

DOI: 10.5748/20CONTECSI/PSE/DSC/7339

eLocator: e207339

**PERFORMANCE ANALYSIS OF INFORMATION SYSTEMS COURSES IN ENADE
USING DATA MINING TECHNOLOGIES ANÁLISE DE DESEMPENHO DE CURSOS
DE SISTEMAS DE INFORMAÇÃO NO ENADE A PARTIR DE TECNOLOGIAS DE
MINERAÇÃO DE DADOS**

Eude Monteiro Da Hora – <https://orcid.org/0000-0001-8771-8733>

Universidade Federal Do Pará

Pedro Henrique Fernandes Rendeiro – <https://orcid.org/0000-0001-9486-750X>

Universidade Federal Do Pará

Reginaldo Cordeiro Dos Santos Filho – <https://orcid.org/0000-0002-0456-8547>

Universidade Federal Do Pará

PERFORMANCE ANALYSIS OF INFORMATION SYSTEMS COURSES IN ENADE USING DATA MINING TECHNOLOGIES

ABSTRACT

This article presents a performance analysis of students enrolled in Information Systems courses in the National Student Performance Examination (Enade) through the application of mining algorithms on the exam's microdata. As a case study, the work uses data relating to students at the Federal University of Pará. The main data used for the analysis include: student performance by subject, overall grade, family income category, and the number of weekly study hours. The analysis revealed areas in which students perform above the national average, as well as areas that offer opportunities for improvement. Furthermore, it was found indication that the analyzed socioeconomic variables may be associated with better performance in the exam.

Keywords: Enade, Information Systems, Data Mining, Student Performance, Higher Education.

ANÁLISE DE DESEMPENHO DE CURSOS DE SISTEMAS DE INFORMAÇÃO NO ENADE A PARTIR DE TECNOLOGIAS DE MINERAÇÃO DE DADOS

RESUMO

Este artigo apresenta uma análise de desempenho de discentes de cursos de Sistemas de Informação no Exame Nacional de Desempenho dos Estudantes (Enade) a partir da aplicação de algoritmos de mineração sobre os microdados da prova. Como estudo de caso, o trabalho utiliza dados referentes aos estudantes da Universidade Federal do Pará. Os principais dados utilizados para a análise incluem: desempenho discente por tema, nota geral, categoria de renda familiar e quantidade de horas semanais de estudo. A análise revelou áreas em que os alunos têm desempenho acima da média nacional, bem como áreas que oferecem oportunidades para melhorias, além do indicativo de que as variáveis socioeconômicas analisadas podem estar associadas a melhores desempenhos no exame.

Palavras-chave: Enade, Sistemas de Informação, Mineração de Dados, Desempenho Discente, Educação Superior.

1. Introdução

Aplicado pela primeira vez no ano de 2004 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), o Exame Nacional de Desempenho dos Estudantes¹ (Enade) tem por objetivo principal avaliar o desempenho dos egressos dos cursos de graduação oferecidos por Instituições de Ensino Superior (IES) de todo o país. Nesse sentido, destaca-se que o resultado do exame proporciona dados importantes acerca da qualidade dos cursos ofertados pelas IES avaliadas no exame, e da educação superior nacional de forma geral.

Sob tal ótica, denota-se que tais informações podem ser multimodalmente utilizadas de forma a: identificar fragilidades de ensino relativas aos componentes curriculares dos cursos; oferecer suporte a tomadas de decisão na reestruturação de projetos pedagógicos de curso em IES; descobrir padrões de correlação entre variáveis como a situação socioeconômica dos discentes, quantidade de horas semanais de estudo e seu desempenho na prova; subsidiar a aplicação de políticas pedagógicas de curso com enfoque em melhorias do ensino superior nacional, entre outras finalidades.

Os principais resultados e estatísticas sobre as edições do Enade constam nos microdados da prova disponibilizados no site do INEP², a partir dos quais é possível realizar análises sobre os índices de proficiência discente no exame. Neste contexto, o presente trabalho tem por fito apresentar e discutir o processo de aplicação de algoritmos de mineração de dados visando a descoberta de padrões úteis e não-triviais acerca do desempenho dos discentes concluintes do curso de Sistemas de Informação (SI) da Universidade Federal do Pará (UFPA) no Enade. O trabalho realiza análises comparativas entre o desempenho geral da UFPA e o desempenho nacional na prova do curso de SI por meio da avaliação da proficiência dos discentes nos componentes específicos da prova, bem como explora o aspecto da influência socioeconômica e do cronograma de estudos dos discentes sobre suas performances no exame.

