

Análise Automática com os Microdados do ENADE para Melhoria do Ensino dos Cursos de Ciência da Computação

Renan Cunha, Claudomiro Sales, Reginaldo Santos

Instituto de Ciências Exatas e Naturais – Universidade Federal do Pará (UFPA)
Rua Augusto Corrêa, 01 – Guamá. CEP 66075-110. Belém – PA – Brasil

renan.cunha@icen.ufpa.br, {cssj, regicsf}@ufpa.br

Abstract. *This article presents the development of an automatic analysis with ENADE microdata for the Computer Science course. The goal is to provide information that can be useful for directors and coordinators who want to improve the quality of their courses. The data inform which are the deficient subjects of the course, what is the change in performance in a given subject over the years, and if the students have low participation in the exam. The fact that the data analysis is automatic makes it possible to generate results for any Computer Science course in the country without additional effort.*

Resumo. *Este artigo apresenta o desenvolvimento de uma análise automática com os microdados do ENADE para o curso de Ciência da Computação. O objetivo é fornecer informações que possam ser úteis para diretores e coordenadores que queiram melhorar a qualidade de seus cursos. Os dados informam quais são as disciplinas deficientes do curso, qual a mudança de desempenho em determinado assunto de um exame ao longo os anos e se os alunos têm baixa participação na prova. O fato da análise de dados ser automática torna possível a geração de resultados para qualquer curso de Ciência da Computação do país sem precisar de esforço adicional.*

1. Introdução

O Exame Nacional de Desempenho dos Estudantes (ENADE) é uma ação aplicada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), órgão vinculado ao Ministério da Educação (MEC), para avaliar o desempenho dos cursos ofertados pelas universidades e faculdades do país [INEP b]. Sabe-se que este exame é uma das formas de avaliar os cursos de nível superior no Brasil e seu resultado oferece evidências empíricas sobre o desempenho dos discentes dos cursos de graduação, que podem ser convertidos em ações voltadas para melhorias do ensino.

Dentre as diversas finalidades deste exame, podem se destacar: avariar debilidades e potencialidades das Instituições de Ensino Superior (IES); servir como uma referência para os estudantes analisarem as principais opções de curso de nível superior, uma vez que representa a qualidade da graduação ofertada pela instituição; orientar políticas educacionais para formar profissionais engajados e comprometidos com a sociedade, e mais habilitados para entrarem no mercado de trabalho; servir como um referencial básico dos processos de regulamentação e supervisão da educação superior; aprimorar o currículo do curso reduzindo, inclusive, índices de evasão [Brito 2008].

A primeira aplicação do exame ocorreu em 2004 e a periodicidade máxima da avaliação é trienal para cada área do conhecimento. O formato da prova consiste em questões discursivas e objetivas, de formação geral e de componente específico. A parte de formação geral compõe 25% da nota final na prova e é idêntica para todos os estudantes que realizam o exame em determinado ano. Já a parte de componente específico corresponde a 75% da nota final e é diferente dependendo de cada curso de graduação.

Os resultados do exame são disponibilizados através dos microdados e também através de relatórios. Vários trabalhos já utilizaram estes dados para obter informações gerais sobre instituições, cursos e a própria estrutura da prova [Lima et al. 2019]. Porém, há poucos trabalhos que realizam análises de dados com o objetivo de fornecer informação para melhorar um curso específico. E ainda, esses trabalhos possuem análises limitadas, já que elas não são executadas de forma automática, dificultando tanto a aplicação para outros cursos, como a exploração de informações mais complexas.

Dessa forma, o presente trabalho tem o objetivo de apresentar uma análise automática, que faz uso dos microdados disponibilizados pelo ENADE para obter informações não triviais acerca do desempenho e participação dos cursos de Bacharelado em Ciência da Computação (BCC) do país. Essas informações podem ajudar diretores a melhorar seus cursos, seja através da reformulação do Plano Pedagógico de Curso (PPC) ou através de implementação de políticas educacionais direcionadas.

