

O impacto de aspectos socioeconômicos no desempenho de estudantes de Sistemas de Informação no Enade

Title: The impact of socioeconomic aspects on the Information Systems students' performance at Enade

Caio César da Silva Rezende
Faculdade Metodista Granbery
caioxcezar@gmail.com

Luiz Antonio Borges Cantarino
Faculdade Metodista Granbery
luizantonix@gmail.com

Phillipe Ferreira de Souza
Faculdade Metodista Granbery
maxzjf2012@gmail.com

Tatiane Ornelas Martins Alves
Faculdade Metodista Granbery
tatiane.alves@granbery.metodista.br

Ricardo Silva Campos
Faculdade Metodista Granbery
ORCID: [0000-0003-3744-0838](https://orcid.org/0000-0003-3744-0838)
ricardosilvacampos@outlook.com

Resumo

O objetivo deste artigo é determinar quais aspectos socioeconômicos são os mais influentes no desempenho dos estudantes do curso de Sistemas de Informação (SI). Para isto, foram utilizados os microdados do Exame Nacional de Desempenho dos Estudantes (Enade) do ano 2017, de forma a se analisar a nota dos estudantes no exame e as respostas do questionário socioeconômico. O levantamento das variáveis mais impactantes foi realizado através do algoritmo de árvore de decisão e o coeficiente de correlação estatístico. Em seguida, as variáveis selecionadas para os estudantes de SI foram amplamente apresentadas, comparadas com as dos participantes dos demais cursos presentes no ENADE 2017 e discutidas à luz da literatura. Mostrou-se que, como esperado, as variáveis mais relevantes foram a renda familiar, a participação em intercâmbio e bolsa acadêmica, o tipo de financiamento da mensalidade e o tipo de escola do ensino médio. Por outro lado, alguns aspectos avaliados não causaram impacto na nota dos estudantes de SI, embora sejam relatados como influentes na literatura, tais quais a modalidade do curso e o fato de trabalhar simultaneamente aos estudos.

Palavras-Chave: Enade; Aspectos socioeconômicos; Sistemas de Informação.

Abstract

This work aims to determine what are the most influential socioeconomic aspects on the Information Systems students' performance. To achieve it, we used the 2017 National Exam of Students Performance (Enade) data, in order to analyze the student's grade at the exam and the answers to the socioeconomic questionnaire. The survey of the most impactful variables was implemented by the decision tree algorithm and the statistical correlation coefficient. Then, the selected variables for Information Systems students were widely presented, compared to other courses attending to Enade 2017 and discussed in the light of literature. We proved that, as expected, the most relevant aspect were family income, exchange and academic scholarship attendance, type of tuition financing and school administrative type. In the other hand, some aspects did not influence the grade of Information Systems students, although they are reported in the literature, such as course modality and working and studying simultaneously.

Keywords: Enade; Socioeconomic aspects; Information Systems.

1. Introdução

Cada vez mais percebe-se que as organizações vêm buscando definições de como se garantirem no mercado. Um dos principais meios de se destacar no mercado competitivo tem sido através da utilização de sistemas de informação, em destaque aqueles para análise de dados. Diante desse cenário se destaca a necessidade da utilização de ferramentas, que trabalhem os dados de interesse das empresas e auxiliem no processo decisório, identificação dos problemas, diagnósticos, listagem das alternativas de soluções, tomada de decisão e avaliação das decisões, entre outros. De posse dessas ferramentas, gestores podem ter uma tomada de decisão mais segura e precisa.

A importância de se trabalhar dados também é fundamental para instituições de ensino, visto que tais ferramentas podem auxiliar no melhor entendimento sobre o aprendizado de seus discentes, bem como saber sobre o desempenho destes em determinado conteúdo e comparar turmas a fim de encontrar pontos de dificuldade no processo de aprendizagem. No Brasil, o INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) está ciente da importância da análise de dados e é atualmente o maior gerador de dados referentes à educação no país. O INEP disponibiliza diversas informações sobre a educação através de microdados, que são bancos de dados cujos registros representam a unidade de coleta mais desagregada. Neste contexto, as informações usualmente estão relacionadas a uma pesquisa ou avaliação e são úteis para as instituições de ensino, pois podem ser utilizadas em processos decisórios. No entanto, existe uma grande dificuldade de se obter conhecimento útil devido à complexidade de extração e análise dos dados disponibilizados, o que dificulta uma análise manual, sendo assim se faz necessário aplicar ferramentas de sistemas de informação para análise (NAMEN; BORGES; SADALA, 2013).

Neste cenário, o presente trabalho propõe a seguinte questão de pesquisa: quais são os aspectos socioeconômicos mais influentes no desempenho acadêmico dos estudantes de Sistemas de Informação? O objetivo geral é conhecer os fatores mais importantes no rendimento estudantil neste curso. Para alcançá-lo, foram utilizados os dados fornecidos pelo INEP referentes ao Exame Nacional de Desempenho dos Estudantes (Enade) de 2017. É realizada uma análise sobre as notas do exame e as respostas do questionário aplicado aos participantes, de forma a se mapear as variáveis de maior impacto no rendimento do discente. Esta análise é realizada através de duas técnicas: um algoritmo de aprendizagem de máquina supervisionado, conhecido como Árvores de Decisão, e o coeficiente de correlação estatístico. Então, estas variáveis selecionadas são apresentadas detalhadamente, comparadas com o desempenho geral dos participantes dos demais cursos presentes no Enade 2017. As variáveis também serão discutidas à luz da literatura, para que se possa melhor compreender o contexto e a causalidade entre fator socioeconômico avaliado e a nota e traçar o perfil do estudante com melhores chance de alto desempenho. Esta é uma pesquisa descritiva em relação ao objetivo, com procedimento *ex-post-facto* de abordagem quantitativa e natureza aplicada.

O desempenho acadêmico é um fenômeno complexo e possui diversas causas, e por este motivo é necessário analisar vários aspectos de vários grupos de indivíduos, tais quais aspectos socioeconômicos, sociodemográficos, institucionais e pedagógicos. O conhecimento adquirido a partir de estudos sobre o assunto podem influenciar o estabelecimento de políticas públicas que visam a melhor qualidade na formação dos estudantes do ensino superior e conseqüentemente na vida social após a conclusão do curso (BARAHONA, 2014; VARGAS, 2014; VALERA et al., 2009). Para as instituições de ensino superior (IES), o conhecimento extraído por este trabalho é de grande interesse, pois pode fomentar os processos decisórios e aprimorar discussões acerca de planos pedagógicos, currículos, melhorias no processo de aprendizagem, programas de permanência, entre outros.

O presente trabalho é desenvolvido da seguinte maneira: Na Seção 2 são apresentados os conceitos necessários para o entendimento deste trabalho e os trabalhos relacionados. Na Seção 3 apresentados os métodos aplicados para obtenção das informações e na Seção 4 estão relatados os resultados obtidos pela presente pesquisa. Finalmente, na seção 5, é apresentada a conclusão.

2. Fundamentação Teórica

Nesta seção serão apresentados o Enade, as técnicas de aprendizagem de máquina e estatística utilizadas para o levantamento das variáveis mais influentes no desempenho dos estudantes no Enade, além dos trabalhos relacionados.

2.1. Enade

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) é um órgão do Ministério da Educação (MEC) responsável pelas políticas de avaliação da educação básica e superior. Entre suas atribuições, o INEP realiza o Enade (Exame Nacional de Desempenho dos Estudantes), com o objetivo de aferir o desempenho dos alunos e alunas em conteúdos programáticos descritos nas diretrizes curriculares dos respectivos cursos de graduação. Além disto, são avaliados os conhecimentos gerais ligados à realidade brasileira e mundial e a outras áreas do conhecimento (BRASIL, 2004).

Segundo Griboski (2012), os docentes e os gestores podem utilizar os resultados do Enade através de dados extraídos por instituição de educação superior e disponibilizados pelas categorias administrativa, organização acadêmica, município, estado e região, de forma a aplicá-los na promoção de ações de melhorias na qualidade dos cursos de graduação, com o objetivo de beneficiar toda a comunidade acadêmica. O exame foi implantado em 2004, e tem sofrido muitas alterações para ampliar a participação discente. Inicialmente, a seleção de estudantes era em caráter amostral e em 2009 passou a ser censitária, ou seja, todos os alunos concluintes passaram a ser convocados para a prova.

Ainda segundo Griboski (2012), o Conceito Enade é atribuído a cada curso de uma instituição de ensino superior, situado em um mesmo município. Este conceito é calculado pela soma de vários itens que compõem a avaliação, como visitas locais aos campi dos cursos avaliados, titulação do corpo docente, produção acadêmica, a qualidade da infraestrutura, incluindo tecnologias educacionais, laboratórios, bibliotecas, entre outros. No entanto, o principal aspecto considerado na aplicação do Enade para determinar a qualidade de um curso é a participação discente no processo avaliativo. Além da prova de conhecimentos gerais e específicos, é aplicado o questionário do estudante, que coleta diversas informações socioculturais e econômicas e o questionário de percepção sobre a prova. A coordenação de curso também deve responder a um questionário sobre a gestão do curso. Desta forma, as informações resultantes desta coleta são de grande valia, pois refletem a opinião e fatores socioeconômico-culturais de estudantes e coordenadores de curso de graduação de todo o Brasil.

O Enade 2017 avaliou os estudantes de bacharelados em: Arquitetura e Urbanismo, Engenharia Ambiental, Engenharia Civil, Engenharia de Alimentos, Engenharia de Computação, Engenharia de Controle e Automação, Engenharia de Produção, Engenharia Elétrica, Engenharia Florestal, Engenharia Mecânica, Engenharia Química, Engenharia e Sistemas de Informação; os estudantes de bacharelado e licenciatura em: Ciência da Computação, Ciências Biológicas, Ciências Sociais, Filosofia, Física, Geografia, História, Letras - Português, Matemática e Química, os estudantes de licenciatura em: Artes Visuais, Educação Física, Letras - Português e Espanhol, Letras - Português e Inglês, Letras - Inglês, Música e Pedagogia; e dos estudantes de cursos tecnológicos em: Análise e Desenvolvimento de Sistemas, Gestão da Produção Industrial, Gestão da Tecnologia da Informação e Redes de Computadores.

2.2. Correlação

A correlação expressa uma medida de associação linear entre duas variáveis, refletindo o grau e a direção de relacionamento entre elas. Exemplificando, um modelo linear com duas variáveis X e Y fortemente correlacionadas significa que o incremento em X gera o mesmo impacto em Y. Existem várias formas de se medir a correlação. Uma delas é chamada de Correlação de Pearson, dado pela equação (FIGUEIREDO FILHO, 2009):

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

O sinal do coeficiente indica a direção do relacionamento e o seu valor indica a força. Os valores ficam entre -1 e 1, sendo que se módulo igual a 1 indica a correlação perfeita. Por outro lado, a relação igual a 0 indica que não há relação linear entre as variáveis. Este intervalo pode ser discretizado de forma a indicar qualitativamente a força da correlação. Para Cohen (2013), valores entre 0,10 e 0,29 indicam fraca correlação. Valores entre 0,30 e 0,49 são considerados correlação intermediária, e valores acima 0,5 indicam correlações fortes. Uma vantagem da correlação é que o seu coeficiente é adimensional, ao contrário da covariância. A covariância também mede a relação linear entre duas variáveis, porém por ser dimensional ela depende das unidades das variáveis.