Dessa forma, a análise dos microdados do curso em questão se caracteriza como ferramenta em potencial para a obtenção de informações relevantes acerca da sua qualidade, as quais podem auxiliar no direcionamento de ações pedagógicas voltadas ao seu aprimoramento.

2. Trabalhos relacionados

A etapa inicial da análise teve por escopo realizar a identificação dos principais enfoques e metodologias de estudo presentes na literatura sobre o Enade, de modo a elencar as pesquisas mais relevantes para complementar e direcionar as análises presentes neste trabalho. O levantamento foi realizado por meio da busca pelo termo “Enade” na plataforma *Google Scholar*³, a qual retornou um conjunto de trabalhos relacionado a análises dos microdados da prova. Considerando o conjunto de estudos retornados,

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

² Link: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enade>

³ Link: <https://scholar.google.com.br>

destacaram-se os que apresentaram maior relação ao tema da análise, enfoque sobre a área de computação e diversidade de metodologias utilizadas, os quais constam destacados nos parágrafos subsequentes.

Em Cortelazzo, Júnior e Rodrigues (2021), é realizada uma análise quantitativa da avaliação dos cursos da área de computação no Enade, sendo eles os cursos de Ciência da Computação, Sistemas de Informação e Análise e Desenvolvimento de Sistemas. São analisados fatores como o número de vezes em que um curso obteve conceito acima da média, classificação jurídica da instituição de ensino (pública ou privada), e desempenho geral no exame.

O trabalho de Mizael e Lima (2018) conduz um estudo de caso referente ao desempenho de alunos do curso de Sistemas de Informação da Universidade Estadual de Goiás no Enade 2014. Denota-se como principal objetivo da pesquisa o estabelecimento de correlações entre variáveis socioeconômicas e desempenhos satisfatórios do exame, traçando assim um perfil médio dos discentes que apresentaram boa proficiência na prova. A metodologia do trabalho faz uso do Processo de Descoberta de Conhecimento em Base de Dados (em inglês, *Knowledge Discovery in Databases— KDD*) (Fayyad; Piatetsky-Shapiro e Smyth, 1996) e o Processo Padrão Intersetorial para Mineração de Dados (em inglês, *Cross Industry Standard Process for Data Mining— CRISP-DM*) (Wirth e Hipp, 2020). Uma das conclusões do estudo é que fatores como o cronograma de horas de estudo e a situação socioeconômica do egresso impactam diretamente em seu desempenho no exame.

O trabalho de Neto (2020) aplica técnicas de mineração de dados sobre as bases das edições de 2014 e 2017 do Enade, visando identificar as características institucionais que mais impactam na performance dos discentes de cinco diferentes cursos de computação. Quanto à metodologia do trabalho, destaca-se a seleção de cinco atributos provenientes do Questionário do Estudante (presente no Enade), os quais permitem a realização de inferências sobre a qualidade da IES frequentada pelo aluno, a saber: Organização Pedagógica, Planejamento Didático, Infraestrutura e Facilidades Físicas, Treinamento Acadêmico e Profissional, e Responsabilidade Institucional. As etapas de mineração dos dados seguem a metodologia do processo KDD, enquanto que a sumarização dos resultados é feita por meio de comentários acerca da classificação de grupo das perguntas do questionário.

Já em Silva *et al.* (2020), é feita uma pesquisa de cunho exploratório e descritivo, a qual visa analisar o histórico de cursos de graduação da área de computação no Enade entre os anos de 2008 a 2014. A análise faz uso da base de dados do Censo da Educação Superior de 2015 e se baseia majoritariamente na utilização de estatística descritiva, fazendo uso de gráficos e tabelas na comparação de cursos por conceito, categoria administrativa da IES e região do país. Dentre as conclusões obtidas na análise, destacam-se i) a atribuição de melhor conceito Enade às instituições públicas e ii) a proeminência de Engenharia e Ciência da Computação (Licenciatura) como sendo as áreas melhor avaliadas no exame.

