

O Perfil Socioeconômico dos Concluintes de Computação do Sul do Brasil

Luis Otávio Oliveira Capelari
André Luis Schwerz
Universidade Tecnológica Federal do Paraná
Campo Mourão, PR, Brasil
luiscapelari@alunos.utfpr.edu.br
andreluis@utfpr.edu.br

ABSTRACT

Arguably, the computing has become one of the most important areas in the world, demanding more and more skilled labor. In this direction, higher education plays an important role in the training of professionals that supply the market demand. To measure the education quality of these professionals, the Enade evaluates the undergraduate's performance from all higher education institutions in Brazil. However, only a few studies have explored the information provided by the exam and its surveys. Based on the 2017 Enade edition, this paper identifies the socio-economic profiles of undergraduate candidates in several computing courses in the southern region of Brazil. As results, we found out that the computing student socio-economic profile consists of white people, male, between 18 and 26 years old and family income of 3 to 6 minimum wages. Also, most students have no scholarship, their parents attended up until high school, the high school completed in public school, and study in the night shift of private institutions. This work summarizes student's profiles, enabling managers in education to introduce new policies to improve higher education.

KEYWORDS

Mineração de Dados, Enade, Perfil, Destilação de Dados para Julgamento Humano, Python

1 INTRODUÇÃO

Recentemente, observa-se o crescimento da demanda de aplicações específicas para processar a grande quantidade de dados que são disseminados principalmente pelos governos com a adesão a políticas de dados abertos. No contexto educacional, a formação de professores [1], a diminuição da evasão [2] e a predição do desempenho dos estudantes [3] são exemplos de aplicações que podem tirar proveito dessa oferta de dados.

No Brasil, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) é uma autarquia federal vinculada ao Ministério da Educação (MEC) responsável por várias avaliações e pesquisas que são disponibilizadas em sua plataforma em formato aberto. Uma delas é o Exame Nacional de Desempenho de Estudantes (Enade) que avalia o desempenho dos concluintes de cursos superiores com base nos conteúdos programáticos previstos nas diretrizes curriculares dos respectivos cursos. Realizado anualmente, o Enade compõe o Sistema Nacional de Avaliação do Ensino Superior (Sinaes).

O Enade possui um ciclo avaliativo de três anos, a cada ano um conjunto diferente de cursos é avaliado. Nesse artigo, o interesse

está no Ano II do ciclo avaliativo, que avalia os cursos da área de computação. Nesse ciclo avaliativo, os cursos de computação foram avaliados em cinco exames (2005, 2008, 2011, 2014 e 2017). Os microdados do exame são disponibilizados de forma aberta pelo Inep em sua plataforma¹.

Embora possuam riqueza de informação sobre a educação superior, os microdados do Inep ainda são pouco explorados pelas instituições de ensino superior como forma de entender e melhorar a educação em seus cursos [4]. O levantamento sistêmico proposto em [5] mostra que a maioria das pesquisas analisam fatores determinantes para o desempenho no exame e, entre os trabalhos que analisam o perfil socioeconômico dos concluintes, nenhum deles trata especificamente dos cursos de computação. Diante deste cenário, a questão de pesquisa que norteia esta trabalho é: *É possível determinar o perfil dos concluintes dos cursos de computação, levando em consideração o resultado do questionário socioeconômico do Enade?*

Portanto, o intuito desse artigo é identificar o perfil socioeconômico dos alunos de vários cursos de computação que estudam em instituições da região sul do Brasil. A análise dos dados foi feita em um código na linguagem de programação Python, utilizando o método Destilação de Dados para Julgamento Humano, que é uma subárea de Mineração de Dados Educacionais [6, 7].

Com os resultados foram descobertas relações entre as variáveis socioeconômicas, além de ser identificado o perfil socioeconômico do discente de computação da região sul do Brasil. Os resultados podem servir como base para tomada de decisões das instituições na inclusão de políticas que visem, por exemplo, a mitigação da evasão, o aperfeiçoamento do ensino, a formação de professores, a melhoria dos currículos, a predição de desempenho, entre outras.

No restante deste trabalho, será explicado o conceito de Mineração de Dados Educacionais na Seção 2, para, em seguida, a Seção 3 apresentar o método utilizado; os resultados obtidos serão apresentados e discutidos na Seção 4, e, a Seção 5 discorrerá sobre a conclusão desse trabalho.

2 MINERAÇÃO DE DADOS EDUCACIONAIS

A Mineração de Dados é uma área que engloba estatística, aprendizagem de máquina, reconhecimento de padrões e visualização de dados. Ela é considerada parte de um processo maior chamado Descoberta de Conhecimento em Base de Dados (KDD, em inglês

¹<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados>

Knowledge Discovery in Databases), que tem como objetivo a descoberta de novos conhecimentos a partir das seguintes etapas: seleção, pré-processamento, transformação, mineração de dados e pós-processamento [8, 9].

Derivada da Mineração de Dados, a Mineração de Dados Educacionais é uma área que explora dados gerados em ambientes educacionais, visando obter conhecimento sobre os alunos e sobre a aprendizagem deles, possibilitando a descoberta de padrões [10, 11].

Existem várias subáreas da Mineração de Dados Educacionais, uma delas é a Destilação de Dados para Julgamento Humano. Essa subárea visa apresentar os dados de forma legível, interativa e bastante visual, para facilitar a compreensão humana e a identificação de padrões, possibilitando que as pessoas façam inferências e tomem decisões com base nesses dados [8, 10, 12].

Os dados são destilados com dois propósitos: identificação e classificação. Na identificação os dados são exibidos de forma que um humano identifique facilmente padrões conhecidos, mas que são difíceis de expressar formalmente. Quando o propósito é a classificação, os dados são destilados para ajudar e acelerar o desenvolvimento futuro de modelos de previsão [6].

3 MÉTODO

Neste artigo, a técnica de Destilação de Dados para Julgamento Humano foi utilizada para identificar o perfil socioeconômico dos discentes de computação da região sul do Brasil. A destilação, com o propósito de identificação, visa apresentar padrões em formas de gráficos e estatísticas. Esse método segue as mesmas etapas do KDD.