A correlação pode ser usada em diferentes aplicações de mineração de dados (ALQALLAF et al, 2002). Por exemplo, o trabalho de Lee *et al.* (2000) propôs uma ferramenta para sistemas de informações médicas, de modo a identificar pacientes com alto risco a partir de informações clínicas, tais quais como idade, sexo, frequência cardíaca, pressão arterial, entre outros. A partir destes dados cadastrados no sistema, foi feita uma análise da correlação entre eles e diversas doenças cardíacas. Desta forma, foi possível analisar quais os fatores mais impactantes para determinadas cardiopatias.

2.3. Árvore de decisão

De acordo com Sharma e Kumar (2016), as árvores de decisão são uma forma de aprendizagem de máquina, bastante usada para problemas de classificação. Para que se configure uma árvore de decisão, deve-se informar um conjunto de variáveis de entrada com os valores dos atributos e uma variável de saída que contenha as classificações para cada tupla de entrada. A partir destes dados, o algoritmo aplica uma estratégia gulosa de dividir para conquistar. Seleciona-se o atributo mais importante, que tem maior impacto na classificação, em primeiro lugar, de forma a dividir os registros para que a árvore tenha a menor altura possível. Este atributo torna-se um nó de decisão e subdivide o conjunto de dados em outros menores. Este processo é repetido até que se obtenha a árvore completa. Os nós internos são testes de um valor de entrada, as ramificações são valores possíveis do atributo e as folhas contém os valores das variáveis de saída, ou seja, a classificação dada a uma entrada. Devido à estratégia de manter a árvore com menor altura possível, as variáveis mais relevantes para a classificação tendem a ficar mais próximas da raiz da árvore.

Conforme Bhargava et al (2013), as árvores de decisão são uma técnica muito popular em mineração de dados por serem de fácil compreensão pelos usuários finais, por lidarem com variáveis categóricas ou nominais, por ter alto desempenho durante a aprendizagem, entre outros.

Uma aplicação deste algoritmo no contexto de mineração de dados foi proposta no trabalho de Rolim et al (2017), em que foi apresentada uma ferramenta para auxiliar a comunidade acadêmica na participação de fóruns em ambientes virtuais de aprendizagem, que automaticamente classifica uma postagem do aluno para tirar dúvidas e então recomenda algum conteúdo para estudo.

2.4. Trabalhos relacionados

Esta seção visa apresentar alguns trabalhos relacionados ao trabalho proposto neste artigo.

O trabalho de Alves (2018) propôs um modelo preditivo do indicador de desempenho de notas na prova de redação do Exame Nacional do Ensino Médio. Foram feitos diferentes experimentos de tratamento dos dados, de forma a resultar nas melhores percentuais de previsões corretas. Foram utilizados algoritmos do software Weka¹, que foram capazes de prever corretamente 61% das amostras da base de dados do ENEM de 2016 (ALVES, 2018).

O trabalho de Mizael e Lima (2018) teve como objetivo realizar uma análise sobre o desempenho no Enade dos estudantes do curso de Sistemas de Informação (SI) da Universidade Estadual de Goiás (UEG). Este trabalho realizou uma análise nos dados do Enade 2014 com técnicas de mineração de dados, porém a análise foi focada no corpo discente da UEG. Os autores concluíram que a situação socioeconômica e a quantidade de horas de estudo impactaram positivamente na nota do exame.

O trabalho de Lima *et al.* (2021) utilizou dados educacionais do Enade para analisar o conteúdo da prova, de forma a demonstrar quais disciplinas interferem mais no resultado do exame. Foram utilizados apenas os dados dos corpos discentes dos cursos de Ciência da Computação, nos exames realizados nos ciclos de avaliação dos anos de 2005, 2008, 2011, 2014 e 2017. As questões da prova foram classificadas conforme os conteúdos pragmáticos das disciplinas da graduação avaliada. Foram aplicadas técnicas estatísticas que permitem a avaliação dos estudantes por temáticas específicas dentro do curso.

O trabalho de Lemos e Miranda (2015) utilizou os dados do Enade para avaliar quais variáveis influenciam o desempenho dos estudantes de Ciências Contábeis. Foram utilizados dados dos exames de 2009 e 2012, em que foram comparados dados entre estudantes das instituições que obtiveram os piores conceitos (1 e 2) e dos estudantes matriculados as instituições com os melhores conceitos (4 e 5). O trabalho revelou que as instituições com melhores desempenhos possuem melhor qualificação docente (titulação e formação pedagógica), melhor infraestrutura e maior quantidade de professores com dedicação exclusiva.

3. Métodos

Foram utilizados dados do Enade do ano 2017 disponibilizados pelo INEP². Os dados estão salvos em formato de arquivo de texto com a extensão (.txt), com valores separados por vírgula. Os dicionários são disponibilizados junto ao arquivo de dados, e descrevem detalhadamente as colunas do arquivo, que refletem as perguntas e alternativas do questionário socioeconômico, percepção da prova, dados pessoais e notas.

Para a realização do presente trabalho, várias etapas foram necessárias. Estas etapas seguem o processo de descoberta de conhecimento em bases de dados de Fayyad et al. (1996) e estão

¹ <https://www.cs.waikato.ac.nz/ml/weka/>

² Disponível em <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enade> (acessado em 25/6/2021)

ilustradas na Figura 1. Inicialmente, foram realizadas pesquisas, levantamentos, análises e estudos sobre os dados obtidos através do INEP. Em seguida, algumas variáveis do *dataset* do INEP foram escolhidas para fazer parte deste trabalho conforme a sua relevância. Depois disso, foram feitos diversos tratamentos nos dados, como a exclusão de dados faltantes e transformações de variáveis categóricas ordinais e nominais. Então, foram aplicados os algoritmos de árvore de decisão e correlação estatística para o levantamento das variáveis mais impactantes no rendimento. Por fim, os resultados obtidos foram analisados e relatados nesse artigo. Todas as etapas serão detalhadas a seguir.

Fonte: Adaptada de Fayyad et al. (1996)

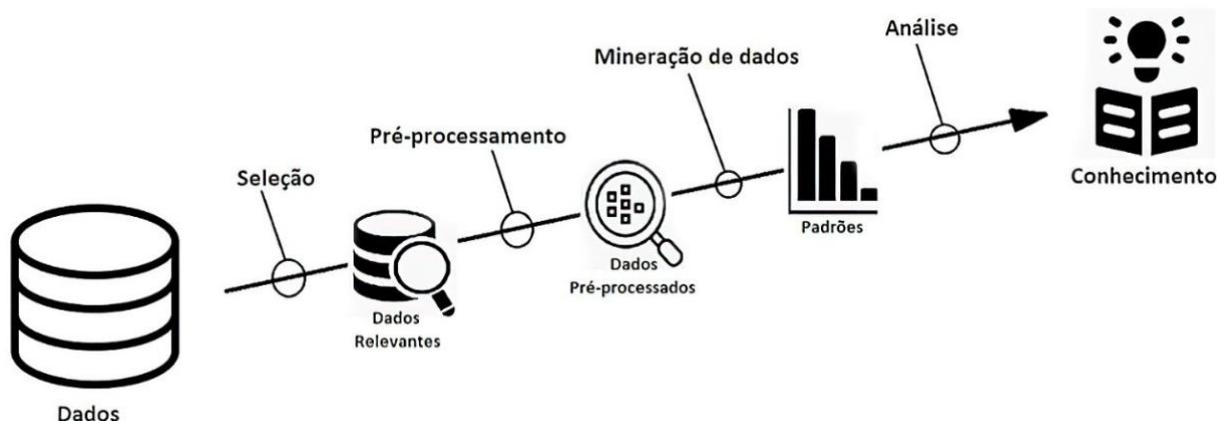


Figura 1: Descrição das etapas da metodologia.

3.1. Seleção e pré-processamento de dados

Inicialmente, foram determinados quais dados seriam utilizados no processo de descoberta de conhecimento. Isso inclui descobrir quais os dados estão acessíveis, obter dados importantes e depois integrá-los para a descoberta de conhecimento em um conjunto, de forma a permitir a sua utilização como base para a criação dos modelos preditivos. Se alguns atributos significativos estiverem faltando, os resultados do estudo podem ser comprometidos. Pode-se tentar adicionar mais atributos ou tentar tratar os dados faltantes, a fim de melhorar a qualidade do banco e, por conseguinte, obter resultados mais confiáveis e gerar conhecimentos mais sólidos. Este processo se inicia com o maior conjunto de dados disponíveis e depois expande e observa o impacto em termos de descoberta e modelagem de conhecimento (RIBEIRO, 2011). Desta forma, foram definidos os critérios de inclusão e exclusão dos dados utilizados nas próximas etapas, representados na Tabela 1.

Tabela 1: Critérios de inclusão e de exclusão.

Critérios de inclusão	Critérios de exclusão
<ul style="list-style-type: none"> • Provas em que o curso de sistema de informação esteve presente (2017); • Alunos que responderam a todas as questões comuns do questionário QE_I01 a QE_I68. 	<ul style="list-style-type: none"> • Código da área de enquadramento do curso no Enade diferentes de 4006 (Sistemas de informação); • Alunos que não possuíam notas finais do exame.

A base de dados foi reduzida, para que fossem avaliadas apenas as informações relevantes ao escopo deste trabalho, que são o código de curso para filtrar apenas os cursos de Sistemas de Informação, dados pessoais do estudante, dados sobre o curso, nota da prova e o questionário do estudante. Foram descartadas, portanto, as variáveis com as respostas das questões da prova, as notas parciais em cada prova e o questionário de percepção da prova. A seguir encontra-se a lista das variáveis utilizadas:

1. CO_IES: Código da IES (e-Mec).
2. CO_GRUPO: Código da área de enquadramento do curso no Enade.
3. CO_CATEGAD: Código da categoria administrativa da IES. Foi simplificada para dois valores apenas: pública ou privada.
4. CO_MODALIDADE: Código da modalidade de Ensino: presencial ou EAD.
5. CO_TURNO_GRADUACAO: Turno do curso.
6. GAP: variável calculada pela diferença entre ANO_IN_GRAD (Ano de ingresso na graduação) e ANO_FIM_EM (Ano de conclusão do ensino médio).
7. NU_IDADE: idade do participante no dia do exame.
8. TP_SEXO: sexo.
9. NT_GER: Nota bruta na prova.
10. QE_I01 a QE_I26: Questionário do estudante.
11. As variáveis QE_I27 a QE_I68 foram marcadas conforme o grau de concordância do aluno para cada assertiva, segundo a escala que varia de 1 (discordância total) a 6 (concordância total). A resposta 7 foi marcada quando julgado "não ter elementos para avaliar a assertiva" e resposta 8 para quando "considerar não pertinente ao curso".