Em Lima *et al.* (2018), foi desenvolvida uma metodologia para classificação das questões do Enade em temas. O trabalho visa viabilizar análises sobre a estrutura do exame, de modo a permitir a identificação dos temas onde os estudantes têm mais facilidade ou dificuldade, além de mostrar como os resultados da prova se correlacionam com a qualidade da estrutura do curso. Foi feito um estudo de caso com a área de Ciência da

Computação no Enade, o qual foi subdividido nas etapas de catalogação dos temas, agrupamento de dados e aplicação de técnicas de análise dos dados.

Tendo em vista as principais preocupações presentes nos trabalhos citados, destaca-se principalmente a importância de realizar a análise de desempenho dos estudantes baseada em procedimentos metodológicos tradicionais da mineração de dados, tal como o KDD. Nota-se também a necessidade de detectar as fraquezas e potencialidades dos discentes que fazem a prova a partir dos temas que são cobrados, além de apontar as melhores instituições de ensino públicas em cada tema. Por fim, apresenta-se também a investigação de possíveis correlações de nota, renda e tempo de estudo.

3. Metodologia

Nesta seção são discutidos os principais pontos relativos à metodologia adotada no desenvolvimento do projeto, a saber: seleção das ferramentas e tecnologias utilizadas na mineração e análise dos dados; coleta e tratamento dos dados de interesse; e por fim, a aplicação dos algoritmos desenvolvidos.

3.1. Seleção das ferramentas utilizadas

Para a produção dos algoritmos de mineração de dados, buscou-se selecionar um conjunto de tecnologias direcionadas à manipulação de grandes volumes de dados, as quais propiciam maior facilidade nas fases de extração, seleção e análise estatística das bases do Enade. Levando em consideração o acervo de *frameworks* disponíveis para a área de ciência de dados, bem como a eficiente integrabilidade existente entre grande parte deles, foi escolhida a linguagem Python para a realização das análises previstas para produção. Ao pesquisar os principais módulos de manipulação de dados presentes na linguagem, buscou-se utilizar os que mais se adequam ao propósito do projeto, sendo assim selecionadas as bibliotecas Matplotlib, Pandas e NumPy como principais ferramentas de aplicação dos algoritmos de análise de dados.

3.2. Coleta e seleção de dados

A base de dados do Enade foi coletada por meio do site oficial do INEP. O curso de Sistemas de Informação foi avaliado no exame nas edições de 2005, 2008, 2011, 2014, 2017 e 2021, sendo os microdados da penúltima os escolhidos para se aplicar a mineração e análise de dados, visto que os dados do curso de SI da UFPA não estão disponíveis na edição de 2021, restando a edição de 2017 como a mais recente em que o curso foi avaliado e recebeu nota.

Com relação aos microdados coletados do Enade 2017, tem-se que eles se apresentam no formato *Comma-Separated Values* (CSV), o qual é o formato padrão de organização dos microdados de todas as edições do exame. Nele, dispõem-se as informações num formato tabular com linhas e colunas separadas por vírgulas, onde as linhas representam os participantes da edição, enquanto as colunas representam diversos atributos distintos relativos aos discentes e.g. idade, sexo, código do curso, cartão-resposta, entre outros.

No que tange à seleção dos dados relevantes ao projeto, destaca-se que foi feita visando pontualmente os estudantes do curso de Bacharelado em Sistemas de Informação. A seleção foi realizada através da filtragem dos discentes baseada no código do curso, o qual

se encontra credenciado no site de Cadastro Nacional de Cursos e Instituições de Educação Superior⁴, mantido pelo Ministério da Educação.

3.3. Tratamento dos dados

A prova do Enade é dividida em questões de formação geral e de componentes específicos, as quais apresentam peso de 25% e 75% no cálculo da nota final do participante, respectivamente. Nesse sentido, dado que o escopo do trabalho incide sobre a análise de desempenho discente nos conteúdos curriculares do curso de Sistemas de Informação, consideraram-se exclusivamente os dados das questões objetivas de componentes específicos para os comparativos de performance por tema da prova, sendo a nota geral obtida no exame considerada apenas para a análise de correlação entre desempenho, situação socioeconômica e cronograma de estudos dos discentes.

