

O sucesso não é apenas uma questão de sorte: um mapeamento sistemático sobre técnicas de análise do Enade da área de Computação

Pedro Luis Saraiva Barbosa, Gabriela Nayara Duarte Oliveira Damazio, Windson Viana de Carvalho, Rafael Carmo
pedro.barbosa@ifce.edu.br, gabrielanayara10@gmail.com, {windson, carmorafael}@virtual.ufc.br
Universidade Federal do Ceará, UFC, BR

RESUMO

Este trabalho apresenta um Mapeamento Sistemático da Literatura (MSL) que busca identificar e discutir especificamente quais técnicas estão sendo usadas para analisar dados do Enade ou implementar ações diagnósticas para o acompanhamento das competências necessárias para o perfil do egresso que realiza tal exame. Dos 32 trabalhos incluídos nesta MSL, este artigo, em especial, analisa os 12 trabalhos que focam na análise do Enade de cursos de Computação. Foram encontrados trabalhos de cada uma das três abordagens, as quais são uso de *machine learning* (ML), uso de estatística e desenvolvimento de sistemas. Percebe-se que o principal foco dos trabalhos é analisar fatores que impactam nas notas dos alunos. A estatística descritiva foi a mais utilizada nos trabalhos. Em relação a ML as abordagens mais utilizadas são a de classificação e agrupamento. Os sistemas encontrados se limitam a aplicação de provas e a análise desses resultados. Em apenas 01 deles, foi realizado uso de gamificação.

CCS CONCEPTS

• **Social and professional topics** → Computing education.

PALAVRAS-CHAVE

Cursos de Computação, Enade, Mapeamento Sistemático da Literatura, Técnicas estatísticas, Técnica de *machine learning*.

1 INTRODUÇÃO

O Exame Nacional de Desempenho dos Estudantes (Enade) é aplicado desde 2004 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) [15] e integra o Sistema Nacional de Avaliação da Educação Superior (Sinaes) [5].

De acordo com Brasil [5] o ENADE afere:

o desempenho dos estudantes em relação aos conteúdos programáticos previstos nas diretrizes curriculares do respectivo curso de graduação, suas habilidades para ajustamento às exigências decorrentes da evolução do conhecimento e suas competências para compreender temas exteriores ao âmbito específico de

sua profissão, ligados à realidade brasileira e mundial e a outras áreas do conhecimento.

Ainda conforme Brasil [5], o ENADE deve ser aplicado de forma periódica com periodicidade máxima de 03 anos e é componente curricular obrigatório para os cursos de graduação. Para garantir essa periodicidade, todos os anos é publicado um edital para tornar pública as diretrizes para a avaliação que será realizada no ano vigente, como por exemplo o Edital número 51, de 24 de junho de 2022 [6].

Após a publicação do edital, são publicadas as portarias contendo diretrizes específicas da prova e componentes específicos dos cursos que realizarão a avaliação [16]. Todas as avaliações contêm 40 questões, sendo 10 com conteúdos de Formação Geral (FG) e 30 do Componente Específico (CE) [6]. Na parte respectiva a FG, são 02 questões discursivas e 08 de múltipla escolha. Na parte referente ao CE, são 03 questões discursivas e 27 questões de múltipla escolha. Além das 40 questões, o Enade também possui um questionário socioeconômico que é respondido pelos estudantes antes de realizar a prova e um questionário de percepção da prova que é respondido no dia do exame [23].

A nota do Enade é um dos indicadores que afeta o Conceito Preliminar do Curso (CPC), e a média dos CPCs dos cursos de uma Instituição de Ensino Superior (IES) afeta o seu Índice Geral de Cursos (IGC) [17]. Todos esses indicadores são mensurados com notas que variam de 1 a 5, sendo 1 um conceito ruim e 5 um conceito de excelência.

Como o Enade avalia competências que devem fazer parte do perfil do egresso, o resultado dessa avaliação é importante para toda a comunidade acadêmica de uma IES. Por meio do resultado dessa avaliação de larga escala, professores, alunos e gestores podem redefinir práticas e assim, melhorar os cursos e consequentemente melhorar a oferta do ensino em geral. Nesse aspecto, as IES podem analisar o currículo dos seus cursos com base nos resultados do Enade e identificar áreas em que os seus estudantes possuem domínio ou não [23].

Diante do exposto, este mapeamento sistemático busca identificar e discutir especificamente quais técnicas estão sendo usadas para analisar dados do Enade ou implementar ações diagnósticas para o acompanhamento das competências necessárias para o perfil do egresso do estudante que realiza tal exame.

O protocolo para este Mapeamento Sistemático da Literatura (MSL), foi baseado no *Guidelines for performing Systematic Literature Reviews in Software Engineering* [18]. Após as buscas nas bases de dados, conseguiu-se 129 trabalhos. Após a análise desses trabalhos em relação aos critérios de seleção restaram 32 trabalhos

Fica permitido ao(s) autor(es) ou a terceiros a reprodução ou distribuição, em parte ou no todo, do material extraído dessa obra, de forma verbatim, adaptada ou remixada, bem como a criação ou produção a partir do conteúdo dessa obra, para fins não comerciais, desde que sejam atribuídos os devidos créditos à criação original, sob os termos da licença CC BY-NC 4.0.

EduComp'23, Abril 24-29, 2023, Recife, Pernambuco, Brasil (On-line)

© 2023 Copyright mantido pelo(s) autor(es). Direitos de publicação licenciados à Sociedade Brasileira de Computação (SBC).

incluídos. Esses 32 trabalhos foram aplicados em diversas áreas. Neste artigo, em especial, foram analisados os 12 trabalhos que focam na análise do Enade de cursos de Computação. Com este MSL, espera-se beneficiar as IES e/ou pesquisadores que pretendem realizar estudos com dados do Enade, ou conhecer ferramentas propostas pela comunidade científica para o acompanhamento dos estudantes que realizam tal prova.

Sobre trabalhos relacionados com esta pesquisa, Lima [21] realizou uma Revisão Sistemática da Literatura (RSL) sobre análise de dados do Enade com 39 trabalhos incluídos. A pesquisa utilizou como base de dados o Google Scholar e incluiu trabalhos publicados entre 2005 e 2016. Os autores concluem com a RSL que as análises geralmente estão relacionadas à estatística descritiva dos dados com a finalidade de descrever e resumir informações que são coletadas. Os autores também encontraram trabalhos que envolviam análises sofisticadas, como a utilização de estatística inferencial e/ou mineração de dados. Com a RSL foi possível identificar que o objetivo da maioria dos trabalhos foi analisar desempenho/rendimento dos estudantes e das IES, e que os problemas mais abordados por esses trabalhos estão associados a verificar como os estudantes se desenvolvem de acordo com seus perfis e a melhoria da qualidade de ensino. Dentre os trabalhos incluídos, os autores encontraram um único trabalho que utilizou mineração de dados, e esse realizava uma análise sobre as notas do Enade. Em relação ao uso da estatística, ela foi utilizada como ferramenta para análise das notas do Enade, questionário socioeconômico, conteúdo da prova, conceito do Enade, questionário do coordenador e questionário de percepção do estudante.

Diferente de Lima [21], esta RSL investiga os trabalhos dos últimos 10 anos (2012-2022) abordando especificamente trabalhos que utilizam dados da área de Computação.