Como ferramenta para a análise dos dados, foi escrito um código² na linguagem de programação Python. Para ler os arquivos da base de dados e executar as análises, foram utilizadas as bibliotecas Pandas e Matplotlib, do Python.

3.1 Seleção

A etapa de seleção serve para eleger apenas as variáveis necessárias de acordo com a informação que se deseja extrair, com a intenção de diminuir o conjunto de dados para melhorar o custo do processamento computacional [8].

O interesse nesse artigo não está apenas nos resultados do Enade, mas, principalmente, nos dados dos participantes. Então, foi coletada a base de microdados do Enade de 2017, disponível no portal do Inep³. Os arquivos utilizados dessa base foram o "Dicionário de variáveis dos Microdados do Enade_Edição 2017.ods" e o "MICRODADOS_ENADE_2017.txt" contendo os dados de cada participante.

Para esse estudo, investigou-se vinte variáveis. Aquelas referentes às IES e ao curso foram a modalidade de ensino, categoria administrativa da IES, código do curso no Enade, área de enquadramento do curso, região de funcionamento do curso e UF de funcionamento do curso. Entre as variáveis referentes ao estudante, as selecionadas foram o sexo, idade, ano de conclusão do Ensino Médio, ano de início da graduação e turno da graduação. Também foram selecionadas uma variável referente ao tipo de presença no Enade e uma referente a nota geral no Enade. Além disso, algumas variáveis foram extraídas do questionário do estudante, como a

cor/raça, escolaridade do pai e da mãe, renda familiar, recebimento de bolsa acadêmica, uso de inclusão social para ingressar e tipo de escola no Ensino Médio.

3.2 Pré-processamento

Nesta etapa, os dados são adequados para a análise que será feita [8], isso envolve filtrar os dados necessários para o objetivo, remover ou preencher valores vazios, além de modificar tipos e formatos de dados [9, 13].

Foram filtrados os alunos de todos os cursos de computação da região sul, que estavam presentes na prova do Enade e que tiveram resultados válidos. Para não haver erros no código, ou influência nos resultados, foram removidos os alunos que deixaram o Questionário do Estudante em branco e também os alunos que não tinham notas registradas.

A Tabela 1 apresenta a nomenclatura adotada às áreas de enquadramento dos cursos selecionados, a quantidade de cursos em cada área, o grau do curso e quantos alunos estão em cada categoria de IES. Inicialmente eram 537.436 alunos registrados na base de dados, sendo 8.856 alunos da área de computação que estudam na região sul. Após a fase de pré-processamento restaram 7.298 alunos para serem analisados, com 71,4% deles estudando em instituições privadas, além de que entre os 313 cursos diferente, 72,2% deles são em instituições privadas.

Importante ressaltar que o Inep em seu ciclo avaliativo não elabora provas para cursos específicos, mas para áreas de enquadramento. Por exemplo, Análise e Desenvolvimento de Sistemas é uma área de enquadramento para o Enade 2017 em que os cursos com o mesmo nome ou, até mesmo, cursos nomeados de Sistemas para Internet podem ser enquadrados. O enquadramento é feito pelos procuradores institucionais baseados nos projetos pedagógicos dos cursos com as diretrizes de prova previamente publicada pelo Inep. No decorrer do texto, usa-se indistintamente o termo curso ao se referir a área de enquadramento.

3.3 Transformação

A etapa de transformação tem o objetivo de melhorar os resultados e a visualização dos dados [13]. Nela os dados são modificados para ficar em um formato diferente do original [8], de acordo com o desejado.

A categoria administrativa da IES, a renda familiar, a idade e o tipo de escola do ensino médio foram as variáveis que sofreram transformação. Essa transformação é necessária para facilitar o entendimento dos dados apresentados. Para isso as alternativas são agrupadas de acordo com características em comum ou colocadas em intervalos de valores.

4 RESULTADOS E DISCUSSÃO

Nesta seção, serão apresentados os resultados da mineração de dados do Enade, a fim de responder a questão de pesquisa e estabelecer o perfil do egresso dos cursos de computação do sul do país.

4.1 Cor ou raça

Após o processo de mineração no conjunto de dados, foi descoberta que a raça branca é a mais comum declarada entre os participantes, representando 80,5% deles, enquanto 11% declararam ser pardos,

²Repositório no Github [URL omitida para manter anonimato]

³Microdados Inep <http://portal.inep.gov.br/microdados>

Tabela 1: Cursos selecionados e suas respectivas quantidades de alunos por categoria de IES.

Sigla	Área de Enquadramento	Grau	# Cursos	# Alunos Pública	#Alunos Privada
BCC	Ciência da Computação	Bacharelado	65	776	811
LCC	Ciência da Computação	Licenciatura	9	67	17
EC	Engenharia da Computação	Bacharelado	24	330	255
SI	Sistemas de Informação	Bacharelado	87	384	1443
ADS	Análise e Desenvolvimento de Sistemas	Tecnologia	86	425	1659
GTI	Gestão da Tecnologia da Informação	Tecnologia	18	40	680
RC	Redes de Computadores	Tecnologia	24	65	346

3,2% pretos, 1,9% amarelos, 0,1% indígenas e 3,3% não declararam sua raça ou cor. Analisando a Figura 1, observa-se que em todos os cursos esse padrão se repete, com a predominância de alunos brancos. Este resultado corrobora com Oliven [14], ao afirmar que mesmo o Brasil sendo um país miscigenado com leis que garantem vagas em universidades para negros e indígenas, o sistema acadêmico brasileiro apresenta um dos maiores quadros de exclusão racial.

