

# Análise exploratória sobre evasão tardia da graduação de uma universidade pública

Caio Vinicius Monteiro Martins<sup>1</sup>, Filipe Carvalho Lacerda<sup>1</sup>, Igor Procópio do Carmo<sup>1</sup>, Edmilson Vitorino Scovino da Silva<sup>1</sup>, Tatiane Ornelas Martins Alves<sup>1</sup>, Ricardo Silva Campos<sup>1,2</sup>

<sup>1</sup>Instituto Metodista Granbery – Bacharelado em Sistemas de Informação

<sup>2</sup>Universidade Federal de Juiz de Fora – CGCO

ricardo.campos@ufjf.br

**Abstract.** *Student dropout in the final periods is a serious problem for higher education institutions due to factors such as the budget received by universities, which depend on the number of students enrolled and graduating, as well as losses in the loss of programs and government plans. Thus, this work performs an exploratory analysis on the database of an academic management system of a federal university, to describe the main factors that cause dropout at different stages. Descriptive statistics and correlation techniques were applied.*

**Resumo.** *A evasão de alunos nos períodos finais é um sério problema para instituições de ensino superior por fatores como o orçamento recebido pelas universidades, que dependem da quantidade de alunos matriculados e de concluintes, assim como prejuízos na perda de programas e planos do governo. Assim, este trabalho realiza uma análise exploratória na base de dados de um sistema de gestão acadêmica de uma universidade federal, para descrever os principais fatores que causam a evasão estudantil em diferentes fases da graduação. Foram aplicadas técnicas de estatística descritiva e correlação.*

## 1. Introdução

A Universidade Federal de Juiz de Fora (UFJF) possui um campus sede localizado em Juiz de Fora - MG e um campus avançado em Governador Valadares - MG. A UFJF oferta cerca de 93 cursos de graduação, 45 cursos de mestrado e 24 de doutorado, em todas as áreas do conhecimento. Ao todo, possui mais de 20 mil alunos na modalidade presencial e cerca de três mil na modalidade à distância.

Para administrar suas informações institucionais, a UFJF tem implementado desde 2003 um sistema próprio, chamado de Sistema Integrado de Gestão Acadêmica (SIGA). Ele foi desenvolvido completamente em software livre, reduzindo custos e promovendo o alinhamento estratégico com o Governo Federal. O SIGA é composto atualmente por 43 módulos, distribuídos em quatro grandes áreas ou sistemas: Acadêmico, Administrativo, Gestão de Pessoas, e Sistemas de Apoio.

A evasão de universitários pode causar prejuízos significativos para a economia, para a gestão institucional e para diversos aspectos sociais. A verba recebida pelas universidades depende da quantidade de alunos matriculados e concluintes. Quando não atingem o objetivo institucional de formar os alunos, as instituições de ensino públicas sofrem prejuízos como a perda de programas e planos fornecidos pelas instituições de

ensino superior (IES), além disso os professores são prejudicados por não conseguirem exercer seu papel [PRESTES, 2014].

A evasão de alunos, principalmente os que estão nos períodos finais, é um grande problema para as escolas, universidades federais, estaduais e privadas, de acordo com Santana (1996), a evasão é um dos mais sérios desafios do sistema educacional, porque causa desequilíbrio, desarmonia e desajustes dos objetivos educacionais.

De acordo com De Oliveira Junior (2017), é importante utilizar técnicas estatística e de aprendizagem de máquina para identificar os atributos mais relevantes para a evasão. Esses atributos podem incluir características socioeconômicas dos estudantes, desempenho acadêmico, dados demográficos e outros fatores relevantes.

Desta forma, o trabalho tem como objetivo realizar uma análise exploratória nos dados de todos os alunos de graduação presencial de uma universidade pública federal, entre os anos de 2003-2020. O intuito é identificar os perfis de alunos universitários no final da graduação que evadem de seus cursos, destacando-se os principais fatores socioeconômicos associados à desistência. Através da análise, pretende-se fornecer informações benéficas à instituição com o fim de buscar pelos padrões mais recorrentes nos aspectos sociais para mapear os grupos com maior risco de evasão, fomentando discussões para auxiliar a tomada de decisões.