Nas seções subsequentes, apresenta-se o Método utilizado nesta pesquisa na seção 2, os Resultados na seção 3 com dados quantitativos e qualitativos sobre os trabalhos incluindo a respostas as questões de pesquisa, a Discussão na seção 4 e por fim as Considerações Finais na seção 5.

2 MÉTODO

2.1 Protocolo

O protocolo deste MSL está apresentado visualmente na Figura 1 e foi baseado em Keele et al. [18], que possui três principais fases nomeadas: *planning*, *conducting the review* e *reporting the review*.

Na primeira fase, a de planejamento, buscou-se identificar a necessidade do estudo e por meio dessa necessidade definir suas questões de pesquisa. A necessidade desta pesquisa está associada a mapear estratégias que são utilizadas para o desenvolvimento de ferramentas para preparar estudantes para o Enade como também para identificar técnicas utilizadas para a análise de dados dos microdados do Enade. Após esta identificação, foi definido o protocolo desta pesquisa, no qual definiu-se os objetivos do mapeamento, as bases de pesquisa, a *string* de busca e os critérios de seleção. O protocolo foi construído de forma incremental, passando pelo processo de revisão por meio de todos os autores desta pesquisa.

Na segunda etapa, executou-se o protocolo definido anteriormente. Primeiramente, foram realizadas buscas nas bases de dados por meio da *string* de busca. Em seguida, foram lidos apenas os

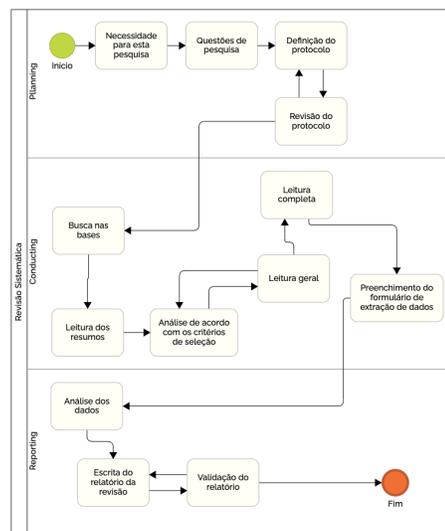


Figura 1: Processo do MSL

resumos dos trabalhos e então o trabalho foi avaliado segundo os critérios de seleção. Essa etapa foi importante para diminuir o número de trabalhos a serem avaliados. Com um número reduzido, pôde-se ler de forma geral os trabalhos e dessa forma revisá-los quanto aos critérios de seleção. Depois dessa etapa, os trabalhos finais já estavam selecionados e foi possível realizar uma leitura completa de todos eles. Sendo assim, foi criada uma planilha e compartilhada com todos os autores. Nessa planilha estavam presentes o título, o tipo de documento, o ano, e as questões de pesquisa de forma detalhada. Isso oportunizou que todos os autores acompanhassem o preenchimento de todos os itens.

Por fim, na terceira etapa, foi realizada uma análise dos dados por meio da geração de gráficos e tabelas a partir dos dados preenchidos no formulário de extração de dados. Logo após, este relatório foi escrito e então validado de forma incremental.

2.2 Questões de pesquisa

As questões de pesquisa para guiar este trabalho foram definidas utilizando a estratégia PICO (População Intervenção Controle Outcomes) [40], que ajudou a estruturar este mapeamento e selecionar questões de pesquisa relevantes.

Como apresentado na Figura 2, primeiramente verificou-se a necessidade do estudo e buscou-se de forma rápida na literatura por meio do Google Scholar trabalhos associados a utilização de tecnologias para o Enade. Dessa forma, pôde-se verificar alguns trabalhos publicados na área de interesse deste estudo. Baseado nessa busca inicial, criou-se uma questão de pesquisa (QP) inicial e foi refinando-a de forma clara e objetiva. Após o refinamento da questão de pesquisa, pôde-se utilizar a ferramenta PICO e definir a questão de pesquisa sintetizada que foi subdividida em várias outras, de forma a facilitar as respostas e dá uma sequência lógica aos resultados. Ao utilizar a ferramenta PICO, definiu-se estudantes do ensino superior e Instituições de Ensino Superior (IES) como população; softwares e análises de dados com direcionamento para

o Enade como intervenção; os microdados do Enade e estudos feitos com estudantes como controle; e as soluções algorítmicas, soluções estatísticas e *softwares* desenvolvidos como *outcomes*. Dessa forma, a QP sintetizada foi: Quais soluções de desenvolvimento de *software*, estatísticas e algorítmicas estão sendo utilizadas para analisar dados dos estudantes e microdados do Enade das IESSs?

Assim sendo, as questões de pesquisa utilizadas para este trabalho são: Quais áreas do ENADE foram o foco da pesquisa?; Qual foi a abordagem e objetivo da pesquisa?; Dos trabalhos de análise de dados, quais foram as técnicas estatísticas mais usadas?; Dos trabalhos de análise de dados, quais foram as técnicas de mineração ou algorítmicas usadas?; Quais os principais resultados e impactos das pesquisas que usam dados do ENADE?; Quais foram as características principais dos sistemas com foco no ENADE?; e Quais os benefícios e limitações dos sistemas apontados pelos autores?

2.3 String de busca

O processo para construção da *string* de busca está apresentado na Figura 2. Para a construção, iniciou-se definindo PICO a qual foi necessária para encontrar palavras-chave que pudessem ser utilizadas na *string*. Por meio da divisão feita a partir da estratégia PICO, definiu-se a combinação das palavras-chave (OR ou AND). Com a *string* construída realizou-se um teste dessa *string* no Google Scholar e logo em seguida realizou-se a análise dos trabalhos que foram retornados como resultado. Após isso, foi verificado se nas três primeiras páginas dos resultados os trabalhos retornados possuíam relação com as questões de pesquisa, caso não possuíssem correspondência era realizada a reformulação da *string* até que os resultados representassem relação com esta pesquisa.

As bases utilizadas para realizar a busca automatizada nesta pesquisa foram: Google Scholar ¹ e Periódicos Capes ². A escolha dessas bases está relacionada com o tema desta pesquisa que está restrita a um exame específico do Brasil e que ao realizar a busca na Scopus pela palavra "Enade" com restrição da subárea *Computer Science* a busca retornou apenas 02 documentos *Open Access*, e nenhum dos 02 documentos possuíam relação com esta pesquisa; no IEEE as buscas retornaram 06 trabalhos, dos 06, 03 possuíam relação com esta pesquisa mas já haviam sido retornados pelo Google Scholar. A busca pelos trabalhos foi realizada entre os dias 15 e 18 de março de 2022.

A *string* de busca usada nas bases de dados foi: ("Educação Superior" OR "Avaliação de larga escala" OR "Avaliação educacional superior" OR "Teoria de resposta ao item" OR "Teoria clássica dos itens" OR "Avaliação do conhecimento") AND "Enade" AND ("Mineração de dados" OR "Aprendizagem de Máquina" OR "análise de dados" OR "Algoritmo" OR "ferramenta de análise" OR "análise da aprendizagem").

2.3.1 Seleção e extração de dados dos estudos. As Tabelas 1 e 2 apresentam os critérios para inclusão e exclusão de trabalhos nesta pesquisa.