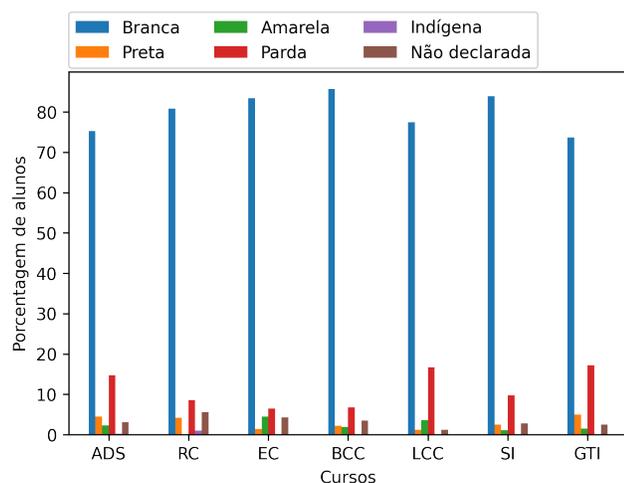


Figura 1: Cor ou raça declarados por curso.

Outra característica adicional analisada é que apenas 6,78% dos alunos declarados pretos e 3,23% dos alunos pardos utilizaram cotas étnico-raciais para ingressar nas instituições e, nenhum aluno indígena utilizou cota étnico-racial. Isso corrobora com o trabalho de Goldemberg e Durham [15] que observam que os indivíduos que fazem uso de políticas de inclusão social têm mais dificuldade em conseguir uma educação básica boa para conseguirem competir igualmente com os outros candidatos.

4.2 Sexo

Outra informação obtida na análise é que a quantidade de alunos do sexo masculino (88,98% dos participantes) prevalece em relação

a quantidade de alunas do sexo feminino na área da computação. Esse fato é observado em todos os cursos, como ilustrado na Figura 2. Uma peculiaridade é que o curso LCC é o que tem uma proporção, entre homens e mulheres, menos desequilibrada.

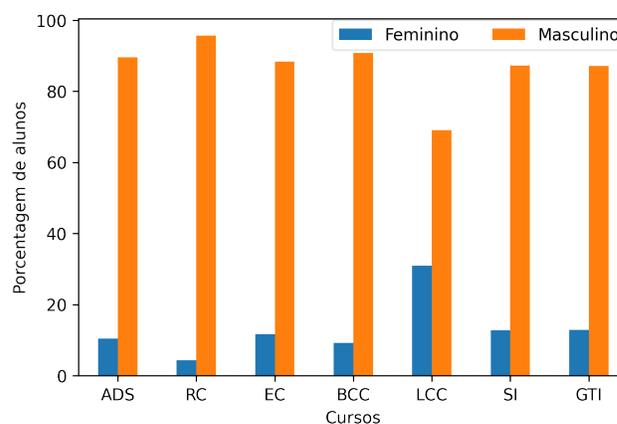


Figura 2: Porcentagem de homens e mulheres por curso.

Esses resultados evidenciam as afirmações de que, apesar de as mulheres terem um papel fundamental na história da computação, a desigualdade de gênero é cada vez mais evidente [16]. A partir da década de 90 houve um declínio do interesse de mulheres em cursos de computação [17]. Os cursos de computação são vistos como "masculinos", o que afasta as mulheres de ingressarem na área. Além disso, os homens têm preferência por cursos da área de ciências exatas, enquanto mulheres preferem a área de ciências humanas [18].

4.3 Idade

Com os dados analisados foi descoberto que a idade mais comum entre os participante é a de 22 anos e, que a maioria dos alunos, 58,46%, se formam na faixa dos 18 aos 26 anos.

Analisando a Figura 3, observa-se que em quase todos os cursos há predominância de concluintes na faixa 18-26 anos, mas os cursos de tecnologia (ADS, RC e GTI) possuem uma maior proporção de alunos acima dos 27 anos. Uma suposição que justifica esse

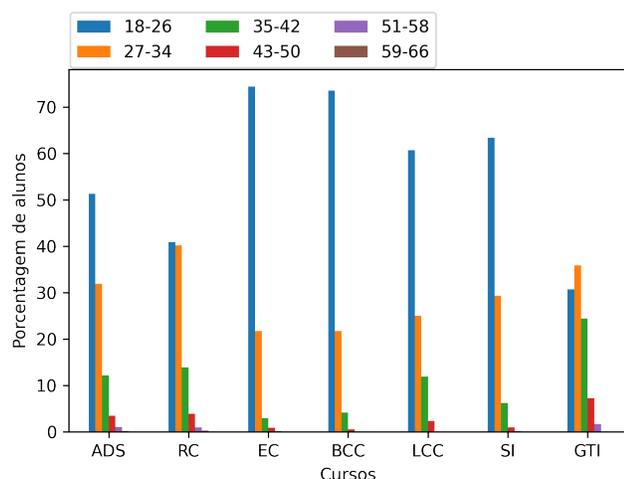


Figura 3: Idade mais comum em cada curso.

comportamento está na natureza dos cursos de tecnologia. Uma vez que eles são focados em preparar profissionais mais rapidamente para o mercado de trabalho e tenham como público alunos que já trabalham antes do ingresso é natural que a segunda faixa de idades tenha um destaque em relação aos cursos de bacharelado. Outra possível explicação para isso estaria no intervalo entre o Ensino Médio e o Ensino Superior. Em uma análise adicional, observa-se que os cursos GTI, RC e ADS são aqueles que os alunos levam mais tempo para ingressar após a conclusão do Ensino Médio, com 9, 7 e 6 anos de demora em média, respectivamente. Com uma maior demora para ingressar no Ensino Superior, os alunos acabam ingressando com uma idade superior que os alunos dos outros cursos.

Ainda sobre as tecnologias, o curso de GTI foge do padrão ao apresentar predominância de formandos na faixa de 27 a 34 anos e ser o curso com a maior porcentagem de alunos nas faixas consecutivas 35-42, 43-50, e 51-58, além de ser o curso com menor porcentagem de alunos entre 18 e 26 anos.

