grupo com a maior média de Conceito ENADE, o *cluster 1* (um), é composto principalmente por instituições de ensino privadas, centros universitários e universidades localizadas na macrorregião Sudeste. Em contraste, o grupo com a média mais baixa, o *cluster 2* (dois), é constituído

por instituições privadas sem fins lucrativos, centros universitários e faculdades da mesma macrorregião.

Essas análises sugerem implicações socioeconômicas significativas. A preponderância de instituições de ensino privadas em cursos de nível superior pode indicar um acesso limitado à educação superior pública, geralmente mais acessível. A concentração na região sudeste pode refletir desigualdades regionais no acesso à educação de qualidade, com outras regiões potencialmente carecendo de opções educacionais similares. Além disso, a associação do baixo Conceito ENADE com instituições privadas em alguns *clusters* levanta questões sobre a qualidade do ensino nessas instituições e destaca a necessidade de uma avaliação rigorosa e melhoria da qualidade da educação superior em todo o país.

Análise crítica dos principais resultados

A análise dos resultados do ENADE 2021 na área de computação revela *insights* importantes sobre a distribuição, qualidade e desempenho dos cursos, bem como a influência de variáveis nas avaliações. Por meio de técnicas como análise exploratória, regressão múltipla e clusterização, identificou-se a necessidade de ampliar a quantidade de cursos avaliados, especialmente em Ciência da Computação (Licenciatura). As médias do Conceito ENADE variam entre as áreas, mas a maioria dos cursos está em faixas médias, indicando desafios na qualidade acadêmica.

A predominância de instituições privadas no ensino superior aponta para limitações no acesso à educação pública, enquanto a concentração de cursos no Sudeste sugere desigualdades regionais no acesso à qualidade educacional, apesar de regiões como Sul e Nordeste terem médias mais altas. A associação do baixo Conceito ENADE às instituições privadas destaca a necessidade de melhorias na qualidade da educação superior.

Por fim, a regressão múltipla identificou variáveis significativas para explicar a variação do Conceito ENADE, contribuindo para a clusterização que resultou em cinco *clusters* distintos de cursos, cada um com características específicas de desempenho e composição. Esses resultados apontam para a importância de estratégias específicas para aprimorar a qualidade da educação superior em todo o país.

CONSIDERAÇÕES FINAIS

A análise abrangente dos dados referentes aos cursos de computação no Exame Nacional de Desempenho de Estudantes (ENADE) de 2021 oferece uma visão aprofundada do panorama educacional nessa disciplina, fornecendo *insights* significativos com potencial para influenciar diretamente diversos aspectos sociais, educacionais, gerenciais e político-públicos. A relevância deste estudo reside na sua capacidade de identificar tendências, desafios e oportunidades que podem contribuir para o aprimoramento dos cursos de computação no contexto brasileiro.

Inicialmente, a análise exploratória dos resultados destaca percepções relevantes sobre o cenário educacional desses cursos. Além de evidenciar demandas específicas por determinados tipos de cursos, observa-se uma distribuição heterogênea nos diversos domínios de avaliação, com variações marcantes nos níveis médios de desempenho do Conceito ENADE entre eles. A concentração dos cursos em faixas médias de conceito sugere desafios comuns relacionados à qualidade acadêmica e aos métodos de ensino.

Ademais, a análise revela a influência de variáveis como a natureza administrativa das instituições, a organização acadêmica, o tipo de modalidade educacional e a localização geográfica no desempenho dos cursos, o que sublinha a importância de políticas e investimentos direcionados para aprimorar a qualidade do ensino de computação em todo o país. Tais descobertas são fundamentais para nortear estratégias de melhoria e garantir que os cursos atendam às demandas do mercado de trabalho e contribuam para o desenvolvimento socioeconômico do Brasil.