Para realizar a extração de dados dos artigos, optou-se por utilizar uma planilha no Google Sheet ³ compartilhada com todos os autores, a qual possui em suas linhas a listagem dos artigos e em suas colunas

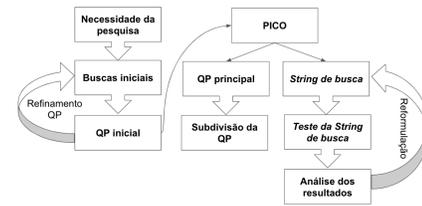


Figura 2: Processo de construção da QP e string de busca

as respostas as questões importantes para esta pesquisa, como por exemplo: os métodos utilizados no trabalho, a origem dos dados utilizados no trabalho, em quais áreas o estudo foi aplicado, quais técnicas foram empregadas para analisar os dados do Enade, dentre outras.

Tabela 1: Critérios de inclusão

Critérios de inclusão	
01	Artigos que apresentam algoritmos para realizar alguma investigação sobre o Enade
02	Artigos que apresentam técnicas estatísticas para realizar alguma investigação sobre o Enade
03	Artigos que apresentam a utilização de softwares de simulados para trabalhar questões do Enade

Tabela 2: Critérios de exclusão

Critérios de exclusão	
01	Artigos que não implementam algoritmos para realizar alguma investigação sobre o Enade
02	Artigos que não possuem o Enade como objeto central do estudo
03	Estudos não escritos em Português ou Inglês
04	Estudos que são short paper (com menos de 2 páginas)
05	Estudos incompletos
06	Não disponíveis
07	Duplicados
08	Estudos secundários
09	Estudos publicados a mais de 10 anos (antes de 2012)

3 RESULTADOS

Após a definição do protocolo do MSL a pesquisa foi executada. As buscas nas bases de dados retornaram 129 trabalhos, desses, 32 foram aceitos e 97 recusados seguindo os critérios de seleção definidos no protocolo (Tabelas 1 e 2).

Dos 32 trabalhos incluídos nesta pesquisa, 14 foram incluídos pelo critério 01, 12 foram incluídos pelo critério 02 e 06 foram incluídos pelo critério 03 (Tabela 1). Foram encontrados, portanto, mais trabalhos que usam técnicas estatísticas com dados do Enade

¹<https://scholar.google.com.br/>

²<https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/index.php?>

³<https://docs.google.com/spreadsheets/u/0/>

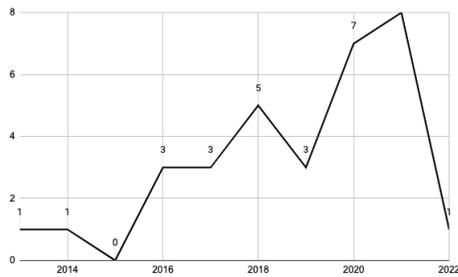


Figura 3: Distribuição dos trabalhos por ano

do que trabalhos de análise pautadas em algoritmos de *machine learning* e também mais do que trabalhos que apresentam *softwares* que usam questões do Enade. A quantidade de trabalhos incluídos nesta pesquisa é inferior ao total de trabalhos encontrados em Lima [21], o que pode estar associado ao período em que a pesquisa foi executada, como também associado a estruturação da *string* de busca. Lima [21] pesquisou trabalhos entre 2005 e 2016, e este trabalho pesquisou trabalhos no período de 2012 a 2022.

A Figura 3 apresenta a distribuição dos trabalhos incluídos por ano. Pode-se perceber que existe um crescente interesse dos pesquisadores por esta área de estudo nos últimos anos, com destaque para o ano de 2021, onde foram encontrados 08 trabalhos. A Tabela 3 apresenta especificamente os 12 trabalhos relacionados à área de Computação, os quais são alvos da seção de discussão desta MSL. As próximas subseções respondem às questões de pesquisa.

3.1 Quais áreas do ENADE foram o foco da pesquisa?

A Tabela 4 apresenta as áreas de conhecimento às quais os trabalhos centravam. A grande parte dos trabalhos não tinha foco em uma área específica. Alguns trabalhos mesclaram áreas, como Melguizo and Wainer [27] que agregaram os cursos de matemática, ciências sociais e ciências biológicas e Coelho [10] que agregou os cursos de engenharia química, administração e estatística. Outros trabalhos utilizaram dados de todos os cursos de um único estado [24] ou de uma determinada região, como Matos et al. [25] que utilizaram dados de todos os cursos da Região Norte do Brasil. Os demais trabalhos apresentam dados de todos os cursos de um ano específico [3, 8] e outros de um período específico [2, 20, 31].

Dos cursos que fizeram uso de dados da área de Ciências Exatas e da Terra, pode-se destacar para esta pesquisa os trabalhos dos autores Charão et al. [9], Cunha et al. [11], Figueiro et al. [12], Lima et al. [22, 23], Rosa et al. [32], Vista et al. [42] que fizeram uso de dados específicos dos cursos de bacharelado em Ciência da Computação (CC) e dos autores Landes and Manhães [19], Silva et al. [36] que fizeram uso de dados de todos os cursos da área de Computação em um determinado período.

Contrariamente, os trabalhos Barbosa and Carvalho [4], França et al. [14], Lima et al. [23], Ribeiro and C Junior [30] não usaram dados do Enade, mas desenvolveram sistemas com foco no Enade para a área de Computação.

3.2 Qual foi a abordagem e objetivo da pesquisa?

Três abordagens foram usadas pelos autores dos trabalhos incluídos nesta pesquisa. Duas delas focavam na análise de dados do Enade, a primeira com técnicas de *machine learning*, como apresentado na Tabela 6; e a segunda abordagem usando técnicas estatísticas, como divulgado na Tabela 5. A última abordagem identificada neste MSL engloba a criação de sistemas com foco na preparação do Enade. As principais características desses sistemas estão apontadas na Tabela 7. Alguns trabalhos, mesclaram o uso da primeira e segunda abordagem, geralmente utilizando estatística descritiva com algum método de *machine learning* [38, 42].

Em relação aos trabalhos que utilizaram técnicas de *machine learning*, foram identificados 05 objetivos. O objetivo mais recorrente foi o de identificar fatores que estão associados à nota e para esta finalidade a maioria dos algoritmos utilizados foram do tipo classificação (Tabela 6). Alguns trabalhos focaram no agrupamento dos dados para verificar o desempenho das IESs de acordo com o Conceito Enade (CE), outros tiveram interesse em investigar padrões entre as questões do Questionário do Estudante (QE).

No que se refere ao uso de estatística percebe-se que os trabalhos incluídos, mesmo usando a mesma técnica, possuem objetivos mais heterogêneos. Foram encontrados 14 objetivos, como mostra a Tabela 5. O objetivo mais recorrente entre esses trabalhos, foi o de descrever o conjunto de dados com uso da estatística descritiva. Acerca dos trabalhos direcionados para a área de Computação, 05 utilizaram estatística, 04 utilizaram algoritmos de *machine learning* e 04 foram propostas de sistemas. O trabalho de Vista et al. [42] utilizou técnicas estatísticas e técnicas de *machine learning*.

3.3 Dos trabalhos de análise de dados, quais foram as técnicas estatísticas mais usadas?

A Tabela 5 apresenta as técnicas estatísticas identificadas nos trabalhos incluídos nesta pesquisa. Percebe-se que grande parte dos trabalhos fizeram uso de estatística descritiva (ED) [8, 9, 19, 22, 23, 26, 38]. Alguns deles fizeram uso da ED para descrever os dados dos microdados do Enade [19, 22, 42], outros para descrever questões [23], outros para descobrir fatores que impactam na nota do estudante [8, 26], outros para a partir da descrição dos dados realizar melhoria do Projeto Pedagógico do Curso (PPC) [9], como também para realizar comparações por meio do Conceito Enade (CE) [38].