A Figura 4 apresenta uma análise da idade dos concluintes por estados. Observa-se que a maioria dos alunos está abaixo dos 26 anos em todos estados, mas o RS é o estado com maior representatividade de pessoas entre 27 e 34 anos. Adicionalmente, destacamos que essa representatividade maior da segunda faixa é distribuída em todos os cursos do estado. Análises específicas com outras bases de dados ainda são necessárias para entender os motivos que fazem a faixa 27-34 ter maior destaque no RS do que nos outros estados.

4.4 Bolsas acadêmicas

Entre os discentes analisados, 76,4% deles não receberam bolsa acadêmica, enquanto 9% receberam bolsa de Iniciação Científica, 3,9% receberam bolsa de extensão, 2,8% receberam bolsa de monitoria, 1,3% receberam bolsa PET e 6,6% receberam outro tipo de bolsa.

Com base na Figura 5, verifica-se que os cursos de bacharelado e licenciatura possuem uma maior proporção de alunos que receberam bolsas acadêmicas em comparação aos outros cursos.

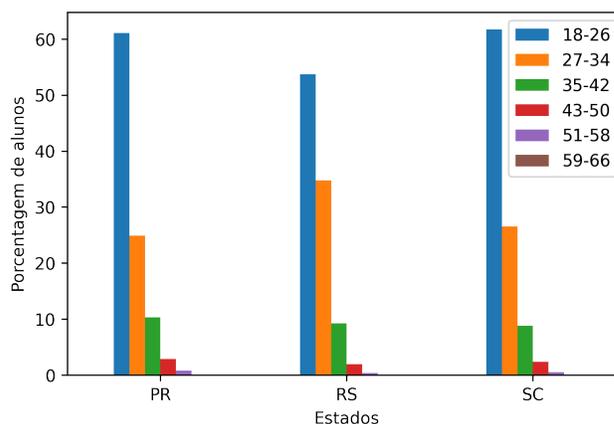


Figura 4: Idade mais comum em cada estado.

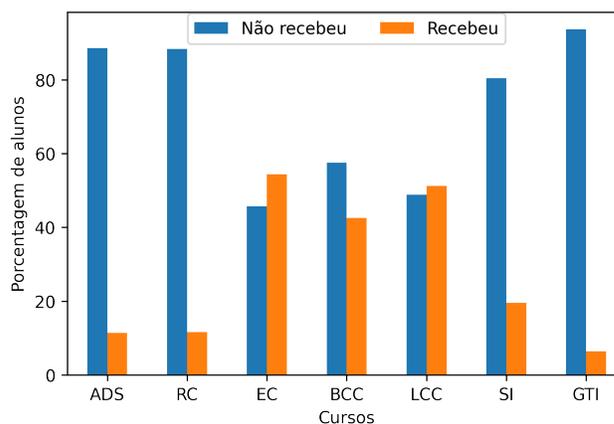


Figura 5: Concluintes em relação a bolsas acadêmicas por curso.

Entre os bolsistas, observa-se na Figura 6 que os cursos de bacharelado são os que tem maior presença de alunos com bolsa de IC e de PET, enquanto nos cursos de tecnologia os alunos declararam ter recebido outros tipos de bolsa.

Observando a Tabela 1 e a Figura 7, tem-se que os alunos que declararam ter recebido outro tipo de bolsa concentram-se de instituições privadas de ensino. Isto pode indicar que bolsas de desconto em mensalidades tenham sido declaradas como bolsa acadêmica na categoria de outro tipo de bolsa.

Por fim, destaca-se ainda que os alunos dos cursos de licenciatura, provenientes de instituições públicas, também declararam receber outro tipo de bolsa. Acredita-se que essas bolsas sejam referentes a bolsas de iniciação à docência, comuns nessa modalidade de ensino.

4.5 Ensino Médio

Uma informação interessante a respeito da trajetória acadêmica dos participantes é o tipo de escola onde cursaram a maior parte do ensino médio. Verificou-se que 73% dos alunos estudaram em

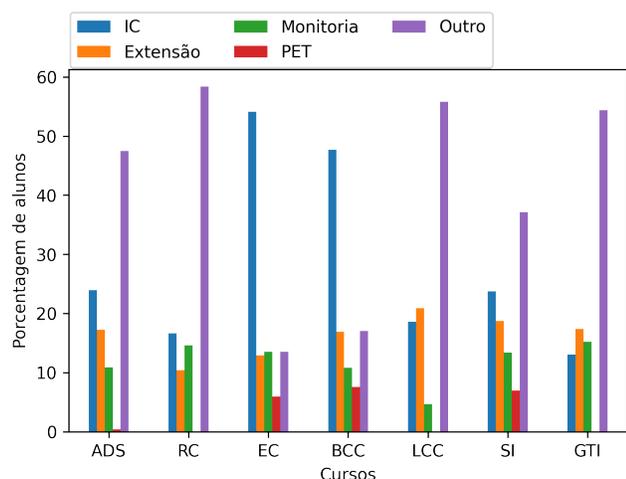


Figura 6: Tipo de bolsa por curso.

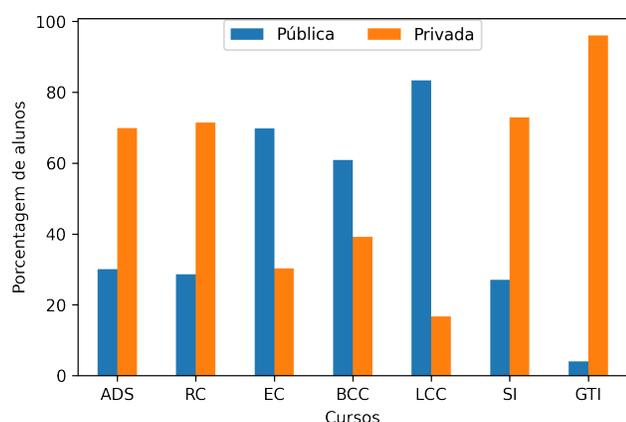


Figura 7: Participantes que receberam outro tipo de bolsa em cada categoria de IES.

escolas públicas, 26,5% em escolas privadas e 0,5% em escolas no exterior.