Paralelamente, a análise multivariada dos cursos de computação no ENADE de 2021 proporciona uma compreensão mais profunda da complexidade e das inter-relações entre as diversas variáveis que influenciam o desempenho acadêmico nessa área. A utilização de técnicas como a regressão múltipla permite identificar variáveis significativas na explicação da variação do Conceito ENADE, fornecendo *insights* sobre a importância de fatores como a pontuação padronizada em conhecimentos específicos, formação geral, regime de trabalho e indicadores de qualidade institucional.

Além disso, a análise de clusterização dos cursos possibilita a identificação de grupos distintos com base em suas características intrínsecas, destacando padrões e tendências antes não perceptíveis nos dados brutos. Tais resultados não apenas oferecem

orientações precisas para a melhoria da qualidade do ensino de computação, mas também podem embasar o desenvolvimento de políticas educacionais mais eficazes, adaptadas às necessidades específicas de cada grupo de cursos e região geográfica. Essa abordagem analítica, fundamentada em técnicas multivariadas, representa uma ferramenta poderosa para a gestão educacional, permitindo uma tomada de decisão mais embasada e estratégica para promover o avanço contínuo do ensino de computação no Brasil.

A aplicação prática desses resultados pode trazer benefícios não apenas para as instituições de ensino, mas também para os estudantes, empregadores e a sociedade em geral. As instituições de ensino podem utilizar esses *insights* para desenvolver programas acadêmicos mais alinhados com as demandas do mercado de trabalho e as necessidades dos estudantes. Os empregadores podem se beneficiar ao recrutar profissionais mais qualificados e preparados para enfrentar os desafios do mundo digital. Por sua vez, os estudantes podem fazer escolhas melhor subsidiadas sobre suas carreiras e se prepararem melhor para o mercado de trabalho.

Ademais, gestores educacionais e formuladores de políticas podem utilizar esses resultados para orientar investimentos, políticas públicas e iniciativas de capacitação profissional, visando a promover o desenvolvimento sustentável e a inclusão social por meio da educação em computação. Isso pode incluir a expansão de programas de bolsas de estudo, o estímulo à pesquisa e inovação tecnológica e o fortalecimento da infraestrutura educacional em regiões menos favorecidas.

Em síntese, este trabalho oferece uma base sólida para a tomada de decisões informadas e a implementação de medidas concretas visando à melhoria contínua da qualidade e relevância dos cursos de computação no Brasil, conforme delineado nos objetivos estabelecidos. Ao alinhar o ensino às demandas do mercado e às necessidades da sociedade, é possível preparar uma força de trabalho mais capacitada e adaptável, capaz de impulsionar a inovação, o crescimento econômico e o desenvolvimento social em todo o país. Em agendas de pesquisas futuras, recomenda-se a expansão de análises de desempenho educacional do ensino superior para as demais áreas de conhecimento, bem como a atualização contínua na área de computação, considerando os novos dados disponíveis a partir dos próximos exames ENADE.

REFERÊNCIAS

ALVES, R. M.; MATOS, A. M.; CANEDO, E. D. Plataforma de visualização dos dados minerados do ENADE dos cursos de Computação nos anos de 2008 a 2014. **Singular Engenharia, Tecnologia e Gestão**, v. 1, n. 2, p. 44–54, 19 out. 2019.

AMARO, E. F. A. **Os Algoritmos C4.5 e Hoeffding Tree Aplicados a Mineração de Dados Educacionais Referente ao Exame Nacional de Desempenho de Estudante (ENADE) em Ciência da Computação**. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade do Extremo Sul Catarinense, UNESC. Criciúma, p. 91. 2019.

ARAUJO, L. R. **Classificação Automática de Questões de Provas: Análise Comparativa de Algoritmos e Aplicação ao ENADE**. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Federal de Santa Maria, UFSM. Santa Maria, p. 50. 2021.

BANIN, S. L. **Python 3 Conceitos e Aplicações: Uma Abordagem Didática**. 1. ed. São Paulo, Saraiva Educação SA, 264p, 2018.