A Teoria de Resposta ao Item (TRI) também esteve presente em algumas análises [1, 10, 13, 29] e foi útil para medir o desempenho dos estudantes, e todos utilizaram o modelo logístico de três parâmetros (ML3P). Outras técnicas utilizadas envolvem balanceamento por entropia e pareamento por escore de propensão para múltiplos tratamentos [2], Análise fatorial confirmatória e Modelagem por Equações Estruturais [33], Cohen's d [27], Alfa de Cronbach, teste U de Mann-Whitney e teste de Kruskal-Wallis [26].

Com destaque aos trabalhos que foram direcionados para cursos da área de Computação, todos que utilizaram estatística fizeram uso de estatística descritiva [9, 19, 22, 23, 42].

Tabela 3: Artigos da área de Computação

Ref.	Título
Silva et al. [36]	Análise do desempenho dos alunos de cursos superiores em computação no Enade – uma abordagem usando mineração de dados
Vista et al. [42]	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
Lima et al. [22]	Análise de conteúdo das provas do Enade para os alunos do curso de Bacharelado em Ciência da Computação
Landes and Manhães [19]	Análise dos exames do Enade para os cursos de computação utilizando o software R
Figueiro et al. [12]	Análise de Agrupamento Hierárquico aplicada aos microdados do ENADE do curso de graduação em Ciência da Computação
Rosa et al. [32]	Estudo Exploratório através de Análises Longitudinais aplicado à Ciência da Computação a partir da Base de Dados do Enade
Ribeiro and C Junior [30]	Plataforma para auxílio na preparação de estudantes para as avaliações do ENADE e POSCOMP
Barbosa and Carvalho [4]	Uso continuado de ferramenta computacional baseada em questões do Enade para a avaliação diagnóstica de alunos de Sistemas de Informação: um estudo de caso
Lima et al. [23]	Content Analysis of Student Assessment Exams
França et al. [14]	ToSabendo: A Platform to Create Engaging Teaching and Learning Experiences
Charão et al. [9]	Explorando Resultados por Questão no Enade em Ciência da Computação para Subsidiar Revisão de Projeto Pedagógico de Curso
Cunha et al. [11]	Análise Automática com os Microdados do ENADE para Melhoria do Ensino dos Cursos de Ciência da Computação

Tabela 4: Áreas de conhecimento

Área	Artigos
Ciências Exatas e da Terra	[36][19][42][22] [12][32][23][9][11] [39][29][13]
Ciências Biológicas	[1]
Engenharias	[27] [28]
Ciências da Saúde	[38]
Ciências Sociais Aplicadas	[7]
Ciências Humanas	[26] [37]

*trabalhos da área de Computação em negrito

3.4 Dos trabalhos de análise de dados, quais foram as técnicas de mineração ou algorítmicas usadas?

A Tabela 6 apresenta os trabalhos que fizeram uso de técnicas de mineração ou algorítmicas. Por meio da coluna objetivo é possível identificar o que o trabalho quis alcançar. Seguidamente, a Tabela 6 apresenta a técnica de *machine learning* utilizada, o tipo dessa técnica, e por fim, quais artigos fizeram uso da técnica com aquele objetivo. Dos 32 trabalhos incluídos, 14 fizeram uso de algoritmos de *machine learning* [3, 11, 12, 20, 24, 25, 28, 31, 32, 36-39, 42]. Desses, 05 trabalhos foram da área de Computação.

Ao analisar a Tabela 6, percebe-se que grande parte dos algoritmos utilizados tiveram a finalidade de classificar e a técnica mais utilizada foi Árvore de Decisão, a qual foi implementada por 05 trabalhos [3, 24, 31, 32, 37].

O Conceito Enade (CE) foi alvo de 05 trabalhos. Martins et al. [24] utilizou o J48 para gerar regras de decisão e partir dessas regras conseguir prever um aumento do CE nos anos subsequentes do Enade. Já Vista et al. [42] e Figueiro et al. [12] utilizaram Agrupamento Hierárquico para agrupar instituições com CE bons e ruins. Nunes [28] utilizou os microdados do Enade para treinar uma Rede Neural Artificial (RNA) capaz de prever o CE das IESs. O K-means foi utilizado por Silva et al. [38] para realizar agrupamentos dos dados em relação ao conceito Enade e descobrir quais categorias administrativas e regiões das IES possuem menores e maiores CE.

Fatores associados a nota do estudante no Enade foi objeto de investigação de 06 trabalhos. Rosa et al. [32] fizeram um estudo para verificar quais dos algoritmos Árvore de Decisão, Floresta Aleatória ou Máquina de Vetores de Suporte (SVM) classificava melhor os fatores que impactam na nota do estudante, os fatores foram determinados usando dados socioeconômicos, no cenário estudado pelos autores o algoritmo J48 obteve melhor acurácia (71.54%). Silva et al. [37] executaram os algoritmos Árvore de Decisão, Floresta Aleatória, Gradient boosting Tree e Naive Bayes e o que teve melhor desempenho para classificar fatores associados a notas foi o Naive Bayes, com 98.21% de acurácia. Os fatores utilizados por Silva et al. [37] foram algumas questões do questionário socioeconômico (QS) respondidas pelos estudantes. Com a mesma finalidade, Silva et al.

[36] utilizou o algoritmo *Apriori* e apresentou 10 regras de associação com confiança maior que 80% e *lift* maior que 3, para isso os autores utilizaram 10 questões do QS. Leão et al. [20] também utilizaram regras de associação com o mesmo objetivo de Silva et al. [36] e identificou 30 regras baseadas no QS. Já Silva et al. [39] utilizou Regressão Linear para a identificação de variáveis associadas as notas dos estudantes, os autores identificaram 13 variáveis dentre 31 selecionadas que impactavam a nota do estudante.

Em relação a Nota do estudante, foi identificado 01 trabalho com este objetivo. Araújo et al. [3] utilizou Árvores de decisão para prever o desempenho dos estudantes, para isso os autores utilizaram os modelos CART, J48 e o Floresta Aleatória analisando vários atributos dos microdados do Enade para prever a Nota bruta da prova. O modelo classificatório que teve maior acurácia no estudo de Araújo et al. [3] foi o CART que obteve 79.78% de acurácia utilizando 70% das instâncias para treino e 30% para teste.

Com a finalidade de verificar relações entre as questões do questionário do estudante (QE), Matos et al. [25] utilizaram o algoritmo *Apriori* e conseguiram identificar 20 associações para o ano de 2014, 72 para 2015 e 84 para o ano de 2016. As regras encontradas possuíam confiança acima de 75%.

O tempo de permanência dos estudantes foi estudado por Rodrigues and Gouveia [31]. Os autores fizeram uso de Árvore de decisão, Floresta Aleatória, *eXtreme Gradient Boosting* (XGBoost) e RNA e perceberam que os modelos XGBoost se destacaram em todos os cenários estudados. Para chegar a essa conclusão os autores levaram em consideração as métricas de acurácia, precisão, *recall* e *f1-score*. Os autores utilizaram no modelo classificatório 30 atributos dos microdados do Enade, e para determinar o tempo de permanência realizaram a diferença entre o ano de conclusão do ensino médio e o ano de início da graduação do estudante.