A Figura 8 mostra em cada tipo de escola de ensino médio, qual o tipo de IES mais optada pelos alunos. Nela vemos que independente do tipo de ensino médio, a maioria dos alunos acaba indo estudar em IES privadas, isso acontece principalmente porque há muito mais IES privadas do que públicas, de acordo com o Censo de Educação Superior de 2019⁴ 88,4% das IES do Brasil são privadas. Também é possível observar que os alunos de escolas privadas tem uma tendência maior de ir estudar em Instituições públicas de Ensino Superior do que os alunos de escolas públicas.

⁴Notas Estatísticas - Censo de Educação Superior 2019 http://download.inep.gov.br/educacao_superior/censo_superior/documentos/2020/Notas_Estatisticas_Censo_da_Educacao_Superior_2019.pdf

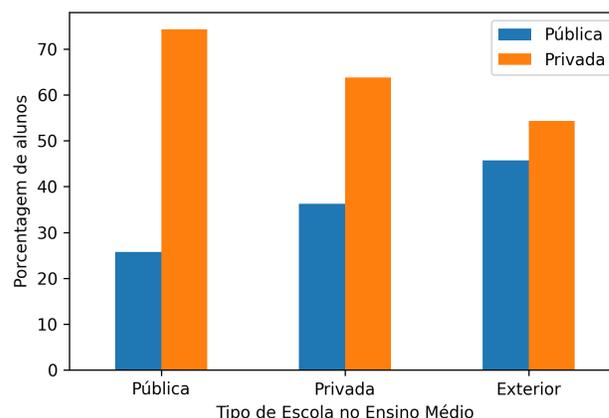


Figura 8: Ensino Médio em cada categoria de IES.

4.6 Modalidade de ensino

Em relação à modalidade dos cursos, observou-se que 83,79% dos alunos fizeram cursos presenciais e também que, entre os cursos das áreas de computação analisados, 14 deles são EAD enquanto 299 são presenciais. A Figura 9 mostra a predominância dos formandos em cursos presenciais e que os alunos do EAD estão principalmente em cursos de tecnologia (ADS e GTI).

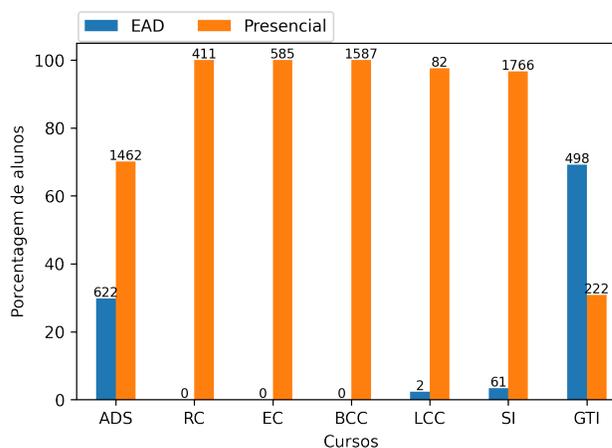


Figura 9: Alunos por modalidade e curso.

Na Figura 10, observa-se que o PR tem a maior parcela dos egressos em EAD; entretanto, todos os estados concentram-se no ensino presencial.

4.7 Turno do curso

Com base nos dados analisados, identificou-se que 75,95% dos alunos frequentou o turno noturno seguido por 17,63% dos alunos no turno integral. Já o turno matutino conta com 5,93% e, em último, o turno vespertino com apenas 0,49% dos alunos. Ainda, a Figura 11 mostra que essa distribuição ocorre similarmente em todos os estados.

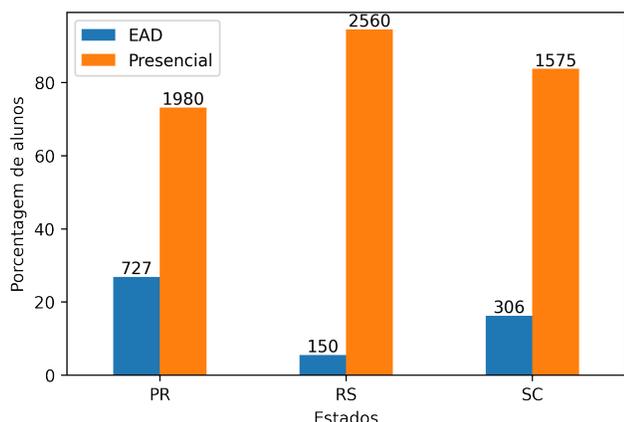


Figura 10: Alunos por modalidade e estado.

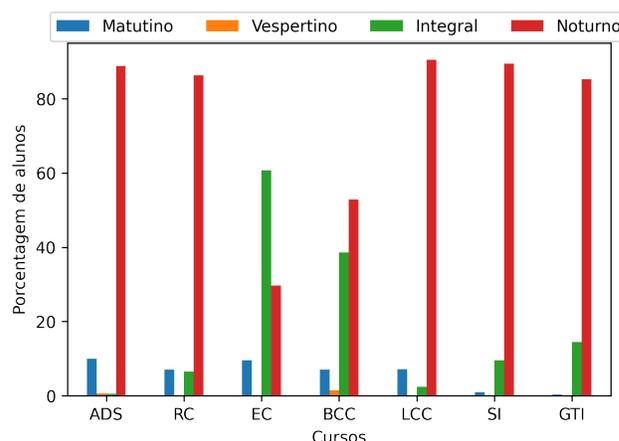


Figura 12: Alunos por turno por curso.

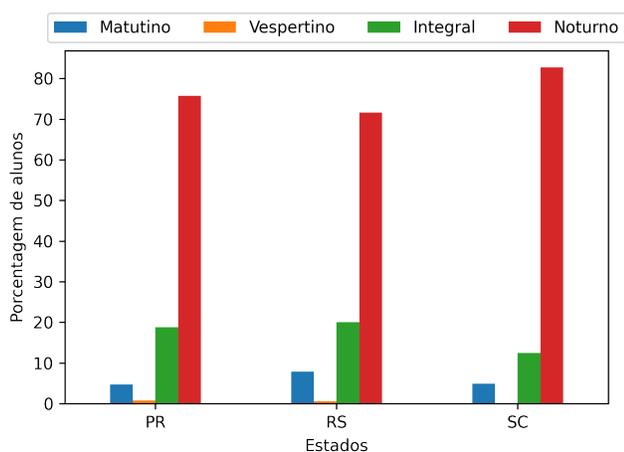


Figura 11: Alunos por turno por estado.