O fato dessa análise ser automática faz com que ela possa ser executada de forma imediata para obter resultados de qualquer curso de BCC do Brasil, desde que este possua conceito no ENADE. Não é necessário realizar manualmente os passos dessa análise de dados. Além disso, os resultados dessa análise são demonstrados com o curso da Universidade Federal do Pará (UFPA). Não é do conhecimento dos autores nenhum outro trabalho que possua o mesmo objetivo.

2. Trabalhos Relacionados

Há diversos trabalhos que realizam análise de dados com os resultados do ENADE. Em [Lima et al. 2018], foi desenvolvida uma metodologia para classificação das questões do ENADE em temas. O objetivo do trabalho é facilitar estudos e análises que levem em conta a estrutura do exame. Um estudo de caso foi feito com as provas do ENADE de Ciência da Computação dos anos 2005, 2008, 2011 e 2014, a fim de verificar quais temas tiveram o maior número de acertos e de questões fáceis.

Em [Araújo 2019], foi desenvolvida uma ferramenta *web* com o objetivo de facilitar a análise de dados com os resultados do ENADE. A ferramenta utilizou dados de todos os cursos do ENADE de 2017 e era possível comparar a nota ENADE de cursos e instituições diferentes. Além disso, o trabalho utilizou um algoritmo de aprendizado de máquina que classificava as notas dos alunos por desempenho, com o objetivo de descobrir quais variáveis estão associadas a uma boa nota no ENADE. Essa ferramenta não se encontra mais disponível.

Apesar dos trabalhos anteriores serem avanços na análise de dados do ENADE, em especial o [Araújo 2019], eles não apresentam dados que podem ser usados para aperfeiçoar um curso específico. O único trabalho encontrado deste tipo foi o de [Charão et al. 2020]. Neste trabalho, os autores realizaram um estudo focado no curso

de Ciência da Computação da Universidade Federal de Santa Maria (UFSM). O desempenho dos alunos do curso é comparado, por disciplina, com as médias nacionais, considerando ao mesmo tempo os cinco anos em que o ENADE de computação foi aplicado. Através disso, os pesquisadores procuram entender quais os possíveis pontos fracos e fortes do curso e assim prover informação para a reformulação do plano pedagógico do mesmo.

Ainda sobre o trabalho de Charão et al., foram utilizados os documentos que o INEP publica em formato PDF. A extração de dados foi feita de forma semi-automática. Além disso, um especialista classificou a área de cada questão, com base nas categorias estabelecidas no ENADE de 2017. Uma limitação do trabalho de Charão et al. é o fato da análise não ser totalmente automática. Isso torna custoso a geração de resultados para cursos de Ciência da Computação de outras instituições. Além disso, os resultados são limitados pela informação presente nos documentos em formato PDF, já que os microdados não foram utilizados.

3. Metodologia

Dessa forma, o presente trabalho possui a motivação de utilizar dados para melhorar a qualidade de ensino. O objetivo dessa pesquisa é apresentar uma análise de dados que provê informações para os cursos de Bacharelado em Ciência da Computação. Informações essas que podem ser utilizadas para tomadas de decisão que melhorem a qualidade do curso em si.

O objetivo específico do trabalho é o desenvolvimento de *software* para análise com os microdados do ENADE. Isso permite dois avanços neste tipo de análise. O primeiro, é que não é necessário realizar os passos da análise de forma manual, é preciso apenas executar o *software* que a análise de dados é realizada automaticamente, permitindo a obtenção dos resultados de qualquer curso de Ciência da Computação de forma imediata. O segundo, é que outros resultados podem ser derivados a partir da exploração dos dados fornecidos pelo INEP, já que o uso de microdados e de código permite a manipulação de dados de forma fácil e rápida.

