

Análise Automatizada de Dados do ENADE com aplicação para os Cursos de Ciência da Computação

Daniel E. S. Trujillo¹, Jan M. Corrêa¹

¹Departamento de Ciência da Computação (CIC) (UnB)
Campus Darcy Ribeiro, Asa Norte – 70.910-900 – Brasília – DF
Universidade de Brasília - UnB

danielerick.sancheztrujillo@gmail.com, jancorrea@unb.br

Abstract. *This study demonstrates the use of a new tool created to perform automated analysis of ENADE 2021 data. It shows how the tool was applied, comparing the Computer Science (Bachelor's Degree) course at the University of Brasília (UnB) with other courses. This work differs from previous ones in that it provides a ranking of relative performance per question within a group of selected institutions. This allows for potentially more detailed, more specific, and relevant comparisons between educational institutions. It also differs in that it provides a direct comparison between two institutions in the distribution of responses per question, allowing for a better understanding of student performance and the detection of biases in responses. With this tool, course coordinators will be able to perform more specific and detailed analyses than those contained in the official ENADE reports.*

Resumo. *Este estudo mostra a utilização de uma nova ferramenta criada para fazer análise automatizada dos dados do ENADE 2021. Mostra como a ferramenta foi aplicada, comparando o curso Ciência da Computação (Bacharelado) da Universidade de Brasília (UnB) com outros cursos. Este trabalho se diferencia dos anteriores por prover o ranqueamento de desempenho relativo por questão dentro de um grupo de instituições selecionadas. Isto permite comparações potencialmente mais detalhadas, mais específicas e relevantes entre as instituições de ensino. Diferencia-se também por prover a comparação direta entre duas instituições na distribuição das respostas por questão, permitindo um entendimento melhor do desempenho dos alunos e a detecção de vieses nas respostas. Com esta ferramenta, os coordenadores de curso poderão fazer análises mais específicas e detalhadas do que as que constam nos relatórios oficiais sobre o ENADE.*

1. Introdução

O ENADE (Exame Nacional de Desempenho dos Estudantes) é uma iniciativa do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) [INEP 2023] que avalia estudantes concluintes de cursos de graduação, analisando seus conhecimentos, habilidades e competências adquiridas durante a formação. Esse exame não apenas mensura o desempenho dos estudantes, mas também serve como base para os indicadores de qualidade do ensino superior no país.

Para auxiliar nesta análise, o INEP [INEP 2023] disponibilizou um documento com a análise de desempenho para o curso de computação avaliado pelo ENADE em

2021. Nesta análise consta o percentual de acerto por questão do curso comparado com a sua unidade da federação, região e nacionalmente. Este percentual de acerto pode ser útil, mas ele mostra basicamente se o desempenho do curso está abaixo ou acima da média, mas não indica o quanto.

Se, por exemplo, um coordenador de curso de uma universidade quiser comparar o desempenho relativo do seu curso com as 20 melhores universidades, os documentos de ENADE não possuem esta informação pois só comparam por média na localização geográfica. Se ele quiser saber se o seu curso está entre os 10% melhores, os documentos do ENADE também não possuem esta informação pois só apresentam a média. Um outro coordenador de uma instituição com poucos recursos financeiros pode querer comparar o seu curso com instituições com os mesmos níveis de recursos, mas os documentos do ENADE não separam por quantidade de recursos. Os dois coordenadores conseguiriam fazer as comparações com a ferramenta descrita neste trabalho.

Outras formas de utilizar a ferramenta descrita neste trabalho seriam: se um diretor de uma rede de universidades particulares quiser comparar o desempenho dos cursos da mesma universidade, mas em estados diferentes. Ele não conseguiria isso com os documentos do ENADE, mas poderia fazer isso com a ferramenta descrita neste trabalho.

Se um professor quiser comparar a sua disciplina com outras semelhantes em outras instituições, ele pode encontrar uma questão do ENADE que cubra especificamente o assunto da sua disciplina e comparar com instituições que ele sabe que têm disciplinas semelhantes para tentar ter uma ideia do seu desempenho relativo. Isto não é mostrado nos documentos do ENADE, mas poderia ser feito utilizando a ferramenta aqui descrita.

As possibilidades destas e várias outras aplicações justificam a criação de uma ferramenta que seja capaz de fazer essas análises de forma automatizada, provendo informações mais específicas do que as disponíveis nos documentos do ENADE.