Os temas da prova do Enade são um conjunto de áreas de conhecimento específicas de cada curso de graduação avaliado, definidos com base nas Diretrizes Curriculares Nacionais de cada área. Sua estruturação tem como objetivo avaliar o desempenho dos estudantes no que se refere às competências e habilidades esperadas ao final da graduação, incluindo conhecimentos teóricos e práticos, bem como a capacidade de aplicação de tais conhecimentos na resolução de problemas e tomada de decisões.

Ademais, os dados presentes na base foram filtrados de modo a considerar na análise somente os discentes que participaram do exame com resultado válido e com cartão resposta não-nulo nas questões de componente específico.

3.4. Desenvolvimento dos algoritmos de mineração de dados

Com a conclusão das etapas suprarreferidas, deu-se início ao desenvolvimento e aplicação dos algoritmos⁵ para a mineração dos microdados do Enade 2017. O objetivo principal dos algoritmos produzidos nesta fase do trabalho consistiu na geração de gráficos e tabelas relativos ao desempenho dos discentes da UFPA em comparação com o desempenho médio nacional e com o desempenho das IES que obtiveram as melhores notas nos temas individuais da prova, além da correlação entre a nota geral dos discentes no exame, a categoria de sua renda familiar e a quantidade de horas de estudo semanais.

4. Resultados e discussões

O cálculo do desempenho dos discentes do curso de SI de uma dada IES em um tema é calculado a partir da razão entre o número total de acertos nas questões pertinentes ao tema e o número total destas questões no exame. Nesse sentido, os itens a) e b) da Figura 1 representam de formas distintas comparativos do desempenho do curso de SI da UFPA com o desempenho nacional no Enade 2017.

Na Figura 1 item a), é ilustrado o percentual de acerto nacional e da UFPA em cada um dos temas abordados nas questões da prova. Nota-se que a UFPA apresentou um desempenho

⁴ Link: <https://emec.mec.gov.br>

⁵ Todos os códigos relativos à análise apresentada neste estudo estão disponíveis em: <http://tinyurl.com/55zpv7hw>

superior a 80% e acima do desempenho médio nacional nos temas de Segurança e Auditoria de Sistemas e Interação Humano-Computador, enquanto apresentou os piores desempenhos nos temas de Probabilidade e Estatística e Gerência de Projetos. Já na Figura 1 item b) é exibida a razão entre o percentual de acertos da UFPA e o percentual de acertos nacional. Nela, evidencia-se o fato de que o desempenho dos discentes do curso de SI da UFPA foi superior ao desempenho nacional em metade dos 20 temas presentes na prova, sendo Pesquisa Operacional o componente específico onde a UFPA melhor se sobressaiu em relação à performance geral dos demais cursos de SI do país, apresentando um desempenho mais de 30% acima da média geral.

Além do comparativo com o desempenho nacional, foram também estruturados comparativos do desempenho da UFPA com o desempenho das melhores IES por tema, os quais são sumarizados nas Tabelas 1 e 2. A Tabela 1 apresenta um comparativo de performance levando em conta todas as instituições cujos discentes prestaram o exame, na qual se nota que múltiplas IES obtiveram a taxa de acerto máxima em diversos temas. Entretanto, ao se observar o número de participantes destas IES, percebe-se que o alto desempenho percentual é, por vezes, consequência da baixa quantidade de discentes que realizaram a prova.

Tabela 1: Cursos de SI com os melhores desempenhos por tema no Enade 2017, considerando todas as IES que participaram do exame.

Tema	IES com o melhor desempenho	Nº de participantes	Desempenho (%)	
			Melhor curso	UFPA
Algoritmos e Estruturas de Dados	Faculdade CNEC Gravataí	7	100	50,7
Arquitetura da Informação	Faculdade de Taquaritinga (FTGA)	4	100	49,3
Arquitetura e Organização de Computadores	Universidade Estadual do Norte do Paraná (UENP)	4	100	47,89
Arquitetura empresarial e da tecnologia da informação	Centro Universitário Facid Wyden (UniFacid Wyden)	5	80	34,51
Banco de Dados	Faculdade Politécnica de Campinas (POLICAMP)	2	75	35,21
Engenharia de Software	Faculdade Politécnica de Campinas (POLICAMP)	2	50	24,65
Fundamentos, Paradigmas e Linguagens de Programação	Universidade Vila Velha (UVV)	17	100	47,89
Gerência de Projetos	Universidade Paulista (UNIP)	2	50	16,9
Gestão do Conhecimento	Centro Universitário Dinâmica das	8	100	60,56