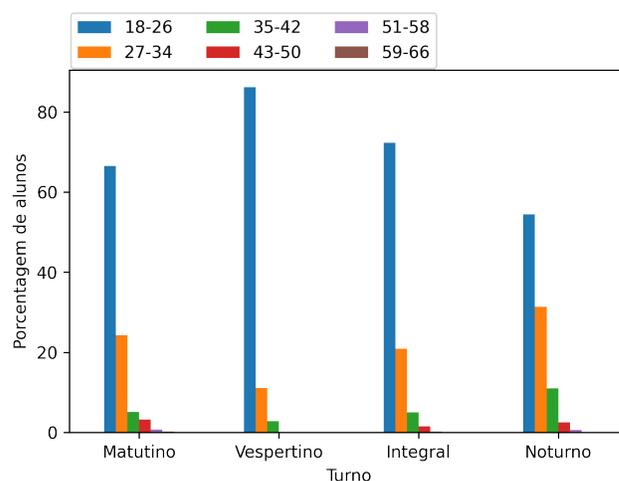


Figura 13: Faixa etária em cada turno de curso.

A Figura 12 indica que a maioria dos alunos é oriundo de cursos noturnos. Entretanto, observa-se uma esperada discrepância em EC e BCC, com maior presença de alunos em cursos de período integral do que em outros cursos.

Como apresentado na Figura 13, em todos turnos a idade predominante esta na faixa entre 18 e 24 anos, mas o turno noturno possui a maior proporção de alunos acima de 24 anos. Isso pode ser explicado pelo fato de que cursos noturno concentram pessoas mais velhas e que trabalham durante o dia. Em uma análise complementar, identificou-se que 87,07% dos alunos nos cursos noturnos já trabalham e 84,66% das pessoas nos cursos noturnos estudam em instituições privadas, o que pode significar que são alunos que trabalham para pagar seus cursos.

4.8 Renda familiar

Dentre todos os participantes, 30,8% têm renda familiar de até 3 salários mínimos, 42,1% têm renda familiar de 3 a 6 salários mínimos, 26,5% têm renda familiar de 6 a 30 salários mínimos, e 0,6% têm renda familiar acima de 30 salários mínimos. A Figura 14 demonstra

que em todos estados a maioria dos alunos possuem uma renda familiar entre 3 e 6 salários mínimos.

A Figura 15 ilustra a relação entre a nota geral dos participantes e sua renda familiar. A distribuição é semelhante entre os estados da região sul. Em geral, conforme a renda familiar aumenta a nota geral também aumenta.

Isso vai de encontro com a informação de que os recursos financeiros que os pais têm disponíveis para os gastos educacionais dos filhos afetam o desempenho do aluno [19]. Alunos com maior renda familiar tendem a estudar em melhores escolas, além de conseguirem fazer cursos extracurriculares [20]. Sendo assim, alunos com maior renda familiar tem uma base educacional melhor, e por consequência, acabam ingressando mais no Ensino Superior, e se destacando como melhores alunos.

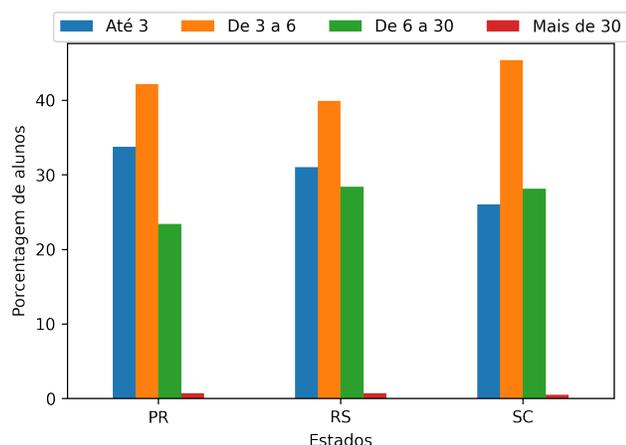


Figura 14: Renda dos participantes em cada estado.

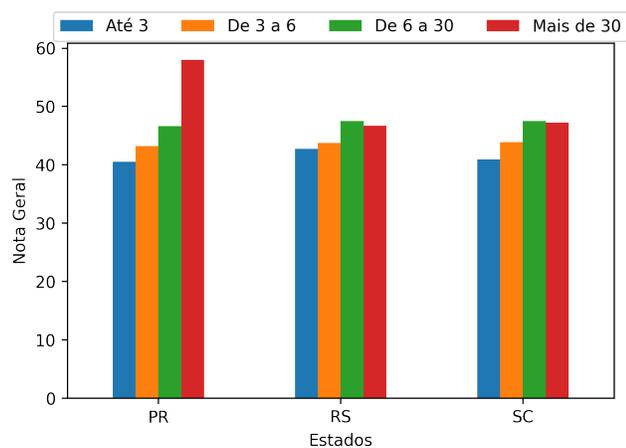


Figura 15: Nota Geral de cada renda familiar nos estados.

4.9 Escolaridade dos pais

A partir da análise das figuras 16 e 17 percebe-se que, em todos os cursos, tanto os pais quanto as mães em sua maioria estudaram até o Ensino Médio, e que a escolaridade dos pais é mais baixa que a das mães, com a porcentagem de mães que possuem pós-graduação sendo maior que a porcentagem de pais, enquanto a porcentagem de pais com nenhuma escolaridade é maior que a das mães. Além disso, também percebe-se nas figuras que alunos dos cursos de bacharelado tem pais e mães com maior nível de escolaridade.