Este trabalho mostra a utilização de uma nova ferramenta que utiliza os dados do ENADE (Exame Nacional de Desempenho dos Estudantes) [INEP 2023] para criar comparações de desempenho entre cursos que não estão disponíveis nos documentos do ENADE. Esta ferramenta permite escolher quais instituições serão comparadas e fazer um ranqueamento relativo para cada questão.

Como exemplo de aplicação prática desta ferramenta, utilizou-se a edição 2021 do ENADE para avaliar os alunos do curso de 'Ciência da Computação bacharelado' da UnB em relação ao mesmo curso em outras instituições de ensino selecionadas.

Este artigo foi organizado da seguinte forma: na seção 2 é investigado se existiam trabalhos que desenvolveram uma ferramenta ou fizeram uma análise semelhante à ferramenta descrita neste trabalho. A seção 3 mostra como os dados foram coletados e processados. A seção 4 mostra um exemplo de aplicação da ferramenta para um caso prático. A seção 5 discorre sobre as principais conclusões. Na seção 6 são propostos trabalhos futuros.

2. Trabalhos relacionados

Uma vez justificada na seção anterior a utilidade da criação de uma ferramenta de análise automatizada que permita análises não contempladas nos documentos do ENADE, procuramos na literatura pela existência de trabalhos semelhantes.

Esta seção tem como objetivo prover uma descrição resumida de cada trabalho e identificar se foi desenvolvida uma solução que permitisse a análise automatizada similar à descrita na seção anterior.

Os trabalhos analisados foram escolhidos consultando o site da Biblioteca Digital da Sociedade Brasileira de Computação [de Computação 2024]. Esta plataforma foi escolhida pela grande quantidade e excelente qualidade dos artigos, e também por ser gratuita. Foram pesquisados todos os documentos em todas as bases contendo a palavra “ENADE”. No momento da procura, foram retornados 32 documentos. Destes, foram selecionados todos os documentos cujo foco foi a análise de dados do ENADE. Destes, foram excluídos documentos que não eram de área relacionada à computação. Assim, com estes critérios, foram analisados 10 artigos no total.

Em [Gondran et al. 2023], foram analisadas quais características pessoais e socioeconômicas contribuem na explicação do desempenho de concluintes de graduação em Engenharia no ENADE. Foram encontrados dentre 14 atributos importantes: idade, situação de trabalho, bolsas, financiamentos e escolaridade dos pais. Verificou-se também que o perfil dos bons estudantes foi mudando ao longo do tempo.

Em [Vieira et al. 2022], procurou-se descobrir como diferentes atributos demográficos afetaram o desempenho dos estudantes no ENADE.

No artigo [Cunha et al. 2021], foi feita uma análise automática para os cursos de Ciência da Computação, mostrando o desempenho nas disciplinas do curso em diferentes anos de aplicação do ENADE. Compararam o desempenho com a média dos cursos no ENADE.

Em [Barbosa et al. 2023], foi feito um mapeamento sistemático de artigos que analisaram os dados do ENADE. O principal foco dos artigos analisados foi a análise de fatores que impactavam nas notas dos alunos.

No artigo [Charão et al. 2020], foram utilizados os relatórios oficiais do INEP sobre o ENADE para identificar pontos de melhoria para o curso. Relata o desafio de classificar as questões por conteúdo e relacioná-las com disciplinas do curso. Realizaram a comparação também com a média dos cursos no ENADE.

Em [Rendeiro et al. 2023] foi feita uma análise do desempenho dos alunos de Engenharia da Computação e, baseado nesta análise, foram feitas sugestões de melhoria do projeto pedagógico do curso. Também realizaram a comparação de acertos por tema com a média dos cursos no ENADE.

Os artigos [Leão et al. 2021] e [Freitas et al. 2019] foram utilizados os dados do ENADE para traçar o perfil das mulheres nos cursos, mas não o desempenho em si.

No artigo [Lima et al. 2021] foi feita uma análise de desempenho do curso de Ciência da Computação no ENADE, levando em conta o conteúdo de cada questão e mapeando para o conteúdo pedagógico relacionado. Os conteúdos das questões foram mapeados com os índices de facilidade e de discriminação, tentando descobrir a dificuldade com que cada conteúdo foi cobrado no exame. Realizaram análise de percentual de acerto por tema.