Cataratas (UDC)				
Governança de Tecnologia da Informação e Comunicação	Universidade Paulista (UNIP)	2	100	35,21
Informática e Sociedade	Universidade do Contestado (UNC)	4	100	61,97
Interação Humano-Computador	Centro Universitário Dinâmica das Cataratas (UDC)	8	100	84,51
Lógica Matemática e Matemática Discreta	Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo (IFES)	7	100	45,07
Modelagem de Sistemas de Informação	Universidade Estadual de Mato Grosso do Sul (UEMS)	4	100	26,76
Pesquisa Operacional	Universidade Paulista (UNIP)	2	100	61,97
Probabilidade e Estatística	Centro Universitário do Sul de Minas (UNIS-MG)	20	80	23,94
Qualidade de Processo e Produto	Faculdade Politécnica de Campinas (POLICAMP)	2	100	61,27
Redes de Computadores e Sistemas Distribuídos	Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ)	2	100	40,85
Segurança e Auditoria de Sistemas	Universidade Paulista (UNIP)	6	100	90,14
Sistemas Operacionais	Universidade do Vale do Taquari (UNIVATES)	7	100	42,25

Considerando este contexto, foi realizado um segundo comparativo, desta vez levando em conta apenas universidades federais públicas, representado na Tabela 2. Tal restrição busca comparar instituições que se enquadram na mesma categoria da UFPA, dadas suas relevâncias no cenário educacional acadêmico e as altas taxas de concorrência por vagas, que geralmente implicam em um maior número de discentes em seus cursos. Nessa perspectiva, a Tabela 2 mostra valores mais diversificados nos melhores desempenhos por tema, destacando-se os temas Segurança e Auditoria de Sistemas e Gerência de Projetos como os que apresentaram o maior e o menor índice de acerto obtido por uma IES na prova, respectivamente. Dentre as IES que constam na Tabela 2, destaca-se a Universidade Federal da Bahia (UFBA) como a universidade que figura mais vezes na lista, apresentando o melhor desempenho em 5 dos 20 temas da prova. Nesse sentido, é válido observar que a performance expressiva da UFBA no exame sugere que seu Projeto Pedagógico de Curso possa ser considerado como um possível elemento norteador para o aperfeiçoamento do curso de SI da UFPA e de outras instituições de ensino do Brasil.

Tabela 2: Cursos de SI com os melhores desempenhos por tema no Enade 2017, considerando somente IES classificadas como universidades federais públicas.

Tema	IES com o melhor desempenho	Nº de participantes	Desempenho (%)	
			Melhor curso	UFPA
Algoritmos e Estruturas de Dados	Universidade Federal do Acre (UFAC)	20	90	50,7
Arquitetura da Informação	Universidade Federal Rural de Pernambuco (UFRPE)	7	85,71	49,3
Arquitetura e Organização de Computadores	Universidade Federal de Mato Grosso do Sul (UFMS)	8	87,5	47,89
Arquitetura empresarial e da tecnologia da informação	Universidade Federal de Mato Grosso do Sul (UFMS)	15	53,33	34,51
Banco de Dados	Universidade Federal da Bahia (UFBA)	18	63,89	35,21
Engenharia de Software	Universidade Federal de Juiz de Fora (UFJF)	19	42,11	24,65
Fundamentos, Paradigmas e Linguagens de Programação	Universidade Federal de Ouro Preto (UFOP)	23	78,26	47,89
Gerência de Projetos	Universidade Federal de Mato Grosso do Sul (UFMS)	8	37,5	16,9
Gestão do Conhecimento	Universidade Federal da Bahia (UFBA)	18	83,33	60,56
Governança de Tecnologia da Informação e Comunicação	Universidade Federal Rural do Rio de Janeiro (UFRRJ)	30	60	35,21
Informática e Sociedade	Universidade Federal de São Carlos (UFSCAR)	24	83,33	61,97
Interação Humano-Computador	Universidade Federal da Bahia (UFBA)	18	94,44	84,51
Lógica Matemática e Matemática Discreta	Universidade Federal da Bahia (UFBA)	18	94,44	45,07
Modelagem de Sistemas de Informação	Universidade Federal da Bahia (UFBA)	18	55,56	26,76
Pesquisa Operacional	Universidade Federal do Estado do Rio de Janeiro (UNIRIO)	53	81,13	61,97
Probabilidade e Estatística	Universidade Federal de Uberlândia	13	69,23	23,94