3.1. Planejamento

Foram definidas quais informações iriam auxiliar os cursos a aumentar o desempenho e a participação no ENADE. Assim, foram estabelecidas as seguintes questões de pesquisa:

1. **Como o desempenho de um curso no ENADE se compara com o desempenho médio nacional?** Essa resposta pode ajudar o diretor a verificar se o curso possui alguma deficiência em alguma das partes do exame. Com isso, a direção pode reformular a estratégia pedagógica para dar mais ênfase em suas deficiências.
2. **Como o desempenho do curso varia em um determinado período de tempo?** Respondida essa pergunta, o diretor de curso pode confirmar se a estratégia pedagógica aplicada está gerando o efeito desejado.
3. **Há uma alta abstenção dos estudantes do curso no exame?** Caso isso seja verdadeiro, a direção de curso pode estudar políticas que incentivem os estudantes a realizar o exame [Miranda et al. 2018].

Para responder essas questões, as seguintes métricas foram definidas:

- Percentual de desempenho total, de formação geral e de componente específico para cada ano em que o curso foi avaliado.
- Razão do percentual de desempenho do curso pelo desempenho médio nacional por tema da parte de Componente Específico do ENADE e considerando ao mesmo tempo, todos os anos em que a prova foi aplicada.
- Percentual de desempenho do curso e do país, por tema do componente específico, nos últimos dois anos em que este tema foi avaliado. Esse dado permite verificar qual a evolução recente do curso em determinada parte da prova, e junto ao desempenho do Brasil, é possível verificar se a variação de desempenho ocorreu por motivos externos como, por exemplo, mudança de dificuldade de prova ou greve.
- Número de alunos presentes e ausentes do curso por ano.
- Número de questões deixadas em branco e rasuradas por ano.

Os percentuais de desempenho são calculados pela média ponderada das notas de questões, considerando tanto o número de alunos e o número de questões.

3.2. Execução e Análise de Dados

O *software* foi desenvolvido na linguagem de programação *Python*¹, em conjunto com os *frameworks* de manipulação de dados *Pandas*² e *Matplotlib*³. A ferramenta *Jupyter Notebook*⁴ é utilizada para visualização dos resultados.

A análise é realizada seguindo o processo *Knowledge Discovery in Databases* (KDD) [Fayyad et al. 1996]. Este processo possui as etapas de coleta, seleção, pré-processamento, transformação e mineração de dados. Todas essas etapas foram implementadas, e são realizadas de forma automática quando o *software* é executado.

A coleta de dados é realizada através do site oficial do ENADE⁵. Os microdados dos anos de 2017, 2014, 2011, 2008 e 2005 são utilizados, já que são respectivos aos anos em que a área de computação foi avaliada. É importante ressaltar, que estava prevista uma aplicação do ENADE para a área de computação em 2020. Porém, o exame foi adiado para 2021, por conta da pandemia do coronavírus [INEP c]. Todos os arquivos coletados estão no formato *Comma-separated values* (CSV), que organiza a informação em forma de tabela, com linhas e colunas. Também é coletado um arquivo CSV disponibilizado em [Charão et al. 2020]. Este arquivo indica quais os temas de cada questão do exame, para o curso de Bacharelado em Ciência da Computação (BCC), considerando todos os cinco anos em que o exame foi aplicado. Essa classificação em temas foi feita por um especialista, que categorizou cada questão em até três temas, com base nas categorias do ENADE de 2017.

Posteriormente, é realizada a seleção dos dados relevantes ao projeto. São selecionados somente os estudantes do curso de BCC. Também é utilizado um filtro adicional para os anos de 2008 e 2005. Nestes anos, o ENADE era realizado tanto com alunos

¹<https://www.python.org/>

²<https://pandas.pydata.org/>

³<https://matplotlib.org/>

⁴<https://jupyter.org/>

⁵<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade/resultados/>

ingressantes como com egressos, mas a partir de 2010, os alunos ingressantes foram dispensados de realizar o exame [INEP a]. Dessa forma, para as bases de dados dos anos de 2008 e 2005, somente os estudantes egressos são selecionados.