Em [Rosa et al. 2021] foi feita mineração de dados para classificar o desempenho

dos alunos baseado nas informações sobre os alunos. Descobriu-se que fatores socioeconômicos têm impacto no desempenho.

No artigo [Ramos et al. 2024] foi feita a extração de informações não triviais para o curso de Redes de Computadores comparando com o desempenho médio nacional dos cursos. Foi feita também uma análise dos questionários de percepção da prova.

Analizando estes artigos, verificou-se que poucos realizaram uma análise automatizada. Mesmo os que a fizeram, como [Cunha et al. 2021] e [Ramos et al. 2024], fizeram a comparação com a média do desempenho dos cursos. Uma comparação com a média nacional dos cursos compara instituições de diferentes tamanhos, localidades, orçamentos e tipos (particulares, federais e estaduais). Um coordenador de curso pode querer comparar sua instituição com instituições que considera semelhantes, mas esse tipo de análise é muito específico para constar nas tabelas de comparação divulgadas pelo ENADE.

Assim, em todos os trabalhos analisados, nenhum fez uma análise de desempenho relativo por questão comparado com um grupo selecionado de instituições. Desta forma, justifica-se a criação de uma ferramenta que faça esta análise de forma automatizada.

3. Coleta e Análise dos Dados

Neste trabalho foi desenvolvida uma ferramenta que permite selecionar um grupo de instituições e verificar o desempenho relativo por questão objetiva. Para se fazer uma comparação mais equitativa entre instituições de ensino, pode-se comparar instituições semelhantes, pois fatores como renda, desigualdades sociais, desigualdades regionais, entre outros, podem afetar o desempenho dos estudantes [Vendrami et al. 2021].

Assim, o usuário da ferramenta pode escolher para a análise o grupo de instituições pelo critério de similaridade entre instituições ou outro que desejar.

Além disto, foi desenvolvida a comparação direta entre duas instituições, comparando questão por questão, mostrando os desempenhos relativos.

Esta ferramenta pode servir para analisar qualquer curso avaliado no ENADE de 2021. A utilização da ferramenta em dados mais recentes pode requerer alguma adaptação em virtude de possíveis mudanças no formato dos dados por parte do INEP.

Devido à sua eficiência e disponibilidade, esta ferramenta foi desenvolvida utilizando a linguagem de programação Python (versão 3.7.5) [G. Van Rossum, F. L. DRAKE 2019] com o Jupyter Notebook [PROJECT JUPYTER 2023]. Foram também utilizadas as bibliotecas Pandas, Matplotlib, Seaborn, NumPy e Unidecode (utilizada para tratar palavras acentuadas).

3.1. Coleta dos Dados

Os microdados do ENADE [INEP 2023] são definidos como todos os dados divulgados para cada respondente individual sobre cerca de 489 mil estudantes divididos em 30 cursos que participaram do exame em 2021. Estes microdados contêm dados sobre os alunos, incluindo dados demográficos como idade, renda, tipo de ensino médio, notas obtidas e o curso correspondente. Eles também contêm as respostas dos alunos para todas as questões objetivas.

A prova do ENADE tem os seguintes dois componentes:

Formação geral (FG): é formada por um conjunto de 10 questões, das quais 8 são de múltipla escolha e 2 são questões discursivas. Seu propósito é avaliar a capacidade do aluno de compreender e refletir sobre questões que vão além da sua área de formação profissional. Os alunos devem demonstrar um entendimento crítico de questões contemporâneas.

Conhecimentos específicos (CE): Composta por 30 questões — 27 de múltipla escolha e 3 discursivas, esta seção visa avaliar os conhecimentos e habilidades específicas relacionadas à área de estudo do curso do aluno.

Neste trabalho, foram utilizados os resultados das questões de múltipla escolha por serem mais diretamente comparáveis.

A adequação dos microdados do ENADE à LGPD [BRASIL 2018] teve o objetivo de preservar a privacidade dos estudantes. Embora ainda seja possível realizar estudos sobre desempenho dos cursos, a nova estrutura impede a agregação de dados no nível do estudante individual. Assim, pode-se identificar o desempenho geral de um aluno, mas não relacionar com características de estudantes específicos. Por exemplo, pode-se ver o desempenho de um estudante na prova, mas não se pode relacionar com a renda do mesmo estudante.