3.5 Quais foram as características principais dos sistemas com foco no Enade?

Esta pesquisa identificou 06 propostas que desenvolveram sistemas para analisar dados do Enade. As principais características encontradas nesses sistemas estão apresentadas na Tabela 7. Das 06 propostas encontradas, 04 delas direcionaram seu estudo para cursos de Computação [4, 14, 23, 30], as quais serão apresentadas abaixo.

Santos et al. [34] propuseram um sistema *desktop* que armazenava um banco de questões inseridos por professores e gerava a prova no formato do Enade. Araújo et al. [3] apresentaram uma ferramenta de análise exploratória automatizada para os microdados do Enade e essa ferramenta junto com um modelo classificatório foi capaz de prever o desempenho do estudante e comparar cursos e IES.

Em relação aos trabalhos que foram direcionados para cursos de Computação, o gerenciamento das questões esteve presente em todos os trabalhos [4, 14, 23, 30]. Em Barbosa and Carvalho [4] e Ribeiro and C Junior [30] o cadastro de questões é realizado por professores ou administradores do sistema. Já em Lima et al. [23], as questões são cadastradas por quem está gerenciando a plataforma e em França et al. [14] as questões podem ser cadastradas por professores e alunos, caso a questão seja cadastrada por um aluno é necessária a validação da questão por um professor. Nos trabalhos

de Barbosa and Carvalho [4], Ribeiro and C Junior [30], as questões são associadas a competências e conteúdos. No trabalho de Lima et al. [23], por exemplo, as questões são cadastradas por temas que possuem dicionários de palavras associadas.

Nos trabalhos Barbosa and Carvalho [4], França et al. [14], Ribeiro and C Junior [30], as questões podem ser organizadas em simulados e os estudantes respondem a avaliação no próprio sistema e são gerados relatórios de desempenho para professores e alunos.

Nesses trabalhos, *Dashboards* disponíveis para os alunos apresentam o desempenho do estudante em uma avaliação específica. Já os professores podem verificar o desempenho da turma por meio de indicadores [4, 14, 30].

Elementos de gamificação foram identificados em França et al. [14]. Nessa plataforma, cada *quiz* é apresentado em um tabuleiro que possui um caminho a ser seguido, no tabuleiro existem casas as quais contém níveis diferentes de dificuldade.

A utilização de dados dos microdados fornecidos pelo Inep foi observado em um dos trabalhos [23]. Lima et al. [23] apresentam um sistema que permite selecionar os microdados a serem processados e então faz a catalogação das questões do exame aos temas cadastrados na plataforma de forma automática por meio dos dicionários, após a catalogação automática o sistema calcula a pontuação de cada aluno de acordo com o tema, e por fim o sistema realiza análises estatísticas e apresenta relatórios.

Dois dos sistemas foram desenvolvidos especificamente para a plataforma *web* [4, 14], 01 dos trabalhos é um sistema *desktop* [23] e o trabalho de Ribeiro and C Junior [30] foi desenvolvido para plataforma *web* e *mobile*.

3.6 Quais os benefícios e limitações dos sistemas apontados pelos autores?

Descreve-se nesta seção os benefícios e as limitações dos 04 trabalhos que direcionaram suas propostas para cursos de Computação.

Barbosa and Carvalho [4] aplicaram um questionário com os professores para avaliar o sistema, por meio das respostas é possível observar que o processo utilizado no curso de estudo para diagnosticar competências utilizando o sistema *Question* foi satisfatório; que foi possível identificar em quais competências era necessário realizar um reforço escolar; que foi possível diagnosticar turmas em específico que necessitavam de apoio de conteúdos; que a aplicação de simulados foi suficiente para diagnosticar os estudantes e servir de apoio para estratégias que precisavam ser colocadas em ação pelo colegiado do curso; e que o sistema facilitou a catalogação das questões por competência, a aplicação dos questionários e a análise das respostas. Foi possível identificar também algumas limitações, como por exemplo a falta de periodicidade da aplicação de simulados para diagnosticar estudantes e que os estudantes devem ser treinados para responder a questões no formato do Enade. Esse ponto foi levantado pelo fato de a maioria dos estudantes não terem familiaridade com o formato das questões do Enade.

Ribeiro and C Junior [30] aplicaram um questionário com os estudantes que utilizaram o sistema e foi possível identificar que após aprender a utilizar o sistema os estudantes afirmaram que seria fácil usá-lo da próxima vez. Também foi possível identificar que o sistema trabalhou muito bem com a adição de instrução

Tabela 5: Técnicas estatísticas

Objetivo	Técnica	Artigos
Analisar as propriedades psicométricas da prova	Regressão logística (TRI)	[1]
Analisar viabilidade de um teste adaptativo	Regressão logística (TRI)	[29][10]
Analisar questões	Regressão logística (TRI)	[13]
Comparar (notas) categorias administrativas (pública e privada)	Teste U de Mann-Whitney	[26]
Comparar (notas) entre as regiões geográfica	Teste de Kruskal-Wallis	[26]
Conceito Enade	Estatística descritiva	[38]
Consistência interna dos itens	Alfa de Cronbach	[26]
Descrever o conjunto dos dados	Estatística descritiva	[42][22][19]
Descrever questões	Estatística descritiva	[23]
Fatores associados a nota	Estatística descritiva	[8][26]
Ganhos em relação a FG e CE	Cohen's d	[27]
Melhoria do PPC	Estatística descritiva	[9]
Nota	Regressão logística (TRI)	[7]
	Balanceamento por entropia	[2]
Satisfação dos estudantes	Análise fatorial confirmatória	[33]

* **trabalhos da área de Computação em negrito**

Tabela 6: Técnicas de machine learning

Objetivo	Técnica	Tipo	Artigos
Conceito Enade	Árvore de decisão	Classificação	[24]
	Clusterização	Agrupamento	[42][12]
	K-means	Agrupamento	[38]
	RNA	Classificação	[28]
Fatores associados a nota	Árvore de decisão	Classificação	[32][37]
	Floresta aleatória	Classificação	[32][37]
	Gradient boosting	Classificação	[37]
	Naive Bayes	Classificação	[37]
	Regras de associação	Associação	[36][20]
	Regressão Linear	Regressão	[39]
Nota	SVM	Classificação	[32]
	Árvore de decisão	Classificação	[3]
	Floresta aleatória	Classificação	[3]
Padrões do QE	Regras de associação	Associação	[25]
Tempo de permanência	Árvore de decisão	Classificação	[31]
	Floresta aleatória	Classificação	[31]
	Gradient boosting	Classificação	[31]
	RNA	Classificação	[31]

* **trabalhos da área de Computação em negrito**

de aula/disciplina; que os estudantes preferem aprender usando ambos ambientes virtuais e aulas presenciais; que o sistema pode ser utilizado para auxiliar na aprendizagem e avaliação de disciplinas da área de Computação; que os estudantes podem utilizar a ferramenta para revisar conteúdos e avaliar seu desempenho nas disciplinas; e que professores também podem utilizar a ferramenta dentro do ambiente presencial. Como limitação, é apresentada a necessidade de criação de uma interface de ajuda para auxiliar os usuários no entendimento do sistema e a necessidade de criação de interfaces que permitam aos alunos e professores registrarem informações sobre as questões que erraram.

Lima et al. [23] apresentam como benefício a identificação da dificuldade das competências e conteúdos nas provas do Enade e porcentagem de estudantes que conseguiriam obter êxito. A limitação colocada pelos autores é a complexidade de generalização da proposta para outros cursos que não sejam do curso de Ciência da Computação.