Após isso, é realizada a limpeza e pré-processamento de dados. Foi percebido que alguns estudantes tinham dados ausentes. Isso ocorria com alunos que eram os únicos, de seus respectivos cursos, a fazer a prova. É importante salientar que estudantes nessa situação não são considerados para o cálculo do conceito ENADE [INEP d]. Dessa forma, esses alunos são ignorados durante a análise de dados.

Na etapa de transformação de dados, novas colunas foram criadas nos arquivos CSV para facilitar a mineração de dados. Foram criadas duas colunas para cada uma das questões da prova do ENADE. Uma coluna armazenava a nota daquela questão. A outra coluna informava a situação daquele aluno com a questão e dizia se ela foi respondida ou deixada em branco. No caso das questões objetivas, essa coluna também podia indicar que a questão foi rasurada.

Por último, foi realizada a etapa de mineração de dados, a fim de produzir informações que eram valiosas para o diretor de curso. Nessa etapa, foi ignorado o desempenho das questões subjetivas do ENADE, já que essas questões possuem uma alta taxa de abstenção e isso prejudicaria a análise. Esse fato é mais explorado na Seção 4.

Toda a análise está disponível de forma *Open-Source* no *GitHub*⁶. Para executá-la, é necessário primeiramente rodar o comando `make first_time` para coletar, selecionar, pré-processar e transformar os dados. Depois é preciso executar o comando `make code_course=<CÓDIGO CURSO> run_notebooks` indicando o código eMEC do curso a ser analisado. Após essas etapas, os resultados estarão disponíveis na pasta `results/`.

4. Resultados e Discussão

Para demonstrar a efetividade da análise, foram obtidos os resultados com o curso de Bacharelado em Ciência da Computação da Universidade Federal do Pará (UFPA).

4.1. Desempenho Total, Geral e Específico

Os percentuais de desempenho total, geral e específico estão disponibilizados na Figura 1, onde pode ser verificado que houve uma melhora na parte de componente específico. O percentual de desempenho, que era de 31,97 em 2005, foi para 44,24 em 2017. Contudo, também podemos ver que houve uma diminuição do desempenho para a formação geral de quase 20 pontos percentuais, variando de 78,57% para 59,56% durante o período analisado. Dessa forma, como o componente específico define 75% da nota final, o percentual de desempenho total subiu de 43,22% para 48,07%.

4.2. Comparação do Desempenho por Tema do Componente Específico.

Para verificar o desempenho do curso considerando os temas de componente específico, o percentual de desempenho histórico do curso é dividido pelo percentual histórico de desempenho nacional, considerando ao mesmo tempo, todos os anos em que a prova foi aplicada.

⁶<https://github.com/renan-cunha/KDD-Enade-Computing>

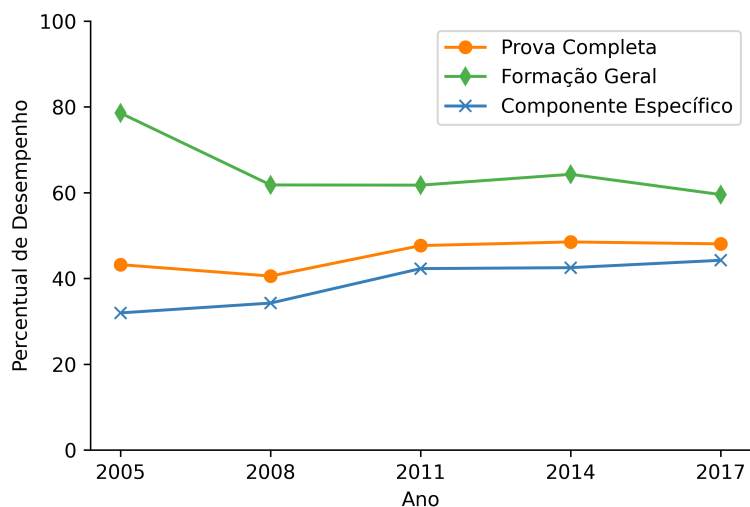


Figura 1. Desempenho no ENADE ao longo dos anos considerando a prova completa, formação geral e componente específico.