Os microdados do ENADE estão disponíveis para *download* no site do INEP. Nele foram obtidos os cadernos de questões, dicionários de dados que explicam o significado de cada campo, bem como os dados em si.

3.2. Processamento dos Dados

Esta seção mostra o processamento dos dados realizado pela ferramenta.

Os microdados do ENADE não podem ser analisados diretamente para comparar o desempenho dos cursos. Para isto, é necessário fazer um pré-processamento dos dados.

O pré-processamento dos dados ocorreu da seguinte forma:

Foram baixados os dados do site do INEP [INEP 2023]. Estes dados, forma de tabela, contêm cerca de 492 mil linhas válidas, cada uma correspondendo a um aluno inscrito nos 30 cursos avaliados pelo ENADE.

Foram filtrados os cerca de 370 mil respondentes que realmente realizaram a prova.

Foram filtrados apenas os alunos de um curso específico para serem comparados, por exemplo, Ciência da Computação.

Cada uma das linhas contém 159 colunas, cada uma correspondendo aos dados demográficos (ex: idade), desempenho no exame e questionário sobre percepção da prova, entre outros.

Como a análise se concentrou no desempenho, foram retirados os dados demográficos de cada respondente. Logo, foram filtradas apenas as colunas com identificador do aluno, curso, instituição de ensino e desempenho nas questões objetivas.

Assim, nos dados utilizados, cada linha na tabela corresponde a um aluno, seu curso, número da instituição de ensino e como ele respondeu a cada uma das questões de múltipla escolha no ENADE. Os dados originais do ENADE não identificam as

instituições de ensino pelo nome, mas pelo seu código. Por exemplo, a UnB não aparece nos dados, mas apenas o seu código. Isto dificulta bastante a análise por uma pessoa que não conhece as instituições pelo número.

Assim, foi criado um *script* em Python que utilizou a tabela de instituições de ensino do MEC, onde constam os códigos e nomes das instituições de ensino para adicionar, para cada aluno, o nome da sua respectiva instituição de ensino superior. Isto permitiu ao usuário da ferramenta selecionar as instituições para análise.

Uma vez que a tabela agora contém os nomes das instituições, o usuário pode selecionar pelo nome o grupo de instituições que deseja comparar. Por exemplo, uma lista com "Unb USP UFRJ UFMG" vai filtrar todos os alunos que pertencem a uma destas 4 instituições.

As respostas dos alunos nas questões objetivas não estão imediatamente disponíveis para avaliação pois as respostas estão dentro de uma *string* de respostas. Esta informação precisa ser comparada com uma *string* que contém o gabarito com a resposta correta. Assim, é necessário percorrer as *strings* um caractere de cada vez e verificar se houve acerto ou erro. Exemplo : o gabarito é "ABC.." e o aluno respondeu "ADE..", isto indica que o aluno acertou apenas a primeira questão dentre as 3 primeiras. Para fazer este processamento de *strings* apenas uma vez e economizar futuros processamentos, foram criadas 70 novas colunas. Duas colunas para cada uma das respostas do aluno em cada questão objetiva (8 da formação geral e 27 da formação específica). Para cada questão objetiva, uma coluna contém qual foi a resposta do aluno e outra se houve acerto ou não. Isto permite calcular rapidamente o percentual de acerto e a distribuição das respostas.

Em seguida, para cada questão objetiva, calculou-se o percentual de acerto médio por curso selecionado. Ou seja, para cada curso, calcula-se qual o percentual de alunos do curso que acertaram aquela questão. Por exemplo, para saber o percentual de acerto na questão 10, basta verificar quantos alunos têm valor 1 na coluna correspondente. Depois, faz-se o ranqueamento relativo entre os cursos selecionados para cada uma das questões do ENADE. A ferramenta também tem a funcionalidade de fazer uma comparação detalhada entre duas instituições, indicando qual acertou mais ou menos para cada questão.

Do ponto de vista do usuário, ele seleciona as instituições de ensino que deseja analisar e a ferramenta é capaz de gerar automaticamente as tabelas com os resultados das análises.

4. Resultados

Como resultados, serão apresentadas análises geradas pela ferramenta para uma aplicação prática na UnB. Neste caso, o coordenador do curso de Ciência da Computação da UnB desejou comparar o desempenho do curso da UnB com cursos em instituições semelhantes.