BARBOSA, P. L. S.; DAMAZIO, G. N. D. O.; DE CARVALHO, W. V.; CARMO, R. **O sucesso não é apenas uma questão de sorte: um mapeamento sistemático sobre técnicas de análise do Enade da área de Computação**. In: Anais do III Simpósio Brasileiro de Educação em Computação. SBC. p. 59-68, 2023a.

BARBOSA, P. L. S.; DE CARVALHO, W. V. **O uso continuado de ferramenta computacional baseada em questões do ENADE para a avaliação diagnóstica de alunos do curso de Sistemas de Informação: um estudo de caso**. In: Anais do XXXI Simpósio Brasileiro de Informática na Educação. SBC, p. 1743-1752, 2020.

BARBOSA, P. L. S.; DE CARVALHO, W. V.; CARMO, R. **Testes Adaptativos de preparação para o Enade em cursos de Computação**. In: Anais Estendidos do III Simpósio Brasileiro de Educação em Computação. SBC. p. 34-35, 2023b.

BARROSO, Lúcia P.; ARTES, Rinaldo. Análise multivariada. **Lavras: Ufla**, p. 151, 2003.

BRASIL. Lei nº 10.861, de 14 de abril de 2004. Institui o Sistema Nacional de Avaliação da Educação Superior – SINAES e dá outras providências. Diário Oficial da República Federativa do Brasil, Brasília, DF, 15 abr. 2004. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/l10.861.htm>. Acesso em: 12 mar. de 2024.

BRASIL. Portaria nº 23, de 21 de dezembro de 2017. Diário Oficial da União, Brasília, DF. Disponível em: <https://www.in.gov.br/materia/-/asset_publisher/KujrwoTZC2Mb/content/id/39380012/do1-2018-09-03-portaria-normativa-n-23-de-21-de-dezembro-2017--39379864>. Acesso em: 12 mar. de 2024.

CANDÃO, J. DE P.; REAL, E. M. **Uma análise do perfil e desempenho de estudantes de Computação através da Mineração de Dados**. Anais do Computer on the Beach, v. 8, p. 529-530, 3 maio 2017.

CAPELARI, L. O. O. **Predição do desempenho no ENADE dos discentes de computação**. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Tecnológica Federal do Paraná, UTFPR, Campo Mourão. 2022.

CAPELARI, L. O. O.; SCHWERZ, A. L. **O perfil socioeconômico dos concluintes de computação do sul do Brasil**. Anais do Computer on the Beach, v. 12, p. 133-140, 2021.

CHARÃO, A. S.; WIECHORK, K.; RODRIGUES, M. L. S.; BARBOSA, F. P. **Explorando Resultados por Questão no Enade em Ciência da Computação para Subsidiar Revisão de Projeto Pedagógico de Curso**. In: Anais do XXVIII Workshop sobre Educação em Computação (WEI), 2020, Cuiabá. Porto Alegre: Sociedade Brasileira de Computação, p. 16-20, 2020.

COELHO, A. C.; CUNHA, J. V. A. Regressão linear múltipla. **Análise multivariada: para os cursos de administração, ciências contábeis e economia**. São Paulo: Atlas, p. 131-231, 2007.

CORTELAZZO, A. L.; JÚNIOR, D. P.; RODRIGUES, M. R. J. B. Resultados do Exame Nacional do Desempenho de Estudantes – ENADE: quantitativos históricos na área da Computação e qualidade dos cursos. **Brazilian Journal of Development**, v. 7, n. 1, p. 865-886, 2021.

COSTA, E. *et al.* Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações. **Jornada de Atualização em Informática na Educação**, v. 1, n. 1, p. 1-29, 2012.

CREPALDE, N. J. B. F.; SILVERA, L. Desempenho universitário no Brasil: estudo sobre desigualdade educacional com dados do Enade 2014. **Revista Brasileira de Sociologia**, v. 4, n. 7, p. 211-240, 2016.

CUNHA, R.; SALES, C.; SANTOS, R. **Análise automática com os microdados do enade para melhoria do ensino dos cursos de ciência da computação**. In: Anais do XXIX Workshop sobre Educação em Computação. Porto Alegre: Sociedade Brasileira de Computação, p. 208-217, 2021.