Assim, o curso possui desempenho maior que o país, no respectivo tema, quando a razão dessa divisão é superior a 1. Caso essa razão seja inferior a 1, o curso possui desempenho inferior a média nacional para o tema comparado. O resultado com o curso da UFPA está sumarizado na Figura 2.

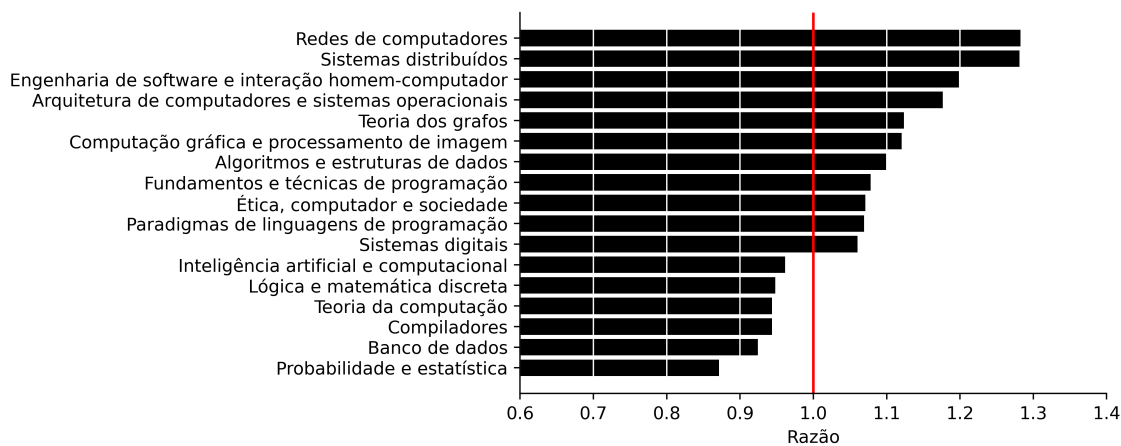


Figura 2. Razão do percentual de desempenho da UFPA pelo desempenho nacional. Razões maiores que 1.0 (à direita da linha vermelha) indicam temas em que o desempenho da UFPA é superior à média nacional. Razões menores que 1.0 (à esquerda da linha vermelha) indicam temas em que a UFPA possui desempenho inferior à média nacional.

É possível verificar que a UFPA possui um desempenho de 28% maior nos temas de Redes de Computadores e Sistemas Distribuídos. Uma possível explicação para isso é a presença de várias disciplinas exclusivas para esses assuntos [FACOMP 2010]. O curso de BCC da UFPA possui duas disciplinas obrigatórias de Sistemas Distribuídos, uma teórica e outra prática. Já para Redes de Computadores, há um total de quatro dis-

ciplinas obrigatórias sobre o assunto: Fundamentos da Comunicação Digital, Redes de Computadores, Laboratório de Redes de Computadores e Redes Multimídia.

Quanto ao pior desempenho, quando comparado à média nacional, o assunto de Probabilidade e Estatística se destaca. O curso da UFPA possui apenas uma disciplina com esse tema e ela normalmente é lecionada por professores que não são da Faculdade de Computação. Além disso, essa disciplina normalmente é ministrada no primeiro ano de curso, o que pode contribuir para que os alunos se distanciem ou esqueçam parte dos assuntos que foram ministrados na disciplina, uma vez que a prova é feita no final do curso.

Já para verificar a evolução recente do curso em cada disciplina, são disponibilizados os desempenhos de cada tema para os últimos dois anos em que este assunto apareceu no exame. Esse resultado é mostrado na Figura 3.