Para fazer a comparação utilizando os microdados do ENADE, o curso de Ciência da Computação (Bacharelado) contou com 72 instituições avaliadas. O coordenador não achou justo comparar a UnB com todas estas 72 Instituições de Ensino Superior (IES) como consta nos documentos do ENADE, pois compararia com instituições menores e com menos recursos. Então, destas 72 IES, o coordenador selecionou para comparação os 20 primeiros cursos ordenados pelo seu conceito no ENADE. Ele teve como objetivo

comparar o seu curso com os melhores cursos do país.

A Tabela 1 mostra as instituições selecionadas conforme obtido no site do INEP, ordenadas por conceito no ENADE.

Tabela 1. Instituições escolhidas Ciência da Computação (Bacharelado)

Código de área	Área de Avaliação	Grupos Avaliados	Código do IES	Nome do IES*	Sigla do IES*	Organização Avaliadora	Conceito Enade (2015)
0004	Ciência da Computação	Bacharelado	575	UNIVERSIDADE FEDERAL DE MINAS GERAIS	UFMG	Universidade	5
0004	Ciência da Computação	Bacharelado	576	UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	UFRRN	Universidade	4.535
0004	Ciência da Computação	Bacharelado	8	UNIVERSIDADE FEDERAL DE VIÇOSA	UFV	Universidade	4.379
0004	Ciência da Computação	Bacharelado	578	UNIVERSIDADE FEDERAL DE JUIZ DE FORA	UFJF	Universidade	4.365
0004	Ciência da Computação	Bacharelado	581	UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL	UFRGS	Universidade	4.338
0004	Ciência da Computação	Bacharelado	571	UNIVERSIDADE FEDERAL DO PIAUÍ	UFPI	Universidade	4.175
0004	Ciência da Computação	Bacharelado	577	UNIVERSIDADE FEDERAL DE ALAGOAS	UFAL	Universidade	4.161
0004	Ciência da Computação	Bacharelado	579	UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO	UFES	Universidade	4.133
0004	Ciência da Computação	Bacharelado	582	UNIVERSIDADE FEDERAL DE SANTA CATARINA	UFSC	Universidade	4.088
0004	Ciência da Computação	Bacharelado	4025	FUNDAÇÃO UNIVERSIDADE FEDERAL DO ABC	UFABC	Universidade	4.064
0004	Ciência da Computação	Bacharelado	584	UNIVERSIDADE ESTADUAL DE CAMPINAS	UNICAMP	Universidade	4.057
0004	Ciência da Computação	Bacharelado	591	UNIVERSIDADE FEDERAL DE SÃO PAULO	UNIFESP	Universidade	3.995
0004	Ciência da Computação	Bacharelado	7	UNIVERSIDADE FEDERAL DE SÃO CARLOS	UFSCAR	Universidade	3.985
0004	Ciência da Computação	Bacharelado	586	UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO	UNESP	Universidade	3.980
0004	Ciência da Computação	Bacharelado	572	UNIVERSIDADE FEDERAL FLUMINENSE	UFF	Universidade	3.971
0004	Ciência da Computação	Bacharelado	588	UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	UFFRJ	Universidade	3.926
0004	Ciência da Computação	Bacharelado	583	UNIVERSIDADE FEDERAL DE SANTA MARIA	UFSM	Universidade	3.819
0004	Ciência da Computação	Bacharelado	271	UNIVERSIDADE DO OESTE PAULISTA	UNOESTE	Universidade	3.818
0004	Ciência da Computação	Bacharelado	2	UNIVERSIDADE DE BRÁSILIA	UNB	Universidade	3.818
0004	Ciência da Computação	Bacharelado	557	UNIVERSIDADE ESTADUAL DE MARINGÁ	UEM	Universidade	3.798

A seguir, será mostrado qual é o desempenho dos alunos da UnB no curso de Ciência da Computação (Bacharelado) quando comparado com o desempenho dos alunos dos cursos de Ciência da Computação (Bacharelado) das 20 instituições com melhor conceito no ENADE. Isto foi feito calculando a nota média dos alunos dos cursos.

Quando os *scripts* em Python são executados, são geradas as tabelas com o ranqueamento comparativo por questão. São geradas automaticamente as tabelas para a formação geral e formação específica. Neste trabalho, será mostrada em detalhes apenas a formação específica.

Algumas questões de múltipla escolha foram anuladas no ENADE e não são mostradas nas tabelas.