As questões QE_I69 a QE_I81 do questionário do estudante são itens exclusivos para os estudantes das licenciaturas e, portanto, foram excluídas desta análise.

A seguir foi realizado o pré-processamento. Nessa fase é realizada a limpeza de dados, como a manipulação de dados ausentes e remoção de ruídos ou *outliers*. Pode incluir técnicas estatísticas complexas ou usar um algoritmo de mineração de dados nesse contexto (GOLDSCHMIDT; PASSOS, 2005). Também podem ser aplicadas outras técnicas de transformação nos dados tais como: normalização, agregação, criação de novos atributos, redução e sintetização dos dados. Isto é feito para que se garanta a qualidade da base de dados, e para que as técnicas computacionais possam ser aplicadas (ROLIM et al. 2017).

Foram excluídos os registros que possuíam uma ou mais variáveis em branco, dentre aquelas variáveis selecionadas conforme o escopo deste trabalho.

Os dicionários disponibilizados pelo INEP foram utilizados para a seleção das variáveis de interesse deste trabalho. Eles também foram usados para classificar corretamente todas as variáveis entre quantitativas, nominais ou ordinais, para que elas recebam o tratamento adequado para bom funcionamento dos algoritmos.

As variáveis numéricas não foram alteradas, uma vez que os métodos utilizados não são sensíveis a dados não normalizados. São elas: NU_IDADE, GAP, e NT_GER.

As variáveis ordinais são QE_I04, QE_I05, QE_I07, QE_I08, QE_I09, QE_I10, QE_I22, QE_I23. As suas respostas em formato de letras foram substituídas por números inteiros, de forma a manter a correspondência de grandezas.

As demais variáveis foram tratadas como nominais. Neste caso, foram acrescentadas colunas com valores binários para cada alternativa possível, ou seja, os valores representam o pertencimento ou não pertencimento a cada categoria, o que é conhecido como variável *dummy*. Inicialmente, o *dataset* contava com 75 colunas. Após este tratamento, passou a ter 479 colunas a mais, além das 75 originais.

3.2. Levantamento das variáveis mais influentes na nota

O objetivo deste trabalho é levantar as variáveis mais influentes na nota dos estudantes no Enade. Para tal, foram utilizados um algoritmo de aprendizagem de máquina e outro estatístico, a fim de

identificar tendências e informações úteis para tomadas de decisões a partir conjunto de dados selecionado acima (PASTA, 2011).

Para gerar a análise dos dados, foram utilizadas as bibliotecas Pandas e SciKit-Learn da linguagem Python. Inicialmente, foi analisado o banco de dados do Enade 2017, com registros de 11892 estudantes do curso de Sistemas de Informação.

A árvore de classificação recebe como entrada as respostas do questionário e a partir delas prevê uma nota categórica, discretizada pelos intervalos determinados pelos quartis: Nota A para valores menores que 35,5, nota B entre 35,5 e 44,6, nota C entre 44,6 e 53,8 e nota D acima de 53,8. A entrada foi separada entre teste e treino na proporção de 20 e 80%, respectivamente. Desta forma, o modelo de classificação obteve cerca 30% de acerto ao classificar o *dataset* de teste.

Desta forma, foi utilizada uma outra técnica para o levantamento das variáveis mais importantes para a nota, chamada de coeficiente de correlação. Foram selecionados os 10 maiores índices, que serão exibidos na próxima seção.

4. Resultados

A árvore de classificação e a correlação foram calculadas a partir do banco de dados selecionado, que contém 11892 registros com 75 variáveis. A partir da metodologia proposta, foi feito o levantamento das variáveis que são mais influentes na nota. Em seguida, estas variáveis foram analisadas estatisticamente, de forma a comparar o desempenho dos grupos de estudantes que responderam a cada alternativa de determinada questão. Por último, algumas variáveis de interesse público, que não constaram na lista de maiores correlações, foram avaliadas.

4.1. Quais são as variáveis mais influentes?

A árvore de classificação obtida pela análise do conjunto de dados dos cursos de Sistemas de Informação encontra-se na Figura 2, cuja altura foi limitada a 3 níveis para a melhor visualização. A raiz da árvore é a coluna QE_I11, que indica se o discente recebeu alguma bolsa ou financiamento pelos estudos. Devido à forma que as variáveis categóricas nominais foram tratadas, nota-se que cada opção da questão foi transformada em uma coluna, com valores binários/booleanos. Em sequência, pode-se observar a variável QE_I08, o que indica a faixa de renda total da família. No nível mais baixo, constam ainda as questões QE_I16, que relata qual a unidade federativa em que o discente concluiu o Ensino Médio, CO_MODALIDADE que indica curso presencial ou à distância e QE_I14, representa a participação de atividades curriculares no exterior.

Todavia, a acurácia na classificação gira em torno de 30%. Assim, fizeram-se necessários estudos adicionais para que fosse possível avaliar melhor o impacto do emprego e renda no desempenho estudantil.

Fonte: Elaboração própria.

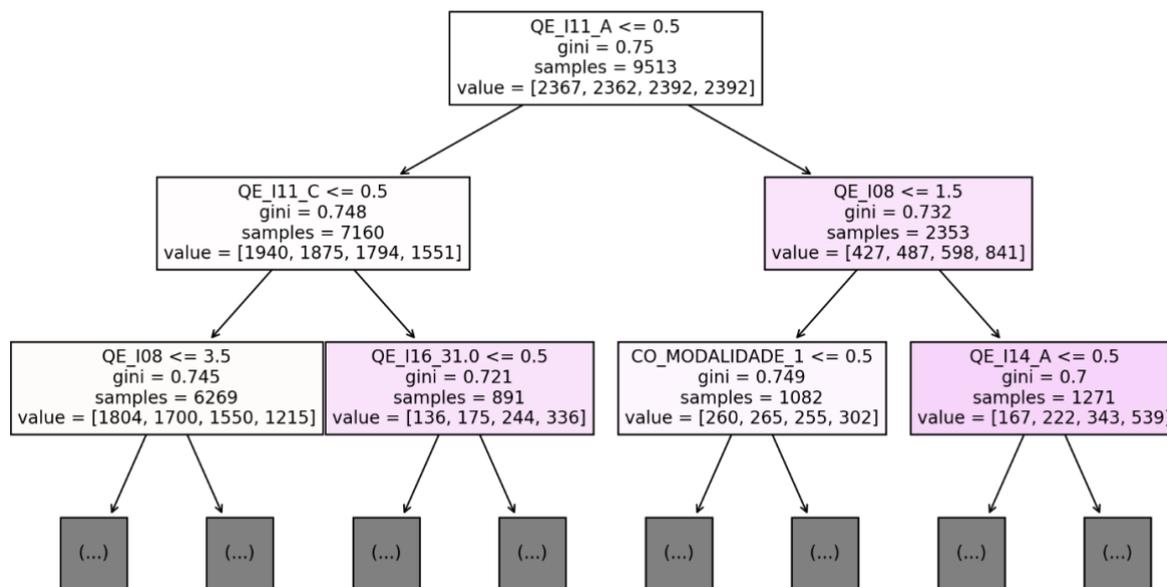


Figura 2:Árvore de decisão.

Desta forma, foi traçado um mapa de correlação entre as variáveis do conjunto de dados. Como são muitas variáveis, foram selecionadas as 10 questões com os maiores valores absolutos de correlação. O resultado pode ser visto na Figura 3.

Nota-se, na última linha da figura, que as variáveis de entrada têm baixa correlação com a nota do estudante. Isto indica que a nota é pouco influenciada por tais dados. Percebe-se que os maiores coeficientes de correlação são as variáveis QE_I08 (renda familiar) e QE_I11 (bolsa e financiamento), o que corrobora as informações obtidas pela árvore de decisão. Pode-se afirmar que uma parte das variáveis aparece em ambos os algoritmos empregados: QE_I08, QE_I11, QE_14. As questões QE_I16 e CO_MODALIDADE aparecem apenas na árvore, enquanto TP_IES, QE_13, QE_23, QE_17 e QE_23 aparecem apenas na matriz de correlação. Todas estas variáveis serão apresentadas e discutidas a seguir.

Fonte: Elaboração própria.

	QE_I17_B	QE_I26_F	QE_I13_B	QE_I11_C	QE_I14_B	QE_I23	QE_I13_A	QE_I11_E	TP_IES_Pub	QE_I11_A	QE_I08	NT_GER
QE_I17_B	1.00	0.06	0.04	-0.15	0.07	-0.04	-0.04	-0.09	0.05	0.05	0.29	0.10
QE_I26_F	0.06	1.00	0.04	-0.02	0.04	0.08	-0.06	0.01	-0.02	-0.02	0.10	0.10
QE_I13_B	0.04	0.04	1.00	-0.05	0.08	0.08	-0.47	-0.06	0.23	0.23	0.02	0.11
QE_I11_C	-0.15	-0.02	-0.05	1.00	0.00	0.02	0.07	-0.15	-0.19	-0.18	-0.13	0.11
QE_I14_B	0.07	0.04	0.08	0.00	1.00	0.01	-0.13	-0.06	0.14	0.14	0.07	0.11
QE_I23	-0.04	0.08	0.08	0.02	0.01	1.00	-0.10	-0.02	0.09	0.08	-0.04	0.11
QE_I13_A	-0.04	-0.06	-0.47	0.07	-0.13	-0.10	1.00	0.13	-0.33	-0.31	0.01	-0.12
QE_I11_E	-0.09	0.01	-0.06	-0.15	-0.06	-0.02	0.13	1.00	-0.27	-0.27	-0.17	-0.13
TP_IES_Pub	0.05	-0.02	0.23	-0.19	0.14	0.09	-0.33	-0.27	1.00	0.94	-0.05	0.14
QE_I11_A	0.05	-0.02	0.23	-0.18	0.14	0.08	-0.31	-0.27	0.94	1.00	-0.05	0.15
QE_I08	0.29	0.10	0.02	-0.13	0.07	-0.04	0.01	-0.17	-0.05	-0.05	1.00	0.15
NT_GER	0.10	0.10	0.11	0.11	0.11	0.11	-0.12	-0.13	0.14	0.15	0.15	1.00

Figura 3: Correlação entre a nota e variáveis socioeconômicas.

4.2. Análise das variáveis mais influentes

Após o levantamento das perguntas mais relevantes no questionário socioeconômico dos estudantes de Sistemas de Informação, é feita uma análise sobre cada uma das variáveis, em que se apresenta uma tabela com o total de estudantes em cada grupo, o percentual, a nota média com o desvio-padrão e a correlação daquela variável com a nota.