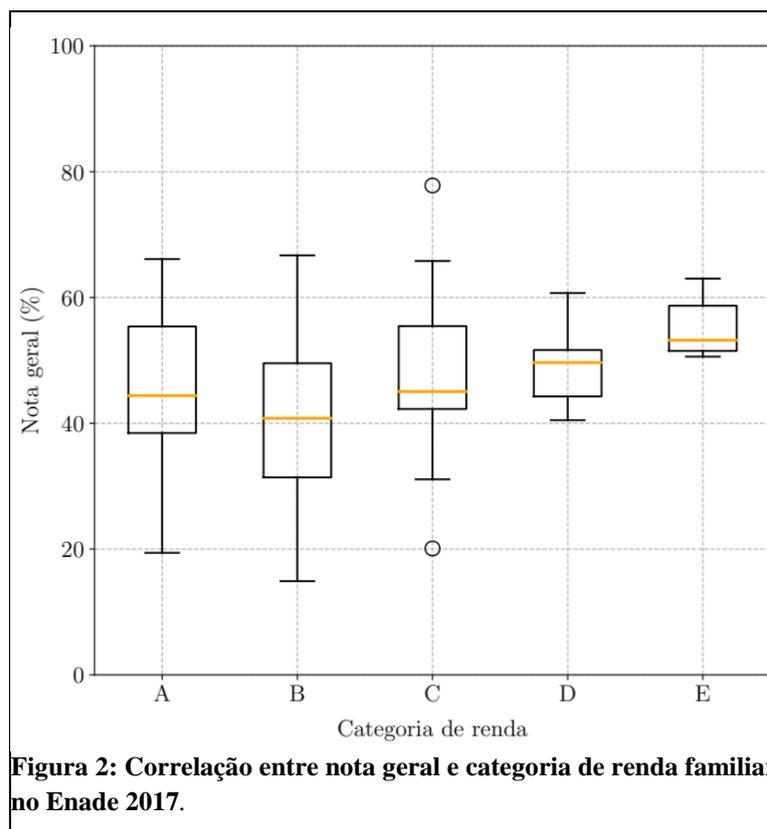
	(UFU)			
Qualidade de Processo e Produto	Universidade Federal de Mato Grosso do Sul (UFMS)	4	87,5	61,27
Redes de Computadores e Sistemas Distribuídos	Universidade Federal de Uberlândia (UFU)	13	69,23	40,85
Segurança e Auditoria de Sistemas	Universidade Federal de Ouro Preto (UFOP)	23	100	90,14
Sistemas Operacionais	Universidade Federal do Espírito Santo (UFES)	13	76,92	42,25

A partir das conclusões depreendidas do trabalho de Mizael e Lima (2018), i.e., que a situação socioeconômica e quantidade de horas de estudo dedicadas semanalmente pelos discentes apresentam alta influência em suas performances no Enade, convém-se averiguar a correlação desses fatores com o desempenho obtido pelos egressos do curso de SI da UFPA. Para o estudo do impacto do primeiro fator sobre a nota dos participantes em questão, foi gerado um gráfico *boxplot* para a análise de correlação entre a renda familiar dos discentes da UFPA e suas notas gerais na prova, ilustrado na Figura 4. Para isso, os dados dos discentes foram agrupados segundo suas categorias de renda, a qual é informada⁶ pelos participantes no Questionário do Estudante. A estruturação das categorias, consoante o dicionário de variáveis presente no arquivo dos microdados, segue a seguinte configuração:

- Categoria A: até R\$ 1.405,50
- Categoria B: de R\$ 1.405,51 a R\$ 2.811,00
- Categoria C: de R\$ 2.811,01 a R\$ 4.216,50
- Categoria D: de R\$ 4.216,51 a R\$ 5.622,00
- Categoria E: de R\$ 5.622,01 a R\$ 9.370,00