As figuras 18 e 19 mostram que a escolaridade dos genitores influenciam na nota do aluno no Enade. Isto é, quanto maior a escolaridade dos pais, maior a nota do aluno no Enade.

Segundo Moreira *et al.*[21], vários estudos apontam que o *background* familiar (escolaridade dos pais e renda familiar) afeta a educação e o desempenho dos filhos em exames educacionais, quanto maior a escolaridade dos pais e maior a renda familiar melhor é o desempenho do estudante. Isso se deve ao fato que famílias com maior

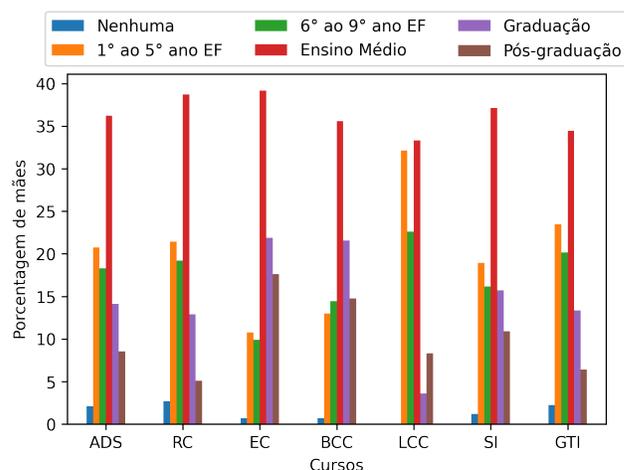


Figura 16: Nível de escolaridade das mães.

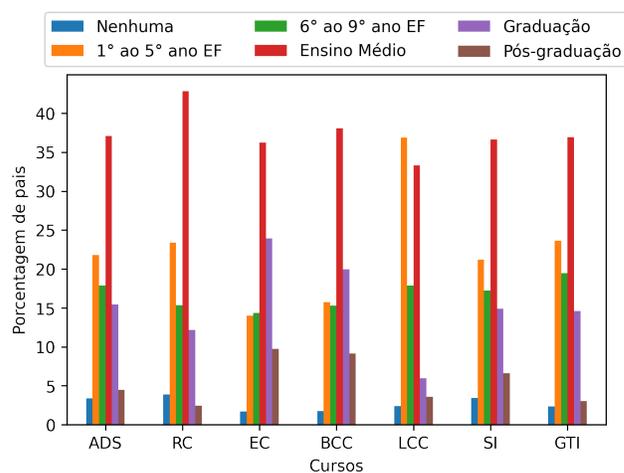


Figura 17: Nível de escolaridade dos pais.

poder aquisitivo investem mais na educação dos filhos, provendo melhores insumos escolares e acesso a melhores professores.

5 CONCLUSÕES

Os resultados contribuíram de forma direta para alcançar o objetivo de detectar o perfil socioeconômico dos concluintes dos cursos de computação da região sul do Brasil, trazendo diversidade de informações e importantes discussões.

Em relação às descobertas, percebe-se que há predominância da raça branca em toda a área de computação, mesmo havendo políticas de inclusão social. Observou-se que o sexo masculino é majoritário em todos os cursos. Também identificou-se que a renda familiar e a escolaridade dos pais influenciam na vida acadêmica e na nota no Enade do participante; e que os alunos de escola privada tem uma tendência maior de ingressar em instituições públicas.

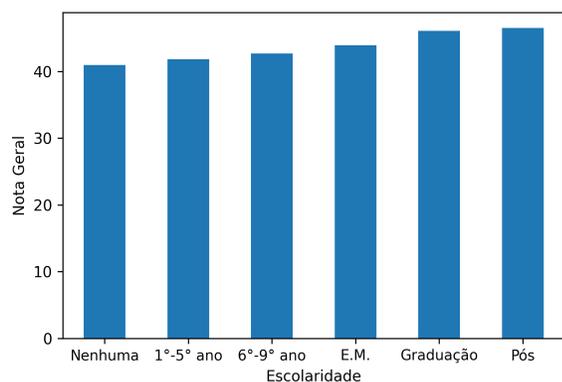


Figura 18: Nota Geral de acordo com a escolaridade das mães.

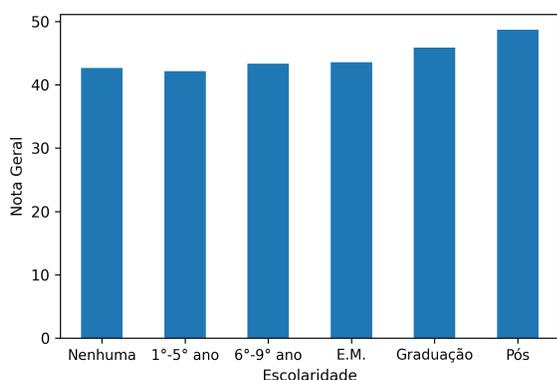


Figura 19: Nota Geral de acordo com a escolaridade dos pais.

Essas informações obtidas por meio da mineração de dados descrevem de forma sintética que o perfil socioeconômico do discente de computação da região sul do Brasil a partir dos dados do Enade 2017 é de uma pessoa branca, do sexo masculino, com idade entre 18 e 26 anos, que não recebe bolsa acadêmica, seus pais estudaram até o Ensino Médio e tem renda familiar de 3 a 6 salários mínimos. Esse aluno faz curso presencial noturno em uma instituição privada e estudou em escola pública durante o ensino médio.

Os dados fornecidos pelo INEP por meio do Enade e do Censo do Ensino Superior ainda precisam ser melhor explorados pela comunidade acadêmica. Há uma vasta variedade de questões em pesquisas em aberto que, se corretamente investigadas, podem contribuir significativamente para a melhoria do ensino superior brasileiro. Como trabalho futuro, por exemplo, pretende-se entender a evolução ao longo dos anos do perfil socioeconômico dos alunos de computação, além de mapear quais variáveis influenciam no desempenho dos alunos no Enade.