França et al. [14] aponta como benefícios o uso da gamificação que tornou a ferramenta mais interativa. A limitação apontada pelos autores é que a ferramenta não fazia uso de estratégias de *machine learning* que poderia aprimorá-la.

Tabela 7: Características encontradas nos sistemas

Característica	[4]	[30]	[23]	[14]	[34]	[3]
Cadastro de Questões	X	X	X	X	X	
Dashboard Aluno	X	X		X	X	
Dashboard Professor	X	X		X	X	
Gamificação				X		
Machine Learning						X
Microdados			X			X
Plataforma <i>desktop</i>			X		X	
Plataforma <i>mobile</i>		X				
Plataforma <i>web</i>	X	X		X		X
Simulados	X	X		X	X	

* trabalhos da área de Computação em negrito

4 DISCUSSÃO

Esta seção apresenta uma discussão referente ao recorte dos 12 estudos encontrados da área de Computação. Em relação a esses trabalhos, percebeu-se que a maior parte estão associados ao uso de estatística descritiva [9, 19, 22, 23, 42] e *machine learning* [11, 12, 32, 36, 42]. No aspecto de que existem muitos trabalhos que usam técnicas estatísticas, este MSL corrobora com Lima [21]. Mas contrário a Lima [21], este trabalho percebeu um incremento de trabalhos que utilizam técnicas de *machine learning* nos últimos anos. Como por exemplo, Vista et al. [42] de 2017, Figueiro et al. [12] de 2018, Silva et al. [36] de 2019, Rosa et al. [32] de 2021 e Cunha et al. [11] de 2021.

Em relação a propostas de sistemas com foco no Enade, esta pesquisa encontrou 04 trabalhos [4, 14, 23, 30]. Alguns desses sistemas apresentaram características similares entre si, como apresentado na Tabela 7, e outros trouxeram características mais inovadoras, como o uso de *machine learning* e gamificação. Os sistemas discutidos foram publicados em forma de artigos e são recentes, como Lima et al. [23] de 2018, Barbosa and Carvalho [4] de 2020, França et al. [14] de 2021 e Ribeiro and C Junior [30] de 2022. Como a publicação desses trabalhos aconteceram em anos recentes, isso pode representar que o crescimento de estudos com dados do Enade utilizando técnicas estatísticas e de *machine learning* podem incrementar esses sistemas em um futuro próximo de acordo com o amadurecimento das pesquisas realizadas nesta área.

4.1 Uso da estatística com dados do Enade de Computação

A Tabela 5 apresenta que todos os trabalhos que fizeram uso de dados da área de Computação e fizeram uso de estatística, utilizaram da estatística descritiva (ED)[9, 19, 22, 23, 42]. Dos 05 trabalhos 03 tiveram como objetivo descrever o conjunto dos dados presente nos microdados [19, 22, 42], 01 fez uso de ED para descrever questões [23] e 01 com a finalidade de Melhoria do PPC [9].

Com esta MSL percebe-se que as análises estatísticas analisadas pelos trabalhos que fizeram uso de dados do Enade de Computação nos últimos anos foram importante para descrever os dados e sintetizar de forma mais conclusiva os dados disponíveis pelos microdados do Enade. Também foi possível identificar a preocupação pedagógica por trás da análise, como por exemplo a melhoria

do processo avaliativo e/ou melhoria da matriz curricular dos cursos de computação. Visto que é por meio das avaliações externas que se consegue identificar as necessidades específicas dos cursos, e entender se esses cursos atendem ou não ao disposto nas Diretrizes Curriculares Nacionais (DCN). As análises realizadas pelos trabalhos possibilitam que a partir dos dados, possa-se chegar a conclusões que beneficiam as discussões para melhoria dos cursos das IESs, como apontado por Charão et al. [9], Lima et al. [23].

Ainda no que se refere a descrever os dados, percebeu-se que as análises estatísticas realizadas permitiram entender aspectos relacionados ao público que realiza o Enade de Computação que não estavam claros nos relatórios disponibilizados pelo Inep. Percebe-se que a liberação dos dados no formato de microdados permite que pesquisadores possam explorá-los. Entre esses aspectos relacionados ao público pode-se apontar as características descritas por Landes and Manhães [19] e exploração das notas dos alunos apresentada por Cunha et al. [11].

De uma maneira geral, os outros trabalhos que não tiveram direcionamento para a área da Computação fizeram mais uso da Teoria de Resposta ao Item (TRI) com a finalidade de analisar as propriedades psicométricas da prova, analisar viabilidade de um teste adaptativo, analisar questões e analisar notas. Um outro ponto, é que percebe-se que alguns trabalhos avançam com a utilização de técnicas híbridas, ou seja, não só o uso de técnicas estatísticas mas também a utilização de técnicas de *machine learning*.

4.2 Uso de *machine learning* com dados do Enade de Computação

Como percebe-se na Tabela 6, 02 trabalhos de Computação fizeram uso de técnicas de agrupamento [12, 42], 01 usou classificação [32] e 01 usou regras de associação [36]. O trabalho de Cunha et al. [11] realizou um estudo com dados do curso de Ciência da Computação que usa técnicas de *machine learning* para produzir informações relevantes para o coordenador do curso, contudo os autores não apresentaram qual técnica foi utilizada. Por este motivo o trabalho não foi listado na Tabela 6. Os autores apresentam uma comparação entre as notas de uma IES e as notas gerais do percentual de acertos dos estudantes entre os anos de 2014 e 2017 catalogados por conteúdos [11].

Com esta MSL, percebeu-se que a partir do uso das técnicas de *machine learning* é possível analisar quais informações presentes nos microdados do Inep podem impactar de forma positiva ou negativa na nota dos estudantes e também analisar os grupos sociais por meio de indicadores de escolaridade da mãe e de dependência administrativa da IES. Os dados podem ser utilizados para melhoria do processo de ensino aprendizagem das IESs tendo em vista a comparação dos seus dados com esses estudos. Outro ponto, é que por meio do agrupamento foi possível verificar semelhanças entre as instâncias e investigar os reais motivos dos bons ou ruins desempenhos dos estudantes.

Dos trabalhos de Computação que fizeram uso de *machine learning* percebe-se que o objetivo mais comum foi identificar fatores que estão associados a nota do estudante no exame e ao agrupamento de IESs por meio do conceito Enade.

4.3 Propostas de sistemas com foco no Enade em Computação

Com esta pesquisa, percebeu-se que existem poucas propostas de sistemas com foco no Enade. As propostas que existem estão limitadas a aplicação de questionários com questões do Enade, análise de microdados do Inep e uso da gamificação.

Nos sistemas encontrados a aplicação das provas apresentavam dados quantitativos de questões por temas, como a porcentagem de acerto e a porcentagem de erro. Também apresentavam relatórios de desempenho dos estudantes nas avaliações de forma individual e em turmas. Pelo fato do Enade trabalhar com competências, percebe-se que esses sistemas precisam evoluir utilizando modelos que consigam rastrear o conhecimento dos alunos e apresentar essa evolução para intervenção no processo de ensino-aprendizagem, como sugerido por França et al. [14].

Em uma das propostas percebeu-se o uso da gamificação. Sobre essa proposta nota-se que a gamificação foi importante para trabalhar a motivação dos estudantes dentro do processo de melhoria das competências. Contudo, o tema da gamificação ainda precisa ser melhor investigado, pois assim como existem estudos que apresentam benefícios [41] existem outros que apontam que não existem evidências suficientes desses benefícios [35].