Para fins de comparação das variáveis mais importantes para o curso de Sistemas de Informação, são exibidos também os dados de todos os participantes do Enade 2017, que não possuíam valores em branco. Assim, são apresentados dados de 442496 estudantes. Todas as variáveis são discutidas à luz da literatura e serão apresentadas na ordem em que aparecem no questionário.

A primeira variável trata sobre tipo de instituição de ensino superior (IES), que pode ser pública ou privada. A Tabela 2 mostra que o público geral se divide entre 64,3% em instituições privadas e 35,7% em públicas. O respectivo desempenho em média é 41,9 e 46,8. Por outro lado, entre os alunos de Sistemas de Informação, 73,9% estão nas instituições privadas e apenas 26,1% estão na rede pública. Também é possível perceber que a nota média das IES públicas é superior, cerca de 47,9 contra 43,6. Esta tendência pode ser observada em cursos de Sistemas de Informação do Enade 2014 (MIZAEL; LIMA, 2018) e em outras graduações, como em Ciências Contábeis no Enade 2012 (FERREIRA, 2015). De acordo com Bittencourt (2008), um dos fatores que pode explicar esta diferença está no processo seletivo. Por terem maior concorrência, as instituições públicas selecionam estudantes com desempenhos mais altos.

Tabela 2: Tipo de instituição de ensino superior.

TP_IES	Todos os cursos				Sistemas de Informação			
	n	%	Nota	Corr.	N	%	Nota	Corr.
Privada	284620	64,3	41,9±13,4	-0,17	8788	73,9	43,6±12,7	-0,14
Pública	157876	35,7	46,8±14,7	0,17	3104	26,1	47,9±13,8	0,14
TOTAL	442496	100	43,6±14,1	-	11892	100	44,7±13,2	-

Fonte: Elaboração própria com dados do INEP-Enade 2017.

A seguir, é analisada a modalidade do curso, que pode ser presencial ou à distância (EAD), conforme consta na Tabela 3. No quadro geral, os estudantes de EAD são 20,1% do total e os estudantes do presencial, 79,9%. A nota média do EAD (40,7) é menor que o presencial (44,4). Desta forma, ao avaliar todos os cursos percebe-se uma tendência relatada na literatura, de que o desempenho do EAD é ligeiramente menor se comparado ao presencial (CAETANO et al., 2016).

Todavia, ao se avaliar os estudantes de Sistemas de Informação separadamente, percebe-se que o desempenho é praticamente igual entre as modalidades. Os cursos EAD possuem apenas 6,4% do total de estudantes do curso, que atingiram nota média 45,0. Os alunos presenciais são 93,6%, que por sua vez obtiveram média muito semelhante aos estudantes EAD, cerca de 44,7, possuindo correlação próxima a 0. A diferença na nota média de 0,3 pontos é muito pequena, principalmente se os desvios-padrões forem considerados, que estão em torno de 13 pontos. Além disto, a proporção de estudantes EAD de SI é menor se comparada aos demais cursos, o que também pode interferir nesta comparação. De qualquer forma, para uma comparação mais justa e adequada entre as duas modalidades seria necessário considerar o capital cultural e outros fatores socioeconômicos dos estudantes em uma pesquisa qualitativa (SCUDELER, CALDERON, PIRES, 2020).

Tabela 3: Modalidade do curso.

MODALID.	Todos os cursos				Sistemas de Informação			
	n	%	Nota	Corr.	N	%	Nota	Corr.
EAD	88821	20,1	40,7±14,2	-0,10	757	6,4	45,0±13,9	0,00
Presencial	353675	79,9	44,4±13,9	0,10	11135	93,6	44,7±13,1	0,00
TOTAL	442496	100	43,6±14,1	-	11892	100	44,7±13,2	-

Fonte: Elaboração própria com dados do INEP-Enade 2017.

Em sequência, avalia-se uma das variáveis mais importantes e influentes: a questão QE_I08, que trata sobre a renda total da família do discente. Na Tabela 4, percebe-se uma tendência de crescimento das notas médias conforme a renda aumenta. Todavia, devido à alta dispersão dos dados notada através do desvio padrão, pode-se dizer que existe uma relação entre renda e nota, porém em um grau baixo. O quadro geral mostra que a nota média aumenta à medida que a renda familiar também aumenta. A diferença entre o grupo mais vulnerável e menor vulnerável é 10,8 pontos, uma das maiores discrepâncias notadas neste trabalho. Também se percebe que os estudantes estão concentrados nas faixas de renda entre 1,5 e 4,5 salários-mínimos.

Ao se avaliar os estudantes de Sistemas de Informação, nota-se que as pessoas em situação econômica mais vulnerável, cuja renda familiar total é de até 1,5 salários-mínimos, representam 14,5% dos estudantes. Estes obtiveram nota média de 41,5. Na faixa de 1,5 a 3 salários-mínimos estão 27,5% dos participantes, que ficaram com nota média 43,7. Outros 24,3% possuem renda familiar de 3 a 4,5 salários-mínimos, grupo cuja nota média foi de 44,9. Entre 4,5 e 6 salários-mínimos, estão 13,3% dos estudantes, com nota média 45,8. Cerca de 13,5% possuem de 6 a 10 salários-mínimos, com nota média 46,7. Nas faixas de renda mais elevadas, encontram-se as faixas de 10 a 30 salários e acima de 30 salários, em que, respectivamente, se encontram 6,4% e 0,5% dos participantes, com notas 49,4 e 51,7. Ou seja, a diferença entre o grupo mais vulnerável e o mais abastado é de 10 pontos, o que demonstra a mesma tendência observada para os estudantes dos outros cursos.

A influência da renda no desempenho discente tem sido amplamente debatida na literatura. Em geral, ela influencia a nota positivamente, o que significa que quanto maior a renda, maior o desempenho (SCUDELER, CALDERON, PIRES, 2020). Uma explicação para isto é que renda familiar elevada permite ao estudante frequentar as melhores escolas particulares no ensino básico e por conseguinte, ter melhores condições de ingresso em instituições públicas, que geralmente são mais concorridas (NOGUEIRA; TSUNODA, 2015). Além disto, os alunos com melhores condições financeiras têm mais acesso às atividades extraclasse, tais quais aulas de informática, idiomas, artes, entre outros, o que torna a sua base acadêmica mais consistente (FERREIRA, 2015). Também pode-se destacar que as rendas maiores proporcionam acesso a mais materiais de estudo, como livros e acesso à Internet, melhor forma de deslocamento entre casa e faculdade e não necessidade de trabalhar para se sustentar (LOPES et al, 2020).

A próxima questão avaliada trata sobre bolsas de estudos ou financiamentos que o estudante recebeu para custear mensalidades (QE_I11), visualizada na Tabela 5. A raiz da árvore de decisão é a alternativa A desta questão, e ainda consta um nó no segundo nível com a alternativa C. Além disto, constam entre as variáveis com maior correlação com a nota as alternativas A, C e E desta mesma questão. Para os alunos do curso de Sistemas de Informação, a alternativa A (Nenhum, pois meu curso é gratuito) abrange 24,4% dos estudantes e tem a segunda melhor média dentre as opções desta questão, com 48,1 pontos. Por motivos óbvios, esta é a alternativa mais escolhida dentre aqueles matriculados nas IES públicas, o que é comprovado pela alta correlação (0,94)

entre esta questão e o tipo de IES, conforme demonstrado na Figura 3. Além disto, por ser a alternativa de escolha para os matriculados em IES públicas, era de se esperar que a nota média seria mais alta, como já demonstrado na Tabela 2.

Tabela 4: Renda total familiar do estudante em termos de salários-mínimos.

QE_I08	Todos os cursos				Sistemas de Informação			
	n	%	Nota	Corr.	N	%	Nota	Corr.
< 1,5	99804	22,6	40,0±13,5	0,21	1719	14,5	41,5±12,5	0,15
De 1,5 a 3	128283	29,0	42,3±13,5	0,21	3275	27,5	43,7±12,6	0,15
De 3 a 4,5	90693	20,5	44,2±13,7	0,21	2895	24,3	44,9±12,7	0,15
De 4,5 a 6	46481	10,5	45,6±13,9	0,21	1582	13,3	45,8±13,3	0,15
De 6 a 10	46816	10,6	47,5±14,2	0,21	1602	13,5	46,7±13,8	0,15
De 10 a 30	26594	6,0	50,2±14,6	0,21	765	6,4	49,4±14,3	0,15
> 30	3825	0,9	50,8±14,9	0,21	54	0,5	51,7±15,4	0,15
TOTAL	442496	100	43,6±14,1	-	11892	100	44,7±13,2	-

Fonte: Elaboração própria com dados do INEP-Enade 2017.

O maior grupo de estudantes contém 28% do total, que representa aqueles que não receberam bolsa, embora o curso não seja gratuito. Para estes, a nota média é de $43,0 \pm 13,0$. Outro destaque na correlação é a alternativa C, que trata dos alunos que receberam bolsa do Programa Universidade para Todos (ProUni) de maneira integral. Eles são 9,3% do total, e obtiveram a maior nota média, cerca de 49,3. Os bolsistas parciais do ProUni representam 2,4% do total, com nota média de 45,4.

Também em destaque na correlação, está a alternativa E, para os bolsistas do Fundo de Financiamento Estudantil (FIES), que são 18,4% dos participantes, com a segunda menor nota média, de 41,3 pontos. Como a nota é menor para este grupo, pode-se perceber que a correlação é negativa, -0,13. O próximo grupo é composto por aqueles com bolsa parcial ProUni e FIES, com 1,9% do total e média 45,6.

Por último, constam grupo de menor representatividade, como bolsas oferecidas por governos estaduais, distritais ou municipais, bolsas oferecidas pela própria instituição, bolsas oferecidas por outras entidades, financiamentos oferecidos pela própria instituição e financiamentos bancários.

Portanto, pode-se afirmar que dentre as modalidades de custeio e financiamento de mensalidades, o melhor desempenho foi obtido entre os estudantes financiados pelo ProUni integralmente, com nota média 49,3. Esta média, inclusive, foi superior aos alunos que não recebem financiamento pois estão matriculados em cursos gratuitos, que por sua vez obtiveram a segunda maior média de 48,1 pontos. Para os alunos de cursos não gratuitos que não recebem financiamento, a média foi de 43 pontos.

Ao se avaliar os estudantes de todos os cursos, nota-se o mesmo comportamento: a maior média foi obtida pelos financiados pelo ProUni (48,5), seguidos pelos não financiados porque estão matriculados em cursos gratuitos (47,0). O FIES (40,2) também obteve desempenho inferior aos não financiados matriculados em cursos não gratuitos (41,4).