Ressalta-se que o Questionário do Estudante apresenta ainda mais duas opções de categorias de renda: F e G, sendo a primeira relativa a uma renda familiar de R\$ 9.370,01 a R\$ 28.110,00; e a segunda, a uma superior à R\$ 28.110,00. Entretanto, visto que apenas um dos estudantes da UFPA se encaixava na classificação F e nenhum na classificação G, optou-se por descartar tais categorias da análise, dado que tais espaços amostrais caracterizavam quantidades insuficientes de dados para depreensão de conclusões. Ademais, nota-se na Figura 4 que a categoria B foi a que apresentou maior variação nos valores das notas gerais, o que pode ser explicado pelo número elevado de discentes que pertencem a este grupo em relação aos demais.

⁶ Refere-se à 8ª questão do Questionário do Estudante, de enunciado: “Qual a renda total de sua família, incluindo seus rendimentos?”



No que tange à análise de correlação entre as variáveis em questão, denota-se haver certa proporcionalidade direta entre o desempenho dos discentes na prova e as categorias de renda B, C, D e E. Contudo, é provável que isso se deva às quantidades significativamente díspares de estudantes pertencentes a cada categoria, o que é exemplificado pela diminuição da variação das notas nas categorias mencionadas.

A Figura 3 apresenta uma análise de correlação entre o desempenho no exame e as categorias de quantidade aproximada de horas semanais dedicadas pelos estudantes (excetuando-se as horas de aula). As categorias de resposta à pergunta⁷ acerca da quantidade de horas semanais de estudos dos discentes, segundo o dicionário de variáveis supracitado, dividem-se da seguinte forma:

- Categoria A: “Nenhuma, apenas assisto às aulas.”
- Categoria B: “De uma a três.”
- Categoria C: “De quatro a sete.”
- Categoria D: “De oito a doze.”
- Categoria E: “Mais de doze.”

⁷ Refere-se à 23ª questão do Questionário do Estudante, de enunciado: “Quantas horas por semana, aproximadamente, você dedicou aos estudos, excetuando as horas de aula?”

A partir do gráfico gerado, ilustrado na Figura 3, denota-se que as melhores notas estão associadas ao intervalo de quatro a sete horas semanais de estudo, apesar da significativa variância do desempenho dos discentes que se enquadram nesta categoria. De modo geral, as categorias B, C, D estão relacionadas com os melhores índices de proficiência na prova, o que indica uma possível correlação direta entre a quantidade de horas de estudo extraclasse e a performance obtida no exame. É válido destacar que a quebra da tendência de proporcionalidade no gráfico denotada pelas notas associadas à categoria E pode ser consequência da baixa quantidade de discentes que a compõem, constituindo assim um espaço amostral demasiadamente diminuto para permitir uma análise de correlação mais conclusiva.

Adicionalmente, buscou-se averiguar a possível existência de correlações entre as duas variáveis analisadas anteriormente, i.e., se a renda familiar dos discentes está associada à quantidade de horas semanais de estudo dedicadas de forma extraclasse. Para a realização desta análise, as informações dos estudantes foram agrupadas de acordo com suas respostas às perguntas do questionário respectivas às variáveis em questão, de modo a organizar a distribuição das quantidades de discentes que compõem as categorias de horas semanais de estudo segundo as categorias de renda familiar. A partir dos resultados ilustrados no gráfico mostrado na Figura 4, denota-se que o intervalo de uma a três horas de estudo semanais é o intervalo modal do conjunto de dados observado, seguido pelo intervalo de quatro a sete horas. Entretanto, não se observam correlações diretas significativas entre os fatores de renda e horas semanais de estudo, de tal modo que, apesar de as análises apresentadas possivelmente indicarem que tais variáveis estão individualmente associadas ao grau de proficiência dos discentes na prova, estas não aparentam expressivamente correlacionadas.

5. Conclusões

O presente trabalho buscou apresentar o processo de análise dos microdados do Enade 2017 a partir da aplicação de algoritmos de mineração de dados, os quais geraram informações dispostas em gráficos e tabelas acerca do desempenho do curso de Sistemas de Informação da UFPA em comparação com o desempenho médio das IES do país. Nesse contexto, as métricas obtidas têm como principal papel fornecer um panorama da qualidade do curso da UFPA com base na performance alcançada nos componentes específicos presentes no exame, podendo ser utilizadas na identificação de debilidades e potencialidades do ensino pedagógico, bem como subsídio do processo de tomada de decisões sobre ações direcionadas à melhoria do status da qualidade do curso e do desempenho dos discentes em futuras edições da prova.