4.4 Implicações

As implicações deste trabalho estão relacionadas ao conhecimento das técnicas estatísticas e de *machine learning* que estão sendo utilizadas com foco na análise de dados do Enade ou preparação de estudantes para o Enade.

Espera-se que os resultados dessa revisão possam beneficiar as IES e/ou pesquisadores de computação que pretendem realizar estudos com dados do Enade, ou conhecer ferramentas propostas pela comunidade científica para o acompanhamento dos estudantes que realizam tal prova.

4.5 Limitações e Ameaças à validade

Este estudo possui as seguintes limitações: É Específico para a área de Computação; Analisa apenas técnicas estatísticas, técnicas de *machine learning* e sistemas desenvolvidos; Analisa apenas trabalhos dos últimos 10 anos; A qualidade dos trabalhos incluídos nesta pesquisa determina a qualidade deste MSL; e O entendimento dos autores sobre os trabalhos incluídos influencia este MSL.

Em relação as ameaças à validade, a utilização de métodos sistemáticos visam desviar de vieses e deixar a análise dos resultados mais objetiva. Dentro deste trabalho as ameaças foram diminuídas pelo estabelecimento de um protocolo de pesquisa e a sua discussão com 02 especialistas do domínio durante cada etapa do processo. No entanto, a execução do protocolo se deu por 02 pesquisadores o que pode acontecer da ameaça do viés permanecer de forma parcial.

Nesse aspecto, é possível que algum estudo importante não tenha feito parte desta análise. Essa ameaça pode acontecer por alguns motivos, como: i) as fontes de estudo não foram indexadas pelas bases de dados que utilizamos; ii) o estudo não foi coberto pela *string* de busca e iii) o estudo está em um idioma diferente do inglês ou português. Para amenizar essa ameaça, utilizamos bases de dados eletrônicas relevantes frequentemente consultadas para

pesquisas e revisões sistemáticas. Além de que, muitas tentativas foram realizadas para construir a versão final da *string* de busca.

5 CONSIDERAÇÕES FINAIS

Este trabalho apresentou um Mapeamento Sistemático da Literatura (MSL) que identificou e discutiu quais técnicas são usadas para analisar dados do Enade ou implementar ações diagnósticas para o acompanhamento das competências necessárias para o perfil do egresso do estudante que realiza tal exame. Após as buscas iniciais foram identificados 129 trabalhos, e com a aplicação dos critérios de seleção foram incluídos 32 estudos. Em especial, este artigo discutiu 12 trabalhos que focaram especificamente na área de Computação.

Em relação à estatística usada, todos os trabalhos da área de Computação fizeram uso de estatística descritiva com a finalidade de descrever os dados dos microdados, descrever as questões e utilizar dados da estatística para melhorar o Projeto Pedagógico do Curso (PPC). Percebe-se nesses trabalhos que por meio da estatística descritiva é possível entender a composição do público que compõe cursos de Computação no Brasil em relação a diversos fatores (sexo, idade, renda). Também identificou-se que é possível por meio dessas análises mapear conteúdos que determinados cursos precisam avançar e assim realizar melhorias no PPC verificando a devida atenção para os mesmos.

No que se refere às técnicas de *machine learning* utilizadas pelos trabalhos, a maioria fez uso de técnicas de agrupamento por meio do algoritmo *Hierarchical Clustering*. Esses agrupamentos foram usados para verificar semelhanças entre as instâncias (IESs). A motivação era a de investigar possíveis causas dos desempenhos bons ou ruins dos estudantes. Dos objetivos, percebeu-se que esses trabalhos se concentraram em agrupar IES por meio do Conceito Enade ou identificar fatores que impactam nas notas dos estudantes. Esses fatores estão associados às respostas que os estudantes deram ao questionário socioeconômico.

No que diz respeito aos sistemas propostos com a finalidade do Enade para a área da Computação, foram identificados poucos sistemas e esses estão limitados a aplicação de questionários com questões do Enade, análise de microdados do Inep e uso da gamificação. Contudo, existe também uma oportunidade de extensão dos mesmo considerando modelos que consigam rastrear o conhecimento dos alunos e apresentar essa evolução para intervenção no processo de ensino-aprendizagem.

5.1 Trabalhos Futuros

Como trabalhos futuros, sugere-se que sejam pensados modelos computacionais que possam diagnosticar estudantes de acordo com as competências exigidas para o perfil do egresso nas Diretrizes Curriculares Nacionais (DCN) e avaliadas pelo Enade.

Analisar o desempenho dos estudantes impacta na verificação do processo de ensino aprendizagem em relação ao perfil do egresso definido pelas DCNs de um determinado curso. Essa preocupação antes de tudo, deve estar presente no processo de ensino aprendizagem desde os alunos ingressante até os alunos egressos. Nesse aspecto faz-se necessário desenvolver sistemas que utilizem de técnicas de recomendação para sugerir questões que um estudante deve responder para avaliar melhor suas competências, como também

sugerir materiais de estudos baseado nas deficiências identificadas desses estudantes.

AGRADECIMENTO

À Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) pelo apoio financeiro para o desenvolvimento deste trabalho.