A Tabela 2 mostra um exemplo de como pode ser feita a comparação de desempenho por questão entre dois cursos. Neste exemplo, foram comparados os cursos de Ciência da Computação da UnB Bacharelado e Licenciatura. Esta comparação pode ser feita para a Formação Geral, pois as questões são as mesmas. A ferramenta permite comparar quaisquer dois cursos na parte da formação geral. Na formação específica, apenas onde os cursos têm as mesmas questões.

A Tabela 2 mostra para cada questão, o percentual de acerto para os cursos de Ciência da Computação da UnB Bacharelado e Licenciatura. A última coluna mostra o percentual relativo a mais de acertos que um curso tem sobre o outro. A cor verde indica que o Bacharelado acertou mais que a Licenciatura. A cor azul indica o contrário. Assim, pode-se ver os pontos fortes e fracos em relação à comparação dos dois cursos. Esta comparação pode ser repetida com quaisquer outros cursos.

A Tabela 3 representa o desempenho dos alunos de ciência da computação da UnB em relação a outras instituições selecionadas, em 27 questões objetivas da formação específica. As questões são identificadas como “forma_spec_q_x”, onde “x” representa o número da questão. Ao lado de cada questão, há uma porcentagem que indica o desempenho relativo dos alunos da UnB, onde 100% significaria que a UnB foi a IES com o melhor desempenho naquela questão e 0% indicaria o pior desempenho. A UnB destaca-se nas questões 2, 6, 16 e 18, com um desempenho maior ou igual a 80%. De forma geral, os dados refletem um desempenho mediano com média no desempenho de 50%.

Tabela 2. Comparação dos acertos (FG) da UnB - Ciência da Computação - Bacharelado e Licenciatura

Questões	Bacharelado	Licenciatura	Comparação Desempenho
Formação Geral Questão 1	60,00%	48,72%	23,16%
Formação Geral Questão 2	36,47%	41,03%	12,49%
Formação Geral Questão 3	71,76%	66,67%	7,65%
Formação Geral Questão 4	58,82%	38,46%	52,94%
Formação Geral Questão 5	71,76%	82,05%	14,33%
Formação Geral Questão 6	68,24%	56,41%	20,96%
Formação Geral Questão 7	71,76%	64,10%	11,95%
Formação Geral Questão 8	31,76%	30,77%	3,24%

4.1. Exemplo de Aplicação Prática

Uma vez que o coordenador do curso de Ciência da Computação utiliza a ferramenta e obteve a tabela 3, ele pode visualizar o desempenho da UnB em cada questão. Nas questões de formação específica, algumas questões foram anuladas pelo ENADE e não aparecem. No caderno de prova, as questões de formação específica começam do número 9. Assim, a questão que consta como 1 nos dados do ENADE é a questão 9 no caderno de prova e assim por diante. Para proceder a uma análise das questões, o coordenador precisou do caderno de prova para Ciência da Computação (bacharelado) que foi obtido em [INEP 2023]. Para saber para qual conteúdo do curso uma questão está relacionada, o coordenador teve de usar sua experiência e o seu conhecimento sobre o curso. Esta é uma parte subjetiva do processo de análise.

Como exemplo de comparação detalhada entre dois cursos, a Tabela 2 mostra que a maior diferença de desempenho na formação geral entre os alunos do bacharelado e da licenciatura em computação da UnB acontece na questão 4. O bacharelado tem um desempenho 52,94% maior que a licenciatura.

Na questão 4 no caderno de prova a alternativa “A” é semelhante ao gabarito “B” pois as duas dizem que as asserções I e II são verdadeiras. A diferença é saber se a asserção II é uma justificativa correta para a asserção I (opção ‘A’) ou não (opção ‘B’). A maioria achou que era uma justificativa válida quando não era. Para melhorar este problema, poderiam ser adicionados conteúdos nas disciplinas que mostram para os alunos o que constitui uma justificativa válida para um argumento.

Nas questões de formação específica do curso de Ciência da Computação, algumas questões foram anuladas pelo ENADE e não aparecem. No caderno de prova, as questões de formação específica começam do número 9. Assim, a questão 1 nos dados do ENADE é a questão 9 no caderno de prova e assim por diante. Na Tabela 3 que mostra a comparação de desempenho na formação específica no curso de Ciência da Computação da UnB em relação às 20 melhores cursos no ENADE. As questões onde o bacharelado apresentou desempenho igual ou superior a 80% das outras instituições foram as 2, 6, 16, 18. Estas questões são, respectivamente, sobre estrutura de dados (listas, pilhas e filas), arquitetura de computadores (arquitetura de Von Neumann), segurança de dados (criptografia) e desenvolvimento de software (interfaces adaptativas).