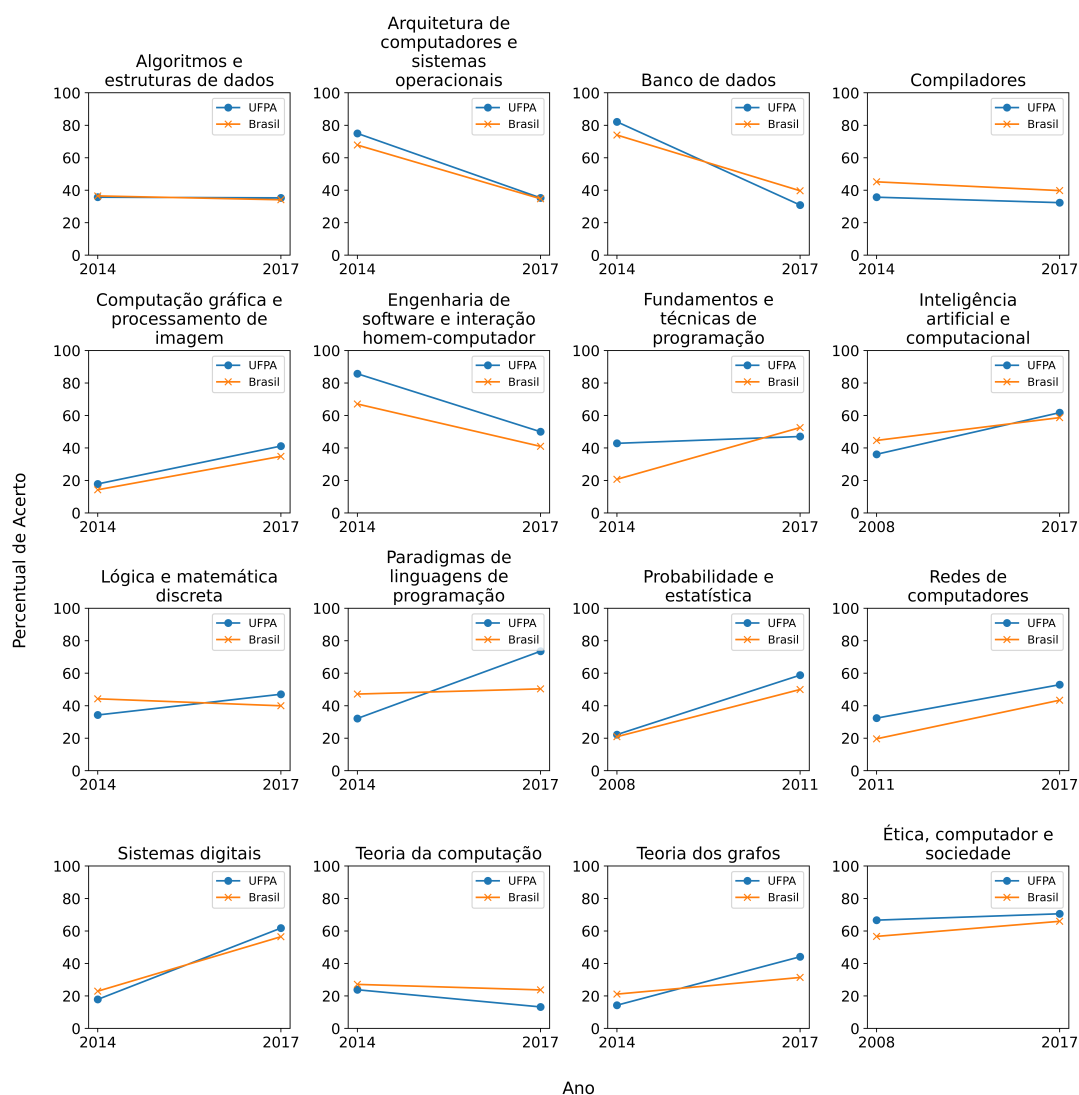


Figura 3. Desempenho nacional e da UFPA nos últimos dois anos em que houveram questões do tema. Essa análise não foi realizada para o tema de Sistemas Distribuídos, pois houve apenas um ano com questões válidas desse tema.