DAMAZIO, G. N. D. O.; DE CARVALHO, W. V. **Gamificando a Preparação para o Enade em Cursos Superiores de Computação**. In: Anais Estendidos do III Simpósio Brasileiro de Educação em Computação. Sociedade Brasileira de Computação, p. 44-45, 2023.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37-37, 1996.

FIGUEIRO, M.; VISTA, N.; BARASUOL, J.; CHICON, P.; ANSUJ, A. Análise de Agrupamento Hierárquico aplicada aos microdados do ENADE do curso de graduação em Ciência da Computação. **Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação**, v. 1, n. 8, 2018.

FRANCELINO, W. L.; MACHADO, L. S. **Mineração de dados nos microdados ENADE Computação**. Relatório técnico, Faculdade Ânima Educação, Tubarão, SC, Brasil, 2020.

FRANKLIN, L. A. S.; EUCLYDES, F. M.; CAMPOS, A. P. T.; FERREIRA, M. A. M. Tecnologias da Informação e Comunicação no contexto democrático brasileiro: uma revisão sistemática de literatura. **Em Questão**, v. 28, n. 2, p. 265-294, 2022.

FREITAS, B.; COSME, L.; NASCIMENTO, M. **Exame Nacional de Desempenho de Estudantes (ENADE): Análise do Perfil das mulheres dos cursos da área de computação**. In: Anais do XIII Women in Information Technology. Sociedade Brasileira de Computação. p. 179-183, 2019.

HINTERHOLZ, O.; VALENZUELA, G.; DA SILVA, W.; FREIRE, A. **Transversal: uma Abordagem Sistêmica para Predição de Desempenho de Estudantes no ENADE por meio do uso do Ciclo PDCA**. In: Anais do XXII Workshop sobre Educação em Computação, p. 140-149. 2014.

INEP. Indicadores de Qualidade da Educação Superior. 2023a. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-de-qualidade-da-educacao-superior>>. Acesso em: 12 mar de 2024.

INEP. Nota Técnica N° 2/2023/CEI/CGGI/DAES: Processo N° 23036.003401/2023-12. Brasília, 2023b. Disponível em: <https://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2023/nota_tecnica_n_2_2023_cei_cggi_daes_descricao_da_metodologia_de_calculo_do_conceito_enade.pdf>. Acesso em: 12 mar. 2024.

INEP. Nota Técnica N° 29/2023/CEI/CGGI/DAES: Processo N° 23036.003401/2023-12. Brasília, 2023c. Disponível em: <https://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2023/nota_tecnica_n_29.pdf>. Acesso em: 12 mar. 2024.

INEP. Nota Técnica N° 4/2023/CEI/CGGI/DAES: Processo N° 23036.003401/2023-12. Brasília, 2023d. Disponível em: <https://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2023/nota_tecnica_n_4_2023_cei_cggi_daes_descricao_da_metodologia_de_calculo_do_cpc.pdf>. Acesso em: 12 mar. 2024.

INEP. Nota Técnica N° 5/2023/CEI/CGGI/DAES: Processo N° 23036.003401/2023-12. Brasília, 2023e. Disponível em: <https://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2023/nota_tecnica_n_5_2023_cei_cggi_daes_descricao_da_metodologia_de_calculo_do_igc.pdf>. Acesso em: 12 mar. 2024.

KUSS, A. C.; MEDAGLIA, J. Turismo e tecnologia da informação: das agências tradicionais às travel techs. **Revista Brasileira de Pesquisa em Turismo**, v. 16, p. e-2668, 2023.

LANDES, F. B.; MANHÃES, L. M. B. Análise dos exames do ENADE para os cursos de computação utilizando o software R. **Revista do Seminário Internacional de Estatística com R**, v. 3, n. 1, 2018.