O trabalho de Mello Neto et al. (2020) apontou que realmente existe uma relação positiva entre a presença de aluno do ProUni e o conceito do ENADE, ao avaliar os estudantes do curso de Direito. Ou seja, além deste capital simbólico no desempenho do exame, ainda existe o retorno de capital econômico, devido à estrutura de financiamento da política governamental. Além disto, ao se aprofundar a análise do perfil do estudante do ProUni, percebeu-se que este grupo “contradiz uma esperada correlação entre baixo desempenho escolar e pertencimento aos estratos econômicos não dominantes” (MELLO NETO et al, 2020), pois apesar destes alunos possuírem renda mais baixa e pais com menor escolaridade, eles possuem melhor desempenho no Enade que os demais estudantes da rede privada, o que se explica devido à maior concorrência para obtenção deste financiamento do que pelo simples ingresso sem auxílio algum.

No trabalho de Bezerra e Tassigny (2018) percebeu-se que há maiores possibilidades de bom desempenho no Enade para os alunos do ProUni do que os das demais modalidades de financiamento, ao analisar os estudantes do curso de Administração participantes do Enade 2015. Além disto, o mesmo trabalho aponta que os alunos do FIES possuem desempenho melhor que a média, porém o impacto positivo é menor do que o causado pelo ProUni. Costa et al. (2020) mostraram que O FIES abrange principalmente estudantes que cursaram o ensino médio em escola pública, e que este financiamento causou impacto positivo na nota do Enade 2016 dos estudantes das IES privadas.

Todavia, o trabalho atual aferiu que o FIES teve desempenho inferior para todos os cursos participantes do Enade 2017 e para o curso de Sistemas de Informação. Desta maneira, é necessária uma investigação mais ampla sobre o impacto desta forma de financiamento, em que se considere qualitativamente as mudanças no regulamento do FIES ao longo do tempo, assim como outras questões sobre o contexto dos estudantes.

Tabela 5: Bolsas ou financiamento para custeio de mensalidades.

QE_I11	Todos os cursos				Sistemas de Informação			
	n	%	Nota	Corr.	N	%	Nota	Corr.
Nenhum, curso gratuito	147009	33,2	47,0±14,7	0,17	2896	24,4	48,1±13,8	0,15
Nenhum, curso não gratuito	133750	30,2	41,4±13,7	-0,10	3329	28,0	43,0±13,0	-0,08
ProUni Integral	27850	6,3	48,5±12,8	0,09	1107	9,3	49,3±12,9	0,11
ProUni parcial	5321	1,2	44,2±12,8	0,00	283	2,4	45,4±11,4	0,01
FIES	64056	14,5	40,2±12,3	-0,10	2190	18,4	41,3±11,9	-0,13
ProUni Parcial e FIES	3355	0,8	43,2±12,4	0,00	229	1,9	45,6±12,7	0,01
Bolsa estad/ munic	13094	3,0	40,8±13,1	-0,04	350	2,9	42,3±12,1	-0,03
Bolsa IES	31119	7,0	42,9±13,6	-0,01	1024	8,6	44,6±12,3	0,00
Outras bolsas	9179	2,1	41,5±13,0	-0,02	311	2,6	44,1±13,1	-0,01

Fin. IES	5760	1,3	38,8±12,8	-0,04	125	1,1	42,0±12,6	-0,02
Fin. bancário	2003	0,5	36,3±12,6	-0,04	48	0,4	40,5±11,0	-0,02
TOTAL	442496	100	43,6±14,1	-	11892	100	44,7±13,2	-

Fonte: Elaboração própria com dados do INEP-Enade 2017.

A seguir, avalia-se a questão QE_I13 na Tabela 6, que pergunta se o aluno recebeu algum tipo de bolsa acadêmica. Ao analisar todos os cursos, conclui-se que 74,5% dos candidatos afirmaram não terem recebido bolsa acadêmica (alternativa A) que obtiveram nota média mais baixa dentre as alternativas desta questão, com 42,1 pontos. A alternativa B (bolsa de iniciação científica) foi preenchida por 9% dos inscritos, que contabilizaram média de 50,5 pontos. As bolsas de extensão foram recebidas por 3,2% os estudantes, cuja média é de 48,9. As bolsas de monitoria/tutoria foram escolhidas por 3,4% dos participantes, que ficaram com média de 50,1 pontos. Ainda há as bolsas PET e outros tipos de bolsas que ficaram com respectivamente, 0,7% com nota 52,1 e 9,1% com nota 44,6. Em suma, demonstra-se que os alunos bolsistas tendem a ter um desempenho melhor em comparação aos não-bolsistas.

Da mesma forma, percebe-se que os estudantes de Sistemas de Informação que possuem alguma bolsa obtiveram desempenho superior aos não-bolsistas. Cerca de 81,1% dos candidatos afirmaram não terem recebido bolsa acadêmica e obtiveram a nota média mais baixa, com 44,0 pontos. Aqueles que receberam bolsa de iniciação científica compreendem 4,9% dos inscritos, que contabilizaram média de 51,0 pontos. As bolsas de extensão foram recebidas por 2,9% os estudantes, cuja média é de 48,3. As bolsas de monitoria/tutoria foram escolhidas por 3,2% dos participantes, que ficaram com média de 50,7 pontos. Ainda há as bolsas PET com 0,6% e nota 52,2, além de outros tipos de bolsas com 7,3% e nota 44,7.

Esta diferença no desempenho de bolsistas e não bolsistas foi também notada no trabalho de Lopes et al. (2020), que utilizou os dados do Enade 2017 dos estudantes de Licenciatura em Educação Física do estado do Ceará. Foi aferido que apenas uma pequena parte dos estudantes são contemplados com algum tipo de bolsa, porém estes contemplados têm melhores desempenhos pois dedicam mais tempo às atividades institucionais. Este engajamento amplia os conhecimentos dos estudantes e, por conseguinte, o seu desempenho acadêmico. Além disto, o desempenho também é positivamente influenciado pelo recebimento do auxílio financeiro, o que facilita a aquisição de livros, materiais didáticos e outros recursos (MENDONÇA et al., 2015; MOREIRA et al., 2019).

Outra variável a ser destacada é a participação de atividades curriculares realizadas no exterior durante a graduação (QE_I14), visualizada na Tabela 7. No caso geral, percebe-se que os estudantes participantes do Programa Ciência Sem Fronteiras (CSF) obtiveram nota média de 54,4 pontos, sendo este o grupo com melhor desempenho entre todos os estudantes. Por outro lado, o menor desempenho foi obtido entre aqueles que não participaram de intercâmbio, cuja nota média foi de 43,2, ou seja, 11,2 pontos menor que o CSF.

Entre os estudantes de Sistemas de Informação, a grande maioria de 96,7% dos participantes declarou não ter participado de atividade no exterior e obteve nota média de 44,5. Os participantes CSF são 1,8% do total, que atingiram nota média de 55,6. Os outros programas de intercâmbio, possuem menor representatividade, que são: programas de intercâmbio financiados pelo Governo Federal, pelo Governo Estadual, pela própria IES do estudante e intercâmbio não institucional. Respectivamente, estes grupos representam 0,06%, 0,002%, 0,44% e 0,94% do total, com notas 49,2, 40,5, 45,8 e 45,4. De uma maneira geral, percebe-se que os participantes de algum programa de intercâmbio possuem médias superiores aos que não participaram de nenhum. Sobretudo, repete-se o comportamento observado para todos os cursos, cujo melhor desempenho foi obtido

pelos participantes do CSF, que atingiram, em média, 11 pontos a mais do que os não-intercambistas.

Tabela 6: Bolsas acadêmicas.

QE_I13	Todos os cursos				Sistemas de Informação			
	n	%	Nota	Corr.	N	%	Nota	Corr.
Nenhuma	329852	74,5	42,1±13,6	-0,19	9646	81,1	44,0±12,9	-0,12
Iniciação científica	39961	9,0	50,5±14,3	0,15	579	4,9	51,0±13,7	0,11
Extensão	13969	3,2	48,9±14,2	0,07	350	2,9	48,3±12,9	0,05
Monitoria	15217	3,4	50,1±14,3	0,09	382	3,2	50,7±14,1	0,08
PET	3244	0,7	52,1±14,5	0,05	70	0,6	52,2±13,3	0,04
Outras	40253	9,1	44,6±14,1	0,02	865	7,3	44,7±12,6	0,00
TOTAL	442496	100	43,6±14,1	-	11892	100	44,7±13,2	-

Fonte: Elaboração própria com dados do INEP-Enade 2017.

O trabalho de Conceição (2017) avaliou amplamente o efeito do CSF no desempenho dos estudantes no Enade 2013, que avaliou os cursos da área da Saúde e Ciências da Terra. Foi constatado que os participantes do CSF têm mais chances de serem mais jovens, de estarem nas faixas de maior renda familiar, de serem solteiros, de estudarem em instituições públicas, de trabalharem menos de 20 horas semanais e de possuírem alguma tipo de bolsa acadêmica, em comparação com os não participantes do programa. Algumas destas características que compõem o perfil do participante do CSF estão relacionadas com as melhores pontuações, conforme relatado nas tabelas apresentadas anteriormente e na literatura (MESSER; WOLTER, 2007). Além disso, o intercâmbio contém características inerentes que estimulam o bom desempenho do aluno, tais quais autonomia e independência (LOPES et al., 2020; CARVALHO et al., 2016).

Tabela 7: Participação em programas de intercâmbio.

QE_I14	Todos os cursos				Sistemas de Informação			
	n	%	Nota	Corr.	N	%	Nota	Corr.
Sem intercâmbio	419183	94,7	43,2±13,9	-0,14	11503	96,7	44,5±13,0	-0,09
Ciência sem Fronteiras	13510	3,1	54,4±13,7	0,14	216	1,8	55,6±13,4	0,11
Intercâmbio Gov. Federal	1214	0,3	52,9±14,9	0,03	7	0,1	49,2±28,5	0,01
Intercâmbio Gov. Estad.	208	0,0	43,4±15,5	0,00	2	0,0	40,5±18,8	0,00
Intercâmbio da IES	3147	0,7	48,6±15,6	0,03	52	0,4	45,8±14,7	0,01

Outro intercâmbio	5234	1,2	46,8±14,2	0,02	112	0,9	45,4±14,6	0,01
TOTAL	442496	100	43,6±14,1	-	11892	100	44,7±13,2	-

Fonte: Elaboração própria com dados do INEP-Enade 2017.

Na árvore de decisão aparece a questão QE_I16, opção 31 (MG) em destaque. Trata-se da unidade federativa em que o participante concluiu o ensino médio. A árvore foi gerada a partir dos dados dos estudantes de Sistemas de Informação, apenas. Assim, percebeu-se que as três maiores notas médias por estado neste curso, foram obtidas por Espírito Santos ($50,0 \pm 12,9$), Mato Grosso do Sul ($47,5 \pm 13,5$) e Minas Gerais ($47,3 \pm 12,9$). Em termos de números de participantes, os três maiores estados são: São Paulo (19,8%), Minas Gerais (16,7%) e Rio de Janeiro (10,6%). Desta forma, o estado de Minas Gerais contém a melhor média entre os estados com maior população, uma vez que Espírito Santo possui apenas 2,4% dos estudantes e Mato Grosso do Sul, 0,73%.