A partir da Figura 3, pode ser verificado que o desempenho no tema de Banco de Dados diminuiu mais que o desempenho nacional de 2014 a 2017. Também podemos ver que o desempenho em Algoritmos e Estrutura de Dados se manteve constante e muito próximo do desempenho médio nacional neste mesmo período.

No tema de Fundamentos e Técnicas de Programação, tanto o desempenho da UFPA como o do país melhoraram. Porém, a melhora da UFPA foi discreta se comparada com a melhora nacional. Por outro lado, a UFPA melhorou, relativamente ao desempenho médio brasileiro, nos temas de Lógica e Matemática Discreta, Paradigmas de Linguagens de Programação e Teoria dos Grafos.

Por último, podemos notar que o pior desempenho percentual da UFPA é no tema de Teoria da Computação, único tema que teve, no último ano em que foi avaliado, um desempenho abaixo de 20%. Mas pode ser notado que o desempenho nacional nessa disciplina também é baixo.

4.3. Dados de Questões em Branco e Alunos Faltantes

Os dados de questões em branco e de alunos faltantes são importantes para inferir a participação dos alunos no ENADE. A Figura 4 possui os dados com o número de alunos presentes e faltantes do exame por ano. Podemos concluir que o ano de 2011 foi o ano com o menor número de estudantes presentes, enquanto o ano de 2017 teve um alto número de estudantes ausentes.

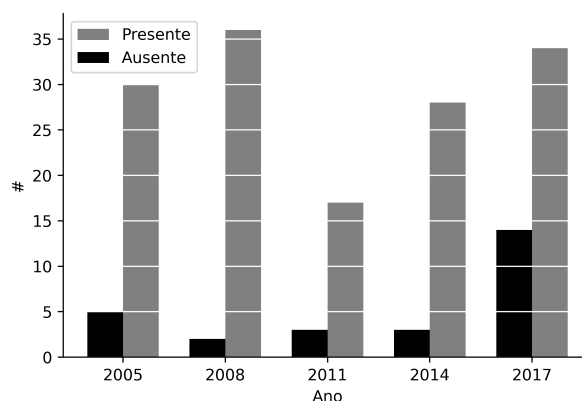


Figura 4. Frequência de estudantes por ano.

Na Tabela 1, são disponibilizados os dados acerca do número de questões deixadas em branco ou rasuradas. Podemos verificar que houve apenas uma questão objetiva em branco no ano de 2017, o que sugere uma falta de atenção de apenas um estudante. Por outro lado, no mesmo ano de 2017, houve 43 questões discursivas em branco e dois alunos deixaram de responder a prova discursiva inteira.

No ano de 2005, podemos ver que o número de questões objetivas deixadas em branco é muito alto, 103 questões no total. O ano de 2005 é conhecido por ter vários estudantes que boicotavam a prova e deixavam propositalmente o cartão resposta em branco [Leitão et al. 2010].

Tabela 1. Dados acerca de questões deixadas em branco e rasuradas.

Ano	Nº de questões objetivas em branco	Nº de questões objetivas rasuradas	Nº de questões discursivas em branco	Nº de alunos que deixaram toda a prova discursiva em branco
2017	1	0	43	2
2014	22	1	36	2
2011	4	0	44	1
2008	9	2	65	2
2005	103	0	44	0

5. Conclusão e Trabalhos Futuros

Esse trabalho apresentou uma análise de dados de desempenho e participação do ENADE, com o curso de Bacharelado em Ciência da Computação. O objetivo é auxiliar as direções das faculdades a melhorar seus cursos, provendo informações que podem ajudá-las a detectar deficiências curriculares ou ainda uma baixa participação dos estudantes no exame.

A principal contribuição deste trabalho é o fato da análise de dados poder ser executada de forma automática, utilizando os microdados disponibilizados pelo INEP. Isso permite a reprodução destes resultados para qualquer curso de Bacharelado em Ciência da Computação do país, desde que esse curso possua um conceito ENADE. O fato dessa análise utilizar os microdados de forma automática também permitiu a geração de dados ainda não explorados na literatura, como a evolução de desempenho por disciplina entre anos e o número de alunos que deixaram a prova discursiva inteira em branco.