Também na Tabela 3 se pode notar quais foram os piores desempenhos comparativos nas questões. O coordenador listou para melhorias todas as questões com desempe-

Tabela 3. Comparação do desempenho (FE) da UnB - Ciência da Computação - Bacharelado com as 20 melhores selecionadas no ENADE

Número da Questão	Desempenho da UnB em relação a outras universidades
forma_espec_q_1	15.00 %
forma_espec_q_2	85.00 %
forma_espec_q_3	30.00 %
forma_espec_q_6	95.00 %
forma_espec_q_7	75.00 %
forma_espec_q_8	25.00 %
forma_espec_q_10	35.00 %
forma_espec_q_11	60.00 %
forma_espec_q_12	30.00 %
forma_espec_q_14	25.00 %
forma_espec_q_15	5.00 %
forma_espec_q_16	80.00 %
forma_espec_q_18	80.00 %
forma_espec_q_19	30.00 %
forma_espec_q_20	45.00 %
forma_espec_q_22	55.00 %
forma_espec_q_23	75.00 %
forma_espec_q_26	65.00 %
forma_espec_q_27	45.00 %
Média	50.26 %

no relativo abaixo de 50%. Verificando as questões no caderno de prova do ENADE, na questão 1 os alunos tiveram desempenho pior que 85% dos cursos comparados, é sobre sistemas operacionais. A questão 15 que tratava sobre estrutura de dados (árvore binária), o desempenho foi abaixo de 95% dos outros cursos. Na questão 8, com desempenho abaixo de 75% versa sobre circuitos lógicos. Na questão 14, com desempenho abaixo de 75% foi sobre Diagrama Entidade-Relacionamento (banco de dados).

Nas questões 3, 12 e 19 o desempenho foi abaixo de 70%. São respectivamente sobre redes neurais, linguagem C e análise estatística. Na questão 10, sobre aprendizado de máquina, o desempenho ficou abaixo de 65%. Nas questões 20 (arquitetura de computadores) e 27 (sistemas distribuídos) o desempenho foi abaixo de 55%.

Para tentar melhorar os desempenhos dos alunos, as disciplinas que cobrem estes conteúdos podem ser revisadas em relação aos conteúdos e à forma em que são lecionadas.

Por exemplo, uma das opções da questão 1 é sobre sistemas em tempo real, conteúdo que não é coberto por disciplinas obrigatórias do curso. Este tópico pode ser adicionado na ementa da disciplina de sistemas operacionais. Outro exemplo de sugestão

é reformular o ensino de inteligência artificial, pois em duas questões sobre este tema (3 e 10) o desempenho dos alunos foi abaixo da média das instituições selecionadas.

5. Conclusões

Neste trabalho foi utilizada uma ferramenta desenvolvida para fazer análises automatizadas utilizando *scripts* em Python sobre os microdados do ENADE do ano 2021. A ferramenta pode ser utilizada para qualquer curso avaliado no ENADE de 2021. A ferramenta foi aplicada para avaliar o curso de Ciência da Computação (Bacharelado) da UnB.

Os microdados inicialmente não estavam em um formato adequado para as análises e foram necessárias transformações como, por exemplo, o mapeamento dos nomes das instituições de ensino.

Foi avaliado o desempenho dos alunos da UnB do curso de Ciência da Computação (Bacharelado) em comparação com o mesmo curso em outras instituições. Nesta análise automatizada, pode-se verificar em detalhes qual foi o desempenho relativo dos alunos para cada uma das questões em relação às outras instituições de ensino selecionadas.

O usuário pode escolher o grupo de instituições de ensino que desejar para fazer as comparações. Isto é importante para fazer comparações mais equitativas e relevantes, pois permite comparar instituições semelhantes. Este tipo de informação é muito específica para constar dos relatórios oficiais do ENADE.