Na Tabela 8, percebe-se que para os alunos de SI, a média geral é de 44,7 e a média dos estados exceto Minas Gerais é 44,3, ambos valores inferiores à média de Minas Gerais. Ao se observar todos os dados do Enade 2017, esta diferença é menor: 43,6 na média total, contra 43,5 de média de todos os estados exceto MG e 44,8 na média somente do estado de MG.

Na literatura são encontrados diversos estudos sobre a diferença de desempenho estudantil entre as regiões geográficas brasileiras, porém com conclusões divergentes. Por exemplo, o trabalho de Sprenger et al. (2018) avaliou os estudantes de Ciências Contábeis, e concluiu que a região geográfica pode interferir no desempenho, devido aos fatores regionais de índice de desenvolvimento humano, expectativa de vida e nível de escolaridade da população local. Neste sentido, Medeiros Filho (2019) também encontrou desempenhos diferentes entre as regiões geográficas ao analisar os dados do Enade dos estudantes de Educação Física. Percebeu-se que há chances de notas mais elevadas para alunos de IES localizadas nas regiões Nordeste e Norte, o que pode estar relacionado com as características socioculturais e demanda quanto ao público investigado. Por outro lado, Abdalla et al. (2016) realizou um estudo sobre o desempenho dos estudantes de Enfermagem no Enade, em que se concluiu que as notas médias não apresentaram grandes disparidades entre as unidades federativas. Na prova de Componentes Específicos, a nota média foi 54,7, sendo que a maior média foi obtida pelas IES do estado de Minas Gerais com 59,0 pontos, e a pior média foi de 50,0 pontos, obtida no estado do Piauí. Portanto, se faz necessária uma pesquisa mais ampla para a melhor compreensão da relação entre regionalidade e desempenho de alunos de um determinado curso.

Tabela 8: UF de conclusão do ensino médio.

QE_I16	Todos os cursos				Sistemas de Informação			
	n	%	Nota	Corr.	N	%	Nota	Corr.
Exceto MG	389174	87,9	43,5±14,1	0,03	9947	83,6	44,3±13,0	0,08
MG	53322	12,1	44,8±14,0	0,03	1945	16,4	47,3±13,4	0,08
TOTAL	442496	100	43,6±14,1	-	11892	100	44,7±13,2	-

Fonte: Elaboração própria com dados do INEP-Enade 2017.

A questão QE_I17 relata o tipo de escola em que o ensino médio foi cursado pelo aluno e seus dados estão exibidos na Tabela 9. O quadro geral indica que a maior parte estudou em escola pública, cerca de 68,6%. Este grupo possui a correlação mais negativa, o que indica a influência

negativa na nota. Fato este ratificado pela média baixa em comparação aos demais grupos: 42,2. A correlação mais positiva (0,18) ocorre nos oriundos de escola particular, que obtiveram nota média 48,1.

Estas características também são observadas na amostra dos alunos de Sistemas de Informação, em que maior parte, cerca de 66,5%, cursou o ensino médio totalmente em escola pública, que obteve nota média de 44,1. O segundo maior grupo cursou o ensino médio totalmente em escola particular, o que inclui 24,3% do total. Estes alunos obtiveram nota média de 47,1, exatamente 3 pontos a mais que a escola pública. A seguir encontram-se outros grupos com menor número de participantes, que cursaram o ensino médio integralmente no exterior (0,1% e nota 39,7), em sua maior parte em escola pública (4,8% e nota 41,8), em sua maior parte em escola privada (4,1% e nota 44,8) e parcialmente no exterior (0,2% e nota 48,0). Nota-se que as diferenças na nota média entre os tipos de escola são pequenas, principalmente ao se relativizar com o desvio-padrão. De qualquer maneira, conclui-se que a maior parte dos estudantes advém de escolas públicas e a maior média foi obtida foi pelos estudantes da rede privada.

A tendência de haver melhores notas em alunos que cursaram o ensino médio em escolas particulares foi percebida em diversos cursos, como Nutrição (ROCHA et al., 2018), Ciências Contábeis (FERREIRA, 2015) e Educação Física (Medeiros Filho, 2019), graças ao acesso a melhores infraestruturas e recursos didáticos. Porém esta tendência não foi encontrada em cursos de Engenharia no Enade 2011 (MORICONI; NASCIMENTO, 2014). Devido à forte competição destes cursos, os selecionados foram os alunos de alto desempenho das escolas públicas, que no geral, são oriundos de escolas federais, que por sua vez possuem bons indicadores.

Tabela 9: Tipo de escola em que cursou o ensino médio.

QE_I17	Todos os cursos				Sistemas de Informação			
	n	%	Nota	Corr.	n	%	Nota	Corr.
Total pública	303741	68,6	42,2±13,7	-0,15	7906	66,5	44,1±12,9	-0,07
Total particular	104418	23,6	48,1±14,3	0,18	2884	24,3	47,1±13,5	0,10
Total exterior	463	0,1	43,9±15,0	0,00	17	0,1	39,7±11,9	-0,01
Maioria pública	18152	4,1	41,5±13,8	-0,03	576	4,8	41,8±13,1	-0,05
Maioria particular	14816	3,3	43,7±13,9	0,00	485	4,1	44,8±12,9	0,00
Parcial exterior	906	0,2	50,5±14,6	0,02	24	0,2	48,0±12,5	0,01
TOTAL	442496	100	43,6±14,1	-	11892	100	44,7±13,2	-

Fonte: Elaboração própria com dados do INEP-Enade 2017.

A questão QE_I23 trata sobre quantas horas por semana, aproximadamente, são dedicadas aos estudos, excetuando as horas de aula. Os dados estão exibidos na Tabela 10. Ao se avaliar todos os registros do Enade 2017, observa-se que a correlação entre esta variável e a nota é de 0,15, o que demonstra uma nota média crescente à medida que as horas de dedicação aumentam,

saltando de 40,4 pontos em 0 horas de dedicação para 47,4 para aqueles que se dedicam mais de 12 horas semanais.

Para os alunos de Sistemas de Informação, este comportamento é similar. A índice de correlação é dada por 0,11. Ao todo, 7,3% dos alunos não se dedicam nenhuma hora por semana e obtiveram nota média 43,5. De uma a três horas de dedicação está a maior parte dos alunos: 43,9%, com média 43,1. Aqueles que se dedicam de oito a doze horas são 29,1% do total, com média um pouco maior que os grupos supracitados: 45,9. Entre aqueles que se dedicam ainda mais, temos o grupo de oito a 12 horas, com 10,8% e média 47,1 e finalmente, tem-se 9% dos participantes que se dedicam mais de doze horas, com média 47,1. De fato, a influência positiva da dedicação na nota tem sido relatada na literatura (LOPES et al, 2020; ARAÚJO, 2017).

Tabela 10: Horas de dedicação semanais aos estudos fora de sala de aula.

QE_I23	Todos os cursos				Sistemas de Informação			
	n	%	Nota	Corr.	n	%	Nota	Corr.
0h	17093	3,9	40,4±14,3	0,15	868	7,3	43,5±13,7	0,11
De 1 a 3h	181222	41,0	41,5±13,6	0,15	5215	43,9	43,1±12,9	0,11
De 4 a 7h	133164	30,1	44,3±13,9	0,15	3458	29,1	45,9±13,2	0,11
De 8 a 12h	57824	13,1	46,1±14,2	0,15	1285	10,8	47,1±12,5	0,11
>12h	53193	12,0	47,4±14,5	0,15	1066	9,0	47,1±13,6	0,11
TOTAL	442496	100	43,6±14,1	-	11892	100	44,7±13,2	-

Fonte: Elaboração própria com dados do INEP-Enade 2017.

Em sequência, analisa-se qual principal razão para o estudante ter escolhido a sua instituição de educação superior (QE_I26). Em ambos os conjuntos avaliados, com todos os cursos e o curso de Sistemas de Informação, a maior média foi obtida pelo grupo que escolheu a IES por Qualidade/reputação, com respectivamente 46,0 e 46,5 de média. Esta opção também obteve o índice de correlação mais positivo em ambos os conjuntos, com 0,12 e 0,10. A segunda maior ocorreu entre os que escolheram a IES por ser gratuita, também em ambos os conjuntos. A pior média ocorreu entre os que escolheram a IES pelo preço da mensalidade, que em ambos os conjuntos obteve o índice de correlação mais negativo.

Segundo Teixeira (2020), os estudantes que escolhem a IES por qualidade e reputação têm um perfil socialmente favorecido, pois este grupo é propenso a ter estudantes oriundos de escolas particulares no ensino médio e com maiores rendas familiares, fatores que já comprovadamente influenciam positivamente no desempenho. Por outro lado, populações menos favorecidas tendem a escolher cursos pela gratuidade e possibilidade de ter bolsa.

Tabela 11: Razão de escolha da IES.

QE_I26	Todos os cursos				Sistemas de Informação			
	n	%	Nota	Corr.	n	%	Nota	Corr.
Gratuidade	77093	17,4	44,5±14,4	0,03	1495	12,6	46,2±13,6	0,04
Preço da mensalidade	37815	8,5	38,6±13,2	-0,11	920	7,7	40,3±12,4	-0,10

Proxim. da residência	71463	16,1	42,6±13,5	-0,03	2198	18,5	43,6±12,8	-0,04
Proxim. do trabalho	6070	1,4	40,0±13,2	-0,03	291	2,4	45,1±13,1	0,00
Facilidade de acesso	31275	7,1	40,3±13,5	-0,06	682	5,7	41,2±2,9	-0,07
Qualidade reputação	150312	34,0	46,0±14,1	0,12	4432	37,3	46,5±13,1	0,10
Única aprovação	6312	1,4	44,5±13,8	0,01	175	1,5	44,1±14,1	-0,01
Possibilid. de bolsa	22769	5,1	44,1±13,4	0,01	783	6,6	45,7±12,4	0,02
Outro	39387	8,9	42,1±14,0	-0,03	916	7,7	42,9±12,9	-0,04
TOTAL	442496	100	43,6±14,1	-	11892	100	44,7±13,2	-

Fonte: Elaboração própria com dados do INEP-Enade 2017.

Finalmente, será analisada uma questão que não apareceu nem na árvore e nem na matriz com as principais correlações, que foi incluída por ser de interesse em diversas discussões acerca do assunto: a questão QE_I10 “Qual alternativa a seguir melhor descreve sua situação de trabalho (exceto estágio ou bolsas)?” que trata sobre a carga horária semanal de trabalho do estudante. A análise de todos os participantes do Enade 2017 mostra que a quantidade de horas trabalhadas por semana impacta negativamente nas notas. Os estudantes que declararam não trabalharem são 38,1% do total e obtiveram a maior média, com 45,1 pontos. Em seguida estão grupos que trabalham eventualmente, menos de 20 hora semanais e entre 21 e 39 horas semanais, que respectivamente contém 7,8%, 8,6% e 11,9% dos participantes, com média 42,8, 43,2 e 44,6. Por último, há o grupo que trabalha 40 horas semanais ou mais, que são 33,6% do total e obtiveram a média mais baixa, 41,9.