Considerando os pontos apresentados, depreende-se que os temas nos quais a UFPA apresentou os melhores desempenhos foram Segurança e Auditoria de Sistemas e Interação Humano-Computador, enquanto que os piores desempenhos alcançados estão relacionados aos temas de Probabilidade e Estatística e Gerência de Projetos, ambos com percentual de acerto em torno de 20%. Ademais, destaca-se Pesquisa Operacional como o tema com maior disparidade positiva de desempenho entre o curso de SI da UFPA e a média nacional, com uma diferença de mais de 15 pontos percentuais de acerto. Entretanto, apesar da boa pontuação obtida em alguns componentes específicos, a UFPA encontra-se abaixo da performance média nacional em metade dos temas abordados na prova, denotando assim um desempenho abaixo do ideal.

Outrossim, a análise das divisões do número de discentes nas categorias de renda familiar e quantidade de horas semanais de estudo extraclasse mostrou a possível existência de uma correlação individual de proporcionalidade entre essas variáveis e o desempenho obtido no exame, embora o espaço amostral do caso apresentado denotasse expressiva heterogeneidade na distribuição quantitativa de participantes nas categorias, o que pode impactar significativamente o resultado observado.

Ademais, é válido pontuar que a metodologia de análise apresentada, embora aplicada neste trabalho como um estudo de caso para o curso de Sistemas de Informação da UFPA, possui amplo potencial de replicação para outros cursos nacionais; além de apresentar espaço para a incorporação de funcionalidades adicionais de análise dos microdados, a exemplo do uso de técnicas adicionais de visualização de dados, análise de correlações entre atributos distintos dos discentes para descoberta de novos padrões associativos, e implementações para automatização de classificação das questões por tema, facilitando assim a aplicação dos algoritmos desenvolvidos aos dados de outros cursos. Dessa forma, ao proporcionar diferentes panoramas de avaliação da performance discente no exame, tais melhorias podem impactar significativamente o processo de aprimoramento de cursos superiores, contribuindo assim para a melhoria da qualidade da educação nacional.

6. Referências

Cortelazzo, A. L., Júnior, D. P., & Rodrigues, M. R. J. B. (2021). **Resultados do exame nacional do desempenho de estudantes – ENADE: quantitativos históricos na área da computação e qualidade dos cursos / Results of the national examination of student performance – ENADE: historical quantitative in the computer area and course quality.** *Brazilian Journal of Development*, 7(1), 865–886. <https://doi.org/10.34117/bjdv7n1-058>.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). **From Data Mining to Knowledge Discovery in Databases.** *AI Magazine*, 17(3), 37. <https://doi.org/10.1609/aimag.v17i3.1230>

Lima, P., Ambrósio, A., Brancher, J., & 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. <http://milanesa.ime.usp.br/rbie/index.php/wcbie/article/view/8267>

Mizael, G. R. d.; Lima, J. L. d. O (2018). **Mineração de dados para análise do desempenho do curso de Sistemas de Informação da UEG no Enade.** *Anais do Simpósio Unificado dos Cursos de Sistemas de Informação da UEG: Computação Cognitiva: Conceitos e Impactos no mundo que conhecemos.*

Neto, W. R. C. (2020). **O uso de mineração de dados educacionais sob o ENADE como apoio ao processo de tomada de decisão de gestores do ensino superior.** Trabalho de Conclusão de Curso, 57 p. <http://repositorio.ufc.br/handle/riufc/58835>.

Silva, L. C. da, Campanelli, A., Silva, L. S. da, Silva, T. V. F. da, Silva, R. C. da, Garcia, R. D. R., & Magalhães, F. L. F. de. (2020). **Graduação em Computação no Brasil: perspectiva a partir do Exame Nacional de Avaliação de Desempenho.** *Revista*

Iberoamericana De Tecnología En Educación Y Educación En Tecnología, (25), e6.
<https://doi.org/10.24215/18509959.25.e6>

Wirth, R. & Hipp, Jochen. (2000). **CRISP-DM: Towards a standard process model for data mining**. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*.