O presente trabalho mostra-se diferente de outros trabalhos na literatura pois ele mostra o resultado de cada questão ranqueado por instituições de ensino de um grupo selecionado. Isto permite comparar o índice de acerto por questão individual em relação a outras instituições. Além disto, diferente de outros trabalhos, mostra-se o desempenho relativo de cada instituição por questão. Isto permite, por exemplo, saber como os alunos de um curso estão em determinados conteúdos em relação a instituições semelhantes. Isto permite comparações mais específicas como, por exemplo, comparar instituições com o mesmo nível de recursos financeiros.

Estas análises podem auxiliar a reformular o ensino das disciplinas relacionadas com as questões. No caso deste trabalho, as análises foram utilizadas para formular sugestões para a melhoria dos cursos.

6. Trabalhos Futuros

Como trabalhos futuros, está em andamento um projeto que tem como objetivo fazer análise de dados do ENADE utilizando uma interface gráfica que ficará disponível *online* para que o usuário possa fazer sua análise de dados de forma mais fácil e acessível.

7. Agradecimentos

Agradecemos a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) pela disponibilização do portal CAFé. Neste trabalho NÃO foi utilizada IA.

Referências

Barbosa, P., Damazio, G., Carvalho, W., and Carmo, R. (2023). 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*, pages 59–68, Porto Alegre, RS, Brasil. SBC.
- BRASIL (2018). Lei geral de proteção de dados pessoais (lgpd). Lei nº 13.709.
- Charão, A., Wiechork, K., Rodrigues, M., and Barbosa, F. (2020). 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*, pages 16–20, Porto Alegre, RS, Brasil. SBC.
- Cunha, R., Sales, C., and Santos, R. (2021). 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*, pages 208–217, Porto Alegre, RS, Brasil. SBC.
- de Computação, S. B. (2024). Biblioteca digital da sociedade brasileira de computação.
- Freitas, B., Cosme, L., and Nascimento, M. (2019). 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*, pages 179–183, Porto Alegre, RS, Brasil. SBC.
- G. Van Rossum, F. L. DRAKE (2019). *Python 3 Reference Manual*. CreateSpace Independent Publishing Platform.
- Gondran, E., Lucca, G., Borges, E., and Santos, H. (2023). A evolução do perfil de estudantes de engenharia com bom desempenho no enade. In *Anais do XXXIV Simpósio Brasileiro de Informática na Educação*, pages 1238–1248, Porto Alegre, RS, Brasil. SBC.
- INEP (2023). Exame nacional de desempenho dos estudantes - enade. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade>. Acesso em: 18 de abril de 2023.
- Leão, J., Araújo, A., Souza, B. M., Araújo, J., Ferreira, K., Barros, M., Lauschner, T., and Nakamura, F. (2021). Uma análise regional sobre o perfil de estudantes concluintes em cursos steam através do enade. In *Anais do XV Women in Information Technology*, pages 200–209, Porto Alegre, RS, Brasil. SBC.
- Lima, P. d. S. N., Ambrósio, A. P. L., Oliveira, J. L. d. S., and Carvalho, C. L. d. (2021). Análise de conteúdo das provas do enade para os alunos do curso de bacharelado em ciência da computação. *Revista Brasileira de Informática na Educação*, 29:385–413.
- PROJECT JUPYTER (2023). Jupyter documentation. Acesso em: 13 de maio de 2023.
- Ramos, M., Oliveira, P., and Braga, A. (2024). Análise do desempenho dos estudantes do curso redes de computadores da universidade federal do ceará no enade. In *Anais do XXXII Workshop sobre Educação em Computação*, pages 658–668, Porto Alegre, RS, Brasil. SBC.
- Rendeiro, P., Monteiro, E., Nicioka, D., and Santos, R. (2023). Avaliação de desempenho em cursos de engenharia de computação baseada nos microdados do enade. In *Anais do XXXI Workshop sobre Educação em Computação*, pages 339–349, Porto Alegre, RS, Brasil. SBC.

- Rosa, E. R., Ferreira, D. J., Silva, N. F. F. d., and Assis, A. (2021). 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*, 29:1463–1486.
- Vendrami, L., Oliveira, F., and Tyeko Anami, K. (2021). A desigualdade social e a educação no brasil: uma revisão de literatura. *Metodologias E Aprendizado*, 4:390–398.
- Vieira, A., Bertolini, D., and Schwerz, A. (2022). Análise do desempenho no enade dos concluintes de computação usando técnica de agrupamento. In *Anais do XXXIII Simpósio Brasileiro de Informática na Educação*, pages 834–845, Porto Alegre, RS, Brasil. SBC.