O trabalho também demonstrou o uso desta análise, através da geração de resultados com a Universidade Federal do Pará. Foi visto que o desempenho do curso na parte de componente específico subiu ao longo dos anos, que o desempenho de formação geral diminuiu, que a UFPA possui deficiência nos temas de Inteligência Artificial, Teoria da Computação, Compiladores, Banco de Dados e nos temas de Matemática. Também foi visto que o número de alunos ausentes foi mais alto que o comum no ano de 2017, e que os alunos em geral deixam muitas questões discursivas em branco. Todas essas informações podem servir de subsídio para melhoria do curso.

Para trabalhos futuros, pretende-se incluir suporte à análise para outros cursos de computação, como Sistemas de Informação e Engenharia da Computação. Além disso, planeja-se desenvolver uma ferramenta *web* para facilitar ainda mais o acesso a esta análise de dados.

Agradecimentos

Esse trabalho foi financiado pelo programa institucional de bolsa de iniciação científica, edital 08/2020, da Pró-reitoria de Pesquisa e Pós-Graduação da UFPA.

Referências

Araújo, R. A. (2019). Análise dos microdados do Enade: proposta de uma ferramenta de exploração utilizando mineração de dados. *Dissertação (Mestrado em Ciência da*

Computação) - Universidade Federal de Goiás, Goiânia.

- Brito, M. R. F. d. (2008). O SINAES e o ENADE: da concepção à implantação. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 13(3):841–850.
- Charão, A. S., Wiechork, K., Rodrigues, M. L., e Barbosa, F. P. (2020). Explorando resultados por questão no enade em ciência da computação para subsidiar revisão de projeto pedagógico de curso. Em *Anais do XXVIII Workshop sobre Educação em Computação*, páginas 16–20. SBC.
- FACOMP (2010). Estrutura curricular do curso de ciência da computação da ufpa. <https://sigaa.ufpa.br/sigaa/link/public/curso/curriculo/152148>. Acesso em 11 de março de 2021.
- Fayyad, U., Piatetsky-Shapiro, G., e Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3):37.
- INEP. Enade: Perguntas frequentes. <https://www.gov.br/inep/pt-br/aceso-a-informacao/perguntas-frequentes/exame-nacional-de-desempenho-dos-estudantes-enade>. Acesso em 21 de Janeiro de 2021.
- INEP. Exame nacional de desempenho dos estudantes (enade). <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade/enade>. Acesso em 13 de janeiro de 2021.
- INEP. Inep adia aplicação do enade 2020. http://inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/inep-adia-aplicacao-do-enade-2020/21206. Acesso em 25 de janeiro de 2021.
- INEP. Nota técnica - cálculo da nota final do enade. https://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2019/nota_tecnica_n20-2019_CGCQES-DAES_calculo_NF_Enade.pdf. Acesso em 18 de janeiro de 2021.
- Leitão, T., Moriconi, G., Abrão, M., e Silva, D. (2010). Uma análise acerca do boicote dos estudantes aos exames de avaliação da educação superior. *Revista Brasileira de Educação*, 15(43):21–44.
- Lima, P. D. S., Ambrósio, A. P., Brancher, J. D., e Felix, I. (2018). SysEnade - Análise das questões de provas do Enade organizadas pelos Temas abordados. *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, 7(1):419.
- Lima, P. d. S. N., Ambrósio, A. P. L., Ferreira, D. J., e Brancher, J. D. (2019). Análise de dados do Enade e Enem: uma revisão sistemática da literatura. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 24(1):89–107.
- Miranda, G. J., Leal, E. A., Gama, M. A. F., e Miranda, A. B. d. (2018). Enade: Os estudantes estão motivados a fazê-lo? *Revista de Educação e Pesquisa em Contabilidade (REPeC)*, 13.