Por outro lado, o desempenho dos estudantes de Sistemas de Informação comportou-se diferentemente. Aqueles que não trabalham totalizam 25,1%, cuja nota média foi de 45,2. Os que trabalham eventualmente são 4,3% dos candidatos, com nota média 43,0. Cerca de 3,7% trabalham até 20 horas semanais, cuja nota é 43,0. Já os estudantes que trabalham de 21 a 39 horas por semana são 10,7% com média 45,3. Por último, consta o maior grupo dos candidatos com 56,3% do total, das pessoas que trabalham 40 horas ou mais, que obtiveram nota 44,70. Desta forma, pode-se afirmar que grande maioria trabalha em tempo integral, ao contrário do que foi apresentado nos demais cursos. Além disto, pode-se concluir que a carga horária de trabalho não interfere na nota, pois todos os grupos alcançaram entre 43 e 45 pontos. De fato, correlação entre as horas de trabalho e nota pode ser considerada muito baixa, cerca de -0,01. A distribuição de notas e os quantitativos entre os grupos constam na Tabela 12.

Existe um número relevantes de universitários que trabalham e estudam simultaneamente, porém há poucos estudos sobre a influência deste aspecto sobre o rendimento discente (NIQUINI et al., 2015). Neste sentido, alguns estudos apontam que o trabalho pode interferir negativamente no desempenho estudantil. Por exemplo, demonstrou-se que os matriculados nos cursos de Serviço Social, Engenharia e Direito e Humanidades no Chile que trabalham e estudam tem o desempenho prejudicado (BARAHONA, 2014). O trabalho de Vargas (2014) foi ao encontro desta mesma relação ao analisar estudantes da Costa Rica. No âmbito nacional, alguns estudos

também encontraram considerável relação entre trabalho e baixo desempenho no Enade (ROCHA et al., 2018).

Em contrapartida, Silva (2016) percebeu que no geral a nota do Enade é influenciada negativamente pelo trabalho, porém há algumas exceções em que o trabalho não interfere no rendimento. Em especial, destaca-se que o aluno-empregado tem mais chances de ter boas notas quando exerce seu ofício em um ambiente favorável à sua condição de estudante. Nesta perspectiva, Gualberto e Rodrigues (2017) discutiram os efeitos na nota do Enade de estudantes de licenciaturas em Física, Química, Matemática e Ciências Biológicas. Conclui-se que apesar de haver uma diferença na média entre os que trabalham e os que não, ela é pouca ou insignificante.

Para fundamentar as discussões sobre as variáveis analisadas, foram utilizados trabalhos que estudaram os aspectos socioeconômicos que impactam no desempenho estudantil, preferencialmente relacionados ao curso de Sistemas de Informação. Todavia, não foram encontradas na literatura referências suficientes para discutir todos os aspectos focados apenas neste curso. Desta forma, foram também utilizadas análises realizadas em outras graduações, mesmo que em alguns casos elas possuam perfis e comportamentos diferentes entre si.

Tabela 12: Horas de trabalho semanais.

QE_I10	Todos os cursos				Sistemas de Informação			
	n	%	Nota	Corr.	n	%	Nota	Corr.
Não trabalha	168631	38,1	45,1±14,2	-0,09	507	4,3	43,0±13,5	-0,01
Eventual	34720	7,8	42,8±14,4	-0,09	442	3,7	43,2±13,2	-0,01
< 20h/sem	38025	8,6	43,2±14,5	-0,09	1273	10,7	45,3±13,3	-0,01
21 -39h/sem	52456	11,9	44,6±14,2	-0,09	6691	56,3	44,7±13,0	-0,01
> 40h/sem	148664	33,6	41,9±13,4	-0,09	8788	73,9	43,6±12,7	-0,14
TOTAL	442496	100	43,6±14,1	-	11892	100	44,7±13,2	-

Fonte: Elaboração própria com dados do INEP-Enade 2017.

5. Conclusão

O artigo investigou a complexa base de dados do Enade, composta por 11892 registros e 75 colunas, apenas composta por estudantes do curso de Sistemas de Informação no ano de 2017. Foi feito um levantamento dos aspectos socioeconômicos mais influentes na nota do Enade, que foram colhidos pelo questionário aplicado no dia do exame.

Notou-se que todas as variáveis possuem baixa correlação com a nota, o que implica em uma relação fraca entre os aspectos socioeconômicos e a nota. Porém podem-se destacar alguns destes aspectos entre os 75 avaliados. Foram selecionados os 10 aspectos com maior correlação com a nota ou que constaram nos três primeiros níveis da árvore de classificação.

O fator mais relevante é a renda, em que se percebe uma diferença crescente entre o grupo mais vulnerável e os mais abastados. A pior nota média é 41,5 e ocorre entre aqueles pertencentes a famílias com renda inferior a 1,5 salários-mínimos. Por outro lado, à medida que a renda aumenta, a nota também aumenta, atingido a maior média de 51,7 no grupo com faixa salarial acima de 30 salários. A maior parte dos estudantes encontra-se nas faixas entre 1,5 e 4,5 salários.

Outro aspecto discutido é o recebimento de bolsas de estudo e financiamento, em que se destacou o desempenho dos alunos das IES públicas que não precisam deste tipo de financiamento, com nota média 48,1 e os alunos que receberam bolsa integral do ProUni, com média 49,3. O destaque negativo ficou entre os bolsistas do FIES, que receberam média 41,3.

O tipo de instituição também se revelou ser um aspecto relevante, em que as particulares contém a maior parte dos alunos de Sistemas de Informação, com uma fatia de 73,9% do total, que obtiveram nota média de 43,6, valor menor que os alunos de IES públicas, que ficaram com média 47,9.

Outros aspectos que também influenciaram a nota positivamente foram a participação em bolsas acadêmicas (iniciação científica, monitoria, extensão, PET) e a participação em intercâmbios, principalmente do Programa Ciência Sem Fronteiras. Por outro lado, notou-se que a dedicação ao trabalho em horas semanais não interferiu na nota dos estudantes.

Ao se analisar a literatura e os dados de todos os cursos do Enade 2017, percebeu-se que o fato do estudante possuir um vínculo empregatício concomitante aos estudos é um fator relacionado com desempenhos mais baixos. Todavia, esta tendência não foi aferida para os estudantes de Sistemas de Informação.

Em vista aos resultados apresentados, foi possível determinar os aspectos mais influentes no desempenho acadêmico dos estudantes de Sistemas de Informação (SI). A principal contribuição deste trabalho foi apresentar uma análise completa sobre todos os estudantes de SI participantes do Enade, o que preenche uma lacuna na literatura, uma vez que estudos específicos para esta finalidade não foram encontrados. Os mais semelhantes realizaram análises sobre o curso de SI em universidade específica (MIZAEL; LIMA; 2018) ou análises mais abrangentes, porém feitas em outros cursos, como Ciências Contábeis (LE MOS; MIRANDA; 2015). Ademais, este trabalho pode auxiliar futuras pesquisas de natureza qualitativa, para que se possa captar aspectos mais subjetivos que influenciam o desempenho estudantil, uma vez que a metodologia apresentada está limitada à sua natureza quantitativa e ao questionário aplicado pelo INEP. Nesta perspectiva, em um trabalho futuro, pretende-se investigar mais detalhadamente a variável de vínculo empregatício, para que se possa compreender os motivos da diferença de comportamento entre os estudantes de SI e os demais. Uma hipótese para este fenômeno é que os estudantes de SI tendem a trabalhar na sua área de formação e que, portanto, o seu ambiente de trabalho é compatível aos estudos.

Além disto, como dito pelos pesquisadores Lima et al., (2021), apenas o Enade não soluciona o problema de defasagem educacional do ensino superior brasileiro, mas é um instrumento de investigação que pode ser usado para estudos e pesquisas para aprimorar tanto o ensino quanto a ferramenta. Nesta lógica, os resultados apresentados são relevantes para fomentar discussões de políticas públicas e/ou institucionais, de forma a buscar por melhorias no processo de aprendizagem, nas formas de financiamento, na concessão de bolsas acadêmicas e projetos de permanência e manutenção dos estudantes.

Referências

Abdalla, I. G., Moreno, L., Abensur, P., Seiffert, O. M. L. B., & Limeira, P. (2016). O padrão oficial de qualidade dos cursos de Enfermagem: uma análise dos resultados do Enade (2004, 2007 e 2010) e da associação com outros indicadores do Sinaes. SEMINÁRIO NACIONAL UNIVERSITAS/BR, 24, 611-634. [[GS Search](#)]

- Alqallaf, F. A., Konis, K. P., Martin, R. D., & Zamar, R. H. (2002). Scalable robust covariance and correlation estimates for data mining. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 4-23. DOI: [10.1145/775047.775050](https://doi.org/10.1145/775047.775050). [GS Search]
- Alves, R. D. (2018). Predição do desempenho da redação do Enem utilizando técnicas de mineração de dados. [Trabalho de Conclusão de Curso, Universidade Federal de Santa Catarina. Campus Araranguá]. Repositório Institucional UFSC. Disponível em: <https://repositorio.ufsc.br/handle/123456789/187760>. [GS Search]
- Araújo, F. H. R. D. (2017). Determinantes do desempenho dos alunos das Instituições de Ensino Superior Brasileiras no ENADE 2013. [Dissertação de Mestrado, Universidade Federal do Ceará]. Repositório Institucional UFC. Disponível em: <http://repositorio.ufc.br/handle/riufc/29295>. [GS Search]
- Barahona, P. (2014). Factores determinantes del rendimiento académico de los estudiantes de la Universidad de Atacama. Estudios pedagógicos (Valdivia), 40(1), 25-39. [GS Search]
- Bezerra, M. E. G., & Tassigny, M. M. (2018). A relação entre a política de financiamento estudantil e o desempenho dos estudantes de administração no Enade. Archivos Analíticos de Políticas Educativas= Education Policy Analysis Archives, 26(1), 5. [GS Search]
- Bhargava, N., Sharma, G., Bhargava, R., & Mathuria, M. (2013.) Decision tree analysis on j48 algorithm for data mining. Proceedings of international journal of advanced research in computer science and software engineering, 3(6). [GS Search]
- Bittencourt, H. R., Viali, L., Casartelli, A. D. O., & Rodrigues, A. C. D. M. (2008). Uma análise da relação entre os conceitos Enade e IDD. Estudos em avaliação educacional, 19(40), 247-262. [GS Search]
- BRASIL (2004). Lei nº 10.861, de 14 de abril de 2004. Institui o Sistema Nacional de Avaliação da Educação Superior-SINAES e dá outras providências. Diário Oficial da União, (72).
- Caetano, C. C. R., de Oliveira Cardoso, T. A., Miranda, G. J., & de Freitas, S. C. (2016). Desempenho no ENADE em Ciências Contábeis: ensino a distância (EAD versus presencial). Revista Universo Contábil, 11(4), 147-165. [GS Search]
- Carvalho, E. J. G. (2016). Política e Gestão da Educação: explorando o significado dos termos. Série-Estudos-Periódico do Programa de Pós-Graduação em Educação da UCDB, 77-96. [GS Search]
- Cohen, J. (2013). Statistical power analysis for the behavioral sciences. Academic press. [GS Search]
- Conceição, O. C. (2017). Uma avaliação do programa Ciência sem Fronteiras: efeitos sobre desempenho, trancamento e conclusão. [Dissertação de Mestrado, Pontifícia Universidade Católica do Rio Grande do Sul]. Biblioteca Digital de Teses e Dissertações PUCRS. Disponível em: <http://tede2.pucrs.br/tede2/handle/tede/7605>. [GS Search]
- Costa, I. S., Rodrigues, I. S., Silveira, S. D. F. R., & Rodrigues, C. T. (2020). O impacto de uma Política Pública Educacional no Enade: uma avaliação do FIES. Revista Meta: Avaliação, 12(35), 437-468. DOI: [10.22347/2175-2753v12i35](https://doi.org/10.22347/2175-2753v12i35). [GS Search]
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37-37. DOI: [10.1609/aimag.v17i3.1230](https://doi.org/10.1609/aimag.v17i3.1230). [GS Search]
- Ferreira, M. A. (2015). Determinantes do desempenho discente no ENADE em cursos de Ciências Contábeis. [Dissertação de Mestrado, Universidade Federal de Uberlândia]. Repositório