LIMA, P. D. S. N.; AMBRÓSIO, A. P. L.; DOS SANTOS OLIVEIRA, J. L.; DE CARVALHO, C. L. Análise de conteúdo das provas do Enade para os alunos do curso de Bacharelado em Ciência da Computação. **Revista Brasileira de Informática na Educação**, v. 29, p. 385-413, 2021.

LIMA, P. S. N.; AMBRÓSIO, A. P. L.; FERREIRA, D. J.; BRANCHER, J. D. Análise de dados do Enade e Enem: uma revisão sistemática da literatura. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 24, n. 1, p. 89-107, 2019.

LUCIANO, E. M.; FANTINEL, L. M.; LÜBECK, R. M. O impacto da institucionalização da governança de Tecnologia da Informação no desempenho da governança de TI e o papel do Job Crafting. **Revista Administração em Diálogo-RAD**, 2021.

MEC. Ministério da Educação. Portaria Normativa nº 23, de 20 de dezembro de 2016. Altera dispositivos da Portaria Normativa MEC nº 40, de 12 de dezembro de 2007, republicada em 2010, para efeitos imediatos nos processos de cálculo e divulgação dos Indicadores de Qualidade da Educação Superior, a partir da edição de 2015. Brasília, 2016. Disponível em: <<https://abmes.org.br/arquivos/legislacoes/portaria-normativa-n-23-2016-12-20.pdf>>. Acesso em: 12 de mar. de 2024.

MEC. Ministério da Educação. Cadastro Nacional de Cursos e Instituições de Educação Superior Cadastro e-MEC. Disponível em: <<https://emec.mec.gov.br/>>. Acesso em: 12 mar. de 2024.

MEDRI, W. Análise exploratória de dados. **Londrina: Universidade Estadual de Londrina**, v. 8, n. 8, p. 151-170, 2011.

MENOLLI, A.; BRAGAGNOLO, A.; GONÇALVES, R. F.; NETO, J. C. Factors Related to the Quality of Computer Science Teacher Education Courses: A Public Data Analysis. **Revista Brasileira de Informática na Educação**, v. 29, p. 209-226, 2021.

MERLOTTO, W. F.; MOORI, R. G.; LOPES, Y. Tecnologia da Informação como um alavancador de desempenho em prestadores de serviços logísticos. **Revista Alcance**, v. 28, n. 1, p. 52-66, 2021.

OLIVEIRA, F. M. de.; PROLO, C. A.; ZORZO, A. F.; DE ANDRADE, G. K. **ENADE Comentado-Computação: Compartilhando Contribuições para a Avaliação de Cursos**. In: INOVAÇÃO, UNIVERSIDADE E RELAÇÃO COM A SOCIEDADE: Boas Práticas na PUCRS, 2009.

OLIVEIRA, F. R. de.; KLEINA, M.; MARQUES, M. A. M.; GAYER, J. A. C. A.; TAMACHIRO, T. S. O. **Clusterização de Clientes: Um Modelo Utilizando Variáveis Categóricas e Numéricas**. X Congresso Brasileiro de Engenharia de Produção: ConBRepro, Abepro, 2020.

PARK, H.; JUN, C. A simple and fast algorithm for K-medoids clustering. **Expert systems with applications**, v. 36, n. 2, p. 3336-3341, 2009.

PEREIRA JÚNIOR, A. Auto-Organização Social no séc. XXI: dinheiro, tecnologia da informação e consciência. **Acta Scientiarum. Human and Social Sciences**, v. 42, n. 3, 2020.

PRODANOV, C. C.; FREITAS, E. C. de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico** - 2ª Edição. Editora Feevale, 2013.

ROCHA, P. C. da S.; JUCÁ, S. C. S.; SILVA, S. A. da. A evolução das Tecnologias da Informação e Comunicação na perspectiva de Touraine, Bell e Castells. **Research, Society and Development**, v. 8, n. 5, p. e1885928, 1 mar. 2019.