REFERÊNCIAS

- [1] Kauê Tortato Alves et al. 2020. Análise das propriedades psicométricas da prova de conhecimentos específicos de licenciatura em ciências biológicas no Enade 2017.
- [2] Antonia Amanda Araujo, Alesandra de Araújo Benevides, Francisca Ziliana Mariano, and Rafael Barros Barbosa. 2020. Diferencial de desempenho dos estudantes cotistas no Exame Nacional de Desempenho de Estudantes: evidências sobre as instituições de ensino superior federais. *Revista Brasileira de Educação* 25.
- [3] Rodrigo Alexandrino Araújo et al. 2019. Análise dos microdados do Enade: Proposta de uma ferramenta de exploração utilizando mineração de dados.
- [4] Pedro Luis Saraiva Barbosa and Windson Viana Carvalho. 2020. 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, 1743–1752.
- [5] Brasil. 2004. LEI N° 10.861, DE 14 DE ABRIL DE 2004. *Diário Oficial [da] República Federativa do Brasil*. http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/110.861.htm
- [6] Brasil. 2022. EDITAL N° 51, DE 24 DE JUNHO DE 2022. *Diário Oficial [da] República Federativa do Brasil*. <https://www.in.gov.br/en/web/dou/-/edital-n-51-de-24-de-junho-de-2022-411675066>
- [7] Raphael Vinicius Weigert Camargo, Rita de Cássia Correa Pepinelli Camargo, Dalton Francisco de Andrade, and Antonio Cezar Bornia. 2016. Desempenho dos alunos de ciências contábeis na prova ENADE/2012: uma aplicação da Teoria da Resposta ao Item. *Revista de Educação e Pesquisa em Contabilidade (REPeC)* 10, 3.
- [8] Bruna Casitagli and Julio Cesar Soares Aragão. 2021. Avaliação do Ensino Superior brasileiro: Desempenho dos estudantes em formação geral. *Revista Portuguesa de Educação* 34, 1.
- [9] Andrea S Charão, Karina Wiechork, Marlon LS Rodrigues, and Fernando P Barbosa. 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*. SBC, 16–20.
- [10] Edy Célia Coelho. 2014. Teoria da resposta ao item: desafios e perspectivas em exames multidisciplinares.
- [11] Renan Cunha, Claudomiro Sales, and Reginaldo Santos. 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*. SBC, 208–217.
- [12] Michele Ferraz Figueiro, Nicolas PB Vista, Patrícia MM Chicon, Angela P Ansj, et al. 2018. 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* 1, 8.
- [13] Luís Carlos Costa Fonseca, Lidinalva de Almada Coutinho, and Lanyllo Araujo dos Santos. 2013. Ferramenta para estimar a proficiência de indivíduos baseada na Teoria de Resposta ao Item. *RENTE* 11, 1.
- [14] Tarsys Fonseca França, Cibele Oliveira Ferreira, Daiane Leandro De Oliveira, Guilherme Tavares De Assis, Anderson Almeida Ferreira, and Elton José Da Silva. 2021. TôSabendo: A Platform to Create Engaging Teaching and Learning Experiences. In *2021 XVI Latin American Conference on Learning Technologies (LACLO)*. IEEE, 275–281.
- [15] Inep. 2022. Exame Nacional de Desempenho dos Estudantes (Enade). [urlhttps://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade](https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade).
- [16] Inep. 2022. Exame Nacional de Desempenho dos Estudantes (Enade). [urlhttps://www.gov.br/inep/pt-br/centrais-de-conteudo/legislacao/enade](https://www.gov.br/inep/pt-br/centrais-de-conteudo/legislacao/enade).
- [17] Inep. 2022. Inep divulga indicadores que avaliam cursos e instituições. [urlhttp://portal.mec.gov.br/](http://portal.mec.gov.br/).
- [18] Staffs Keele et al. 2007. Guidelines for performing systematic literature reviews in software engineering.
- [19] Flávio Barreiro Landes and Laci Mary Barbosa Manhães. 2018. Análise dos exames do ENADE para os cursos de computação utilizando o software R. *Revista do Seminário Internacional de Estatística com R* 3, 1.
- [20] Heloise Acco Tives Leão, Edna Dias Canedo, Marcelo Ladeira, and Fabiano Fagundes. 2018. Mining enade data from the ulbra network institution. In *Information Technology-New Generations*. Springer, 287–294.
- [21] Priscila da Silva Neves Lima. 2018. *Análise das Provas do Enade Agrupadas por Tema: Um Estudo de Caso para Estudantes de Ciência da Computação*. Ph.D. Dissertation. Dissertação (Mestrado)-Universidade Federal de Goiás.[GS Search].
- [22] Priscila da Silva Neves Lima, Ana Paula Laboissière Ambrósio, João Lucas dos Santos Oliveira, and Cedric Luiz de Carvalho. 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.
- [23] Priscila da Silva Neves Lima, Ana Paula Laboissière Ambrósio, Igor Moreira Félix, Jacques Duilo Brancher, and Deller James Ferreira. 2018. Content Analysis of Student Assessment Exams. In *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–9.
- [24] Ernane Rosa Martins, Davi Taveira Alencar Alarcão, and Sibelius Lellis Vieira. 2021. Descoberta de conhecimento aplicado ao Enade.
- [25] Alexandre Moraes Matos, Lucas Ribeiro Reis de Sousa, Ismael Pontes Torres Júnior, Renato Marinho Alves, Heloise Acco Tives Leão, and Fabiano Fagundes. 2020. Mineração dos Dados do Enade: Avaliação do questionário do estudante das Instituições do Norte do Brasil.
- [26] Antonio Evanildo Cardoso Medeiros Filho. 2019. *Percepção discente e desempenho dos estudantes dos cursos de licenciatura em Educação Física no Enade 2017*. Master's thesis. Universidade Estadual do Ceará - UECE.
- [27] Tatiana Melguizo and Jacques Wainer. 2016. Toward a set of measures of student learning outcomes in higher education: evidence from Brazil. *Higher Education* 72, 3, 381–401.
- [28] Ricardo Tocantins Modugno Nunes. 2018. Cálculo preditivo de classificação das notas do Enade utilizando Redes Neurais Artificiais. *Revista de Tecnologia Aplicada* 7, 2.
- [29] Jean Piton-Goncalves. 2020. Testes adaptativos para o Enade: uma aplicação metodológica. *Revista Meta: Avaliação* 12, 36, 665–688.
- [30] Tiago C Ribeiro and Wilson F C Junior. 2022. Plataforma para auxílio na preparação de estudantes para as avaliações do Enade e POSCOMP.
- [31] Ebony M Rodrigues and Roberta MM Gouveia. 2021. Técnicas de Machine Learning para Predição do Tempo de Permanência na Graduação no Âmbito do Ensino Superior Público Brasileiro. In *Anais do VI Congresso sobre Tecnologias na Educação*. SBC, 128–137.
- [32] Eliene Ribeiro Rosa, Deller James Ferreira, Nádia Félix Felipe da Silva, and Alfredo Assis. 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.
- [33] Agenor Sousa Santos, Maria José Pereira Dantas, and Ricardo Luiz Machado. 2017. Structural equation modeling applied to assess industrial engineering students' satisfaction according to Enade 2011. *Production* 27.
- [34] Perycles Jannser Lopes Santos, José Hélio Luna Neto, Noberto Carvalho Rocha, and Frago Florentin. 2016. E-prova: Sistema para Elaboração de Avaliações no Padrão Enade.
- [35] E Schnepf. 1993. From prey via endosymbiont to plastids: comparative studies in dinoflagellates. In *Origins of Plastids* (2nd ed.), R A Lewin (Ed.). Chapman and Hall, New York, 53–76.
- [36] Alexander Figueiredo Silva, Raphael Magalhães Hoed, and Pedro Fábio Saraiva. [n.d.]. Análise do Desempenho dos Alunos de Cursos Superiores em Computação no Enade—Uma abordagem usando Mineração de Dados. *WWW/INTERNET 2019*, 207.
- [37] Ivonaldo Silva, Marcia Terra Silva, and Pedro Luiz Oliveira Costa Neto. 2021. Descoberta dos fatores preditivos de sucesso dos alunos do Curso de Tecnologia em Logística, nos exames do ENADE, por meio da Mineração de Dados Educacionais. *2º International Conference on Network Enterprises and Logistics Management*.
- [38] Jesaias Carvalho Pereira Silva, Everton Jales de Oliveira, Francisco Milton Mendes Neto, Araken Medeiros Santos, Lenardo Chaves e Silva, Leonardo Torres Marques, Igor Melo Mendes, Ademar França de Sousa Neto, Jesreel Carvalho Pereira Silva, and Salatiel Dantas Silva. 2020. Data analysis of the performance of Brazilian higher education medicine courses. In *Proceedings of the 10th Euro-American Conference on Telematics and Information Systems*. 1–8.
- [39] LGF Silva, MEPS Rocha, and RAA Fagundes. 2017. Enade: Math and Science Students' Performance Analysis. *IEEE Latin America Transactions* 15, 9, 1742–1746.
- [40] P Stone. 2002. Popping the (PICO) question in research and evidence-based practice. *Nurs Res* 15, 3, 197–198.
- [41] Jan van Elderen and Esther van der Stappen. 2019. The potential impact of gamification elements on the acceptance of technology in the context of education: a literature review.
- [42] Nicolas Pastorio Boa Vista, Michele Ferraz Figueiro, and Patricia Mariotto Mozzaquatro Chicon. 2017. 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* 1, 1.