- Institucional UFU. Disponível em: <http://repositorio.ufu.br/handle/123456789/12620>. [GS Search]
- Figueiredo Filho, D. B., & Silva Junior, J. A. (2009). Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r). *Revista Política Hoje*, 18(1), 115-146. [GS Search]
- Goldschmidt, R., & Passos, E. (2005). *Data mining: um guia prático*. Gulf Professional Publishing. [GS Search]
- Griboski, C. M. (2012). O Enade como indutor da qualidade da educação superior. *Estudos em avaliação educacional*, 23(53), 178-195. DOI: [10.18222/eaec235320121920](https://doi.org/10.18222/eaec235320121920). [GS Search]
- Gualberto, L., & Rodrigues, A. (2017). Impacto do trabalho no desempenho dos licenciandos no ENADE. *Ensenanza de las Ciencias, (Extra)*, 2675-2680. [GS Search]
- Lee, I. N., Liao, S. C., & Embrechts, M. (2000). Data mining techniques applied to medical information. *Medical informatics and the Internet in medicine*, 25(2), 81-102. DOI: [10.1080/14639230050058275](https://doi.org/10.1080/14639230050058275) [GS Search]
- Lemos, K. C. S., & Miranda, G. J. (2015). Alto e Baixo Desempenho no Enade: que variáveis explicam? *REVISTA AMBIENTE CONTÁBIL-Universidade Federal do Rio Grande do Norte*, 7(2), 101-118. [GS Search]
- Lima, P. D. S. N., Ambrósio, A. P. L., dos Santos Oliveira, J. L., & de Carvalho, C. L. (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. DOI: [10.5753/rbie.2021.29.0.385](https://doi.org/10.5753/rbie.2021.29.0.385). [GS Search]
- Lopes, J. M. R., SOUSA, L., & SANTOS, M. (2020). Fatores associados ao desempenho acadêmico de estudantes de licenciatura em Educação Física do Ceará no ENADE. *Revista Educação & Linguagem*, 7, 62-74. [GS Search]
- Medeiros Filho, A. E. C de., Junior, J. A. D. F. P., Santa Rosa, L. F. R., de Santana, K. S., & Magalhães, R. D. C. B. P. (2019). Desempenho de estudantes de Educação Física em Educação Inclusiva no Enade/Brasil. *Research, Society and Development*, 8(7), 29. [GS Search]
- Mello Neto, R. D, Medeiros, H. A. V., & Catani, A. M. (2020). ProUni e retornos de capitais para instituições privadas de ensino superior: uma leitura para além dos benefícios diretos com o financiamento público. *Revista Internacional de Educação Superior*, 6, e020028. DOI: [10.20396/riesup.v6i0.8655907](https://doi.org/10.20396/riesup.v6i0.8655907). [GS Search]
- Mendonça, C. M. C., Guerra, L. C. B., de Moura Mendonça, A. V. P., de Sousa Neto, M. V., & Fernandes, A. S. A. (2015). Políticas de acesso à Educação Superior e desempenho de alunos bolsistas: o caso de uma universidade privada de Natal/RN. *RAUnP*, 7(2), 15-29. DOI: [10.21714/raunp.v7i2.677](https://doi.org/10.21714/raunp.v7i2.677). [GS Search]
- Messer, D., & Wolter, S. C. (2007). Are student exchange programs worth it?. *Higher education*, 54(5), 647-663. DOI: [10.1007/s10734-006-9016-6](https://doi.org/10.1007/s10734-006-9016-6). [GS Search]
- Mizael, G. R., & Lima, J. L. 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*, 2. [GS Search]
- Moreira, G. D. O., Passeri, S., Velho, P. E. N. F., Ferraresi, F., Appenzeller, S., & Amaral, E. (2019). The academic performance of scholarship students during medical school. *Revista*

- Brasileira de Educação Médica, 43, 163-169. DOI: [10.1590/1981-52712015v43n3RB20180180](https://doi.org/10.1590/1981-52712015v43n3RB20180180). [[GS Search](#)]
- Moriconi, G. M., & Nascimento, P. A. M. M. (2014). Fatores associados ao desempenho dos concluintes de engenharia no Enade 2011. *Estudos em avaliação educacional*, 25(57), 248-278. DOI: [10.18222/eaec255720142831](https://doi.org/10.18222/eaec255720142831). [[GS Search](#)]
- Namen, A. A., Borges, S. X. D. A., & Sadala, M. D. G. S. (2013). Indicadores de qualidade do ensino fundamental: o uso das tecnologias de mineração de dados e de visões multidimensionais para apoio à análise e definição de políticas públicas. *Revista Brasileira de Estudos Pedagógicos*, 94(238), 677-700. DOI: [10.24109/2176-6681.rbep.94i238.353](https://doi.org/10.24109/2176-6681.rbep.94i238.353). [[GS Search](#)]
- Niquini, R. P., Teixeira, L. R., Sousa, C. A. D., Manelli, R. N., Luz, A. A. D., Turte-Cavadinha, S. L., & Fischer, F. M. (2015). Características do trabalho de estudantes universitários associadas ao seu desempenho acadêmico. *Educação em Revista*, 31, 359-381. DOI: [10.1590/0102-4698122477](https://doi.org/10.1590/0102-4698122477). [[GS Search](#)]
- Nogueira, E. D. A., & Tsunoda, D. F. (2015). Mineração de dados para análise da relação entre as características socioeconômicas de concluintes do ensino superior e o desempenho desses estudantes no Enade 2012. *Percurso*, 15(1), 245-268. [[GS Search](#)]
- Pasta, A. (2011). Aplicação da técnica de Data Mining na base de dados do ambiente de gestão educacional: um estudo de caso de uma instituição de ensino superior de Blumenau, SC. [Dissertação de Mestrado, Universidade do Vale do Itajaí]. Repositório Univali. Disponível em: <https://siaiap39.univali.br/repositorio/handle/repositorio/1044>. [[GS Search](#)]
- Ribeiro, L. D. S. (2010). Uma abordagem semântica para seleção de atributos no processo de KDD. [Dissertação de Mestrado, Universidade Federal da Paraíba]. Repositório Institucional UFPB. Disponível em: <https://repositorio.ufpb.br/jspui/handle/tede/6048>. [[GS Search](#)]
- Rocha, A. L. D. P., Leles, C. R., & Queiroz, M. G. (2018). Fatores associados ao desempenho acadêmico de estudantes de Nutrição no Enade. *Revista brasileira de Estudos pedagógicos*, 99(251), 74-94. DOI: [10.24109/2176-6681.rbep.99i251.3162](https://doi.org/10.24109/2176-6681.rbep.99i251.3162). [[GS Search](#)]
- Rolim, V., Ferreira, R., & Costa, E. (2017). Utilização de técnicas de aprendizado de máquina para acompanhamento de fóruns educacionais. *Revista Brasileira de Informática na Educação*, 25(03), 112-130. DOI: [10.5753/rbie.2017.25.03.112](https://doi.org/10.5753/rbie.2017.25.03.112). [[GS Search](#)]
- Scudeler, M. A., Calderon, A. I., & Pires, A. (2020). O desafio da educação a distância no ensino superior brasileiro: a expansão dos cursos de pedagogia a luz do Enade como índice de qualidade. *Jornal de Políticas Educacionais*, 14(7). DOI: [10.5380/jpe.v14i0.70138](https://doi.org/10.5380/jpe.v14i0.70138) [[GS Search](#)]
- Silva, D. F. D. (2016). Efeito do Trabalho no Desempenho de Alunos no Enade. [Dissertação de Mestrado, Universidade Federal de Santa Maria]. Repositório Digital da UFSM. Disponível em: <https://repositorio.ufsm.br/handle/1/9634>. [[GS Search](#)]
- Sharma, H., & Kumar, S. (2016). A survey on decision tree algorithms of classification in data mining. *International Journal of Science and Research (IJSR)*, 5(4), 2094-2097. [[GS Search](#)]
- Sprenger, K. B., Kronbauer, C. A., Silvestre, A. O., de Azevedo, E. R., & Alves, T. W. (2018). Fatores explicativos dos índices de aprovação no exame de suficiência contábil. *ConTexto*, 18(38). [[GS Search](#)]
- Teixeira, B. P. (2020). Capital cultural e background na escolha de um curso de graduação e de uma instituição de educação superior: uma análise dos microdados do Enade. [Trabalho de

- Conclusão de Curso, Universidade de Passo Fundo]. Repositório UPF. Disponível em: <http://repositorio.upf.br/handle/riupf/1968> [GS Search]
- Valera, J., Sinha, S., Varela, J., & Balaguer, E. P. (2009). Una explicación del rendimiento estudiantil universitario mediante modelos de regresión logística. *Visión gerencial*, (7)2, 415-427 [GS Search]
- Vargas, G. M. V. (2014). Factores asociados al rendimiento académico tomando en cuenta el nivel socioeconómico: Estudio de regresión múltiple en estudiantes universitarios. *Revista Electrónica Educare*, 18(1), 119-154. [GS Search]