AGRADECIMENTOS

Agradecemos as valiosas contribuições do Prof. Dr. Marco Aurélio Graciotto Silva.

REFERÊNCIAS

- [1] Roseli Constantino Schwerz, Natalia Neves Macedo Deimling, Cesar Vanderlei Deimling, and Daniele Cristina da Silva. Considerações sobre os indicadores de formação docente no Brasil. *Pro-Posições*, 31, 2020. doi: 10.1590/1980-6248-2017-0199.
- [2] Sandro J. Rigo, Silvio C. Cazella, and Wagner Cambruzzi. Minerando dados educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. In *Anais do Workshop de Desafios da Computação Aplicada à Educação, DesafIE! 2012*, pages 168–177, Porto Alegre, RS, BR, 2012. SBC.
- [3] Paulo J.L. Adeodato, Mailson M. Santos Filho, and Rodrigo L. Rodrigues. Predição de desempenho de escolas privadas usando o enem como indicador de qualidade escolar. In *Simpósio Brasileiro de Informática na Educação, SBIE'2014*, pages 891–895, Porto Alegre, RS, BR, 2014. SBC. doi: 10.5753/cbie.sbie.2014.891.
- [4] Philippe Santos, Rafael Ferreira, and Péricles Miranda. Dados abertos educacionais: Uma revisão da literatura brasileira. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28 of *SBIE 2017*, page 11, Porto Alegre, RS, BR, 2017. SBC. doi: 10.5753/cbie.sbie.2017.11.
- [5] Priscila da Silva Neves Lima, Ana Paula Laboissière Ambrósio, Deller James Ferreira, and Jacques Duílio Brancher. Análise de dados do enade e enem: uma revisão sistemática da literatura. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 24(1):89–107, 2019. doi: 10.1590/s1414-40772019000100006.
- [6] Ryan S. Baker. Data mining for education. *International encyclopedia of education*, 7(3):112–118, 2010.
- [7] Cristobal Romero and Sebastian Ventura. Data mining in education. *WIREs Data Mining and Knowledge Discovery*, 3(1):12–27, 2013. doi: 10.1002/widm.1075.
- [8] Evandro Costa, Ryan Baker, Lucas Amorim, Jonathas Magalhães, and Tarsis Marinho. Mineração de dados educacionais: Conceitos, técnicas, ferramentas e aplicações. *Jornada de Atualização em Informática na Educação*, 1(1):1–29, 2013.
- [9] Cássio Oliveira Camilo and João Carlos da Silva. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. Technical report, Universidade Federal de Goiás, Goiânia, GO, BR, 2009.
- [10] Ryan Baker, Seiji Isotani, and Adriana Carvalho. Mineração de dados educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, 19(02):03, 2011. doi: 10.5753/rbie.2011.19.02.03.
- [11] Leandro Silva, Ismar Silveira, Luciano Silva, Rodrigo Rodrigues, and Jorge Ramos. Ciência de dados educacionais: definições e convergências entre as áreas de pesquisa. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 6 of *WCBIE'2017*, page 764, Porto Alegre, RS, BR, 2017. SBC. doi: 10.5753/cbie.wcbie.2017.764.
- [12] Leandro A. Silva and Luciano Silva. Fundamentos de mineração de dados educacionais. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 3 of *WCBIE'2014*, page 568, Porto Alegre, RS, BR, 2015. SBC. doi: 10.5753/cbie.wcbie.2014.568.
- [13] Nicollas Nogueira Cretton and Georgia Rodrigues Gomes. Aplicação de técnicas de mineração de dados na base de dados do enade com enfoque nos cursos de medicina. *Acta Biomedica Brasiliensis*, 7(1):74–89, 2016. doi: 10.18571/acbm.100.
- [14] Arabela Campos Oliven. Ações afirmativas, relações raciais e política de cotas nas universidades: Uma comparação entre os estados unidos e o Brasil. *Educação*, 30(61):29–51, 2007.
- [15] José Goldemberg and Eunice R. Durham. *Cotas nas universidades públicas*, pages 167–172. Civilização Brasileira, Rio de Janeiro, RJ, BR, 2007.
- [16] Wilk Santos. Mulheres na computação: Uma análise da participação feminina nos cursos de licenciatura em computação. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação, WCBIE'2017*, page 814, Porto Alegre, RS, BR, 2017. SBC. doi: 10.5753/cbie.wcbie.2017.814.
- [17] Carolina Santana Louzada, Wesckley Faria Gomes, Maria Augusta Silveira Netto Nunes, Edilayne Meneses Salgueiro, Beatriz Trinchão Andrade, and Patricia Soares de Lima. Um mapeamento das publicações sobre o ingresso das mulheres na computação. In *Conferência Latino-americana em Informática-VI Congresso da Mulher Latino-americana na Computação, CLEI'2014*, page 16, Montevideo, Uruguay, 2014. IEEE.
- [18] Prícila Castellini and Marília Abrahão Amaral. Mulheres na computação: que cursos queremos. In *Anais eletrônicos do Seminário Internacional Fazendo Gênero & Women's Worlds Congress*, volume 11 of *WW'2017*, page 12, Florianópolis, SC, BR, 2017. UFSC.
- [19] Juliana de Lucena Ruas Riani and Eduardo Luiz Gonçalves Rios-Neto. Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros? *Revista Brasileira de Estudos de População*, 25(2):251–269, 2008. doi: 10.1590/S0102-30982008000200004.
- [20] Brenda Cristina de Oliveira Rodrigues, Marise Santana Resende, Gilberto José Miranda, and Janser Moura Pereira. Determinantes do desempenho acadêmico dos alunos dos cursos de ciências contábeis no ensino a distância. *Enfoque: Reflexão Contábil*, 35(2):139–153, 2016. doi: 10.4025/enfoque.v35i2.30105.
- [21] Romilson do Carmo Moreira. *Três ensaios sobre educação, desigualdades raciais e políticas de ações afirmativas*. PhD thesis, PUC-RS, Porto Alegre, RS, BR, 2019.