ROSA, E. R.; FERREIRA, D. J.; SILVA, N. F. F.; ASSIS, A. Estudo exploratório através de análises longitudinais aplicado à ciência da computação a partir da base de dados do ENADE. **Revista Brasileira de Informática na Educação**, v. 29, p. 1463-1486, 2021.

SANTOS, C. A. P. dos. **Uma Análise Exploratória da Influência dos Projetos Pedagógicos dos Cursos Superiores no Resultado do Enade por meio de Mineração de Textos e Aprendizado de Máquina**. 2022. Dissertação (Mestrado em Ciência da Computação) - Faculdade de Computação, Universidade Federal de Mato Grosso do Sul, Campo Grande, 2022.

SILVA, A. F.; HOED, R. M.; SARAIVA, P. F. **Análise do desempenho dos alunos de cursos superiores em computação no enade – uma abordagem usando mineração de dados**. Conferências IADIS Ibero-Americanas WWW/Internet e Computação Aplicada, v. 2019, p. 207, 2019.

SILVA, L. C. D.; CAMPANELLI, A.; SILVA, L. S. D.; SILVA, T. V. F. D.; SILVA, R. C. D.; GARCIA, R. D. R.; MAGALHÃES, F. L. F. D. Graduação em Computação no Brasil: perspectiva a partir do Exame Nacional de Avaliação de Desempenho. **Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología**, (25), 57-67, 2020.

SILVA, L. C. da; CAMPANELLI, A.; SILVA, L. S. da; SILVA, T. V. F. da; SANTOS, P. V. A. dos. **Série histórica dos conceitos do ENADE em cursos de TI: houve mudanças ao longo de três ciclos avaliativos?** In: Anais do XVI Simpósio de Excelência em Gestão e Tecnologia, 2018.

SILVEIRA, M. E. R.; BOLL, D. C. C.; MATOZINHO, F. L.; NEGRIZOLI, I. F.; VANZIN, L.; CAMARA, M. K.; OYAMADA, M. S. Classificação por matérias das questões do Enade e POSCOMP. **Revista ComInG-Communications and Innovations Gazette**, v. 5, n. 2, p. 9-19, 2021.

SOARES, R. A. M. D. C.; ALBUQUERQUE, T. V. D.; MENDES-FILHO, L.; ALEXANDRE, M. L. Revisão sistemática da produção científica brasileira sobre turismo e tecnologia da

informação e comunicação (TIC). **Revista Brasileira de Pesquisa em Turismo**, v. 16, e-2629, 2023.

VISTA, N. P. B.; FIGUEIRÓ, M. F.; CHICON, P. M. M. **Técnicas de mineração de dados aplicadas aos microdados do ENADE para avaliar o desempenho dos acadêmicos do curso de Ciência da Computação no Rio Grande do Sul utilizando o software R**. I Seminário de Pesquisa Científica e Tecnológica, v. 1, n. 1, 2017.

NOTA DE REVISÃO

Este estudo teve revisão de linguagem realizada por profissional habilitado, Maurício Lima Bastos, Graduado em Letras Português e Inglês pela Universidade Federal do Ceará (UFC) e Mestre em Letras pela Universidade Estadual do Ceará (UECE).

NOTA DE AUTORIA

1. Josicleyton Azevedo dos Santos: Concepção da ideia, curadoria de dados, metodologia, escrita do rascunho original, revisão e edição do manuscrito, administração do projeto.
2. Macilene Maria Monteiro Maia: Análise formal, escrita do rascunho original, revisão e edição do manuscrito.
3. Marcos Filho Lima Bastos: Análise formal, escrita do rascunho original, revisão e edição do manuscrito.
4. Mileno Alexandre Barbosa Epifanio: Análise formal, escrita do rascunho original, revisão e edição do manuscrito.
5. Fábio Francisco da Costa Fontes: Supervisão do projeto, análise formal, escrita do rascunho original, revisão e edição do manuscrito.

Recebido em: 20/02/2024

Parecer em: 24/05/2024

Aprovado em: 10/06/2024