Este trabalho possui objetivo descritivo e de procedimento *ex post facto*, com natureza aplicada e abordagem quantitativa. O método consiste na obtenção e preparação dos dados, aplicação de correlação e discussão das variáveis. Além da análise na base com todos os estudantes, foi feita uma análise com alunos nos períodos finais, de maneira a se avaliar separadamente os fatores socioeconômicos que levam à evasão tardia. Assim, pretende-se diferenciar as diferentes causas de ambas.

## **2. Revisão de literatura**

Segundo Vasquez et al. (2003), a evasão é uma situação enfrentada por um estudante que deseja, mas não consegue alcançar êxito em seu projeto educacional. Em termos mais práticos, considera-se como evadido o aluno de uma instituição de ensino superior que não apresenta atividades acadêmicas por três semestre consecutivos. Sendo assim, é possível classificar a evasão estudantil sob duas perspectivas: tempo e espaço. A evasão temporal é separada em três categorias devido a causalidades diferentes: evasão precoce: indivíduo que foi aprovado no curso, porém não realizou a matrícula; evasão inicial: aluno que evade do curso nos primeiros quatro períodos e evasão tardia: aluno que evade a partir do quinto período.

Por outro lado, ainda segundo Vasquez et al. (2003) a evasão espacial se divide em: deserção institucional, que ocorre quando o estudante abandona a universidade, evasão interna, que se dá quando o estudante trocar seu programa acadêmico por outro oferecido dentro da mesma instituição de ensino, e evasão do sistema educacional, em que o estudante interrompe os estudos.

No Brasil, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), vinculado ao Ministério da Educação, é responsável pela realização anual do censo escolar da educação superior, e por conseguinte, é responsável pelo

levantamento da evasão. O INEP considera evasão como a saída antecipada, antes da conclusão, por desistência independentemente do motivo. Além disto, em seus indicadores, o censo considera como desistentes todos aqueles estudantes que encerraram seu vínculo com o curso, por transferência para outro curso, abandono ou desligamento. Ou seja, os indicadores de evasão do censo abrangem os três tipos de evasão espacial apontados por Vasquez et al. (2003).

De acordo com Brito Junior (2018), o impacto gerado pela evasão nos cursos de graduação é significativo, visto que esse fenômeno atrapalha o desenvolvimento do mercado de trabalho criando um déficit de mão de obra qualificada para os setores com maior evasão de cursos superiores, bem como uma diminuição da verba arrecadada tanto pelas universidades públicas quanto pelas privadas a cada aluno evadido. Assim, Pereira (2003) conclui que a evasão gera prejuízos financeiros para os setores públicos e privados.

Segundo o Mapa do Ensino Superior do Instituto Semesp<sup>1</sup>, os índices de evasão em 2019 foram de 30,7% na rede privada e 18,4% na rede pública. Entre cursos presenciais, o maior percentual de evasão é 37,6% em Sistemas de Informação, seguido de 35,9% em Administração. Em cursos à distância, as maiores evasões ocorrem em Marketing (44,7%) e Matemática Formação de Professor (44,3%).

Já foram realizados diversos estudos relacionados para buscar entender os motivos mais frequentes que levam os alunos a evadirem de seus cursos durante sua graduação no Brasil. Buscando informações que possam levar a uma estimativa para contribuir com a resolução deste problema. Por exemplo, Brito Junior (2018) desenvolveu um estudo com o uso da mineração de dados para identificar os perfis dos graduandos que evadem do curso de sistemas de informação, utilizando os dados da UFRN.

Couto e Santana (2017) discutiram a aplicação de técnicas de mineração de dados educacionais para identificar fatores associados à evasão em um curso de Engenharia de Produção e utilizaram técnicas de classificação e associação para identificar as variáveis mais relevantes para a previsão da evasão. Os resultados indicaram que as variáveis mais importantes para a evasão foram o desempenho acadêmico, a carga horária de trabalho e a distância entre a casa e a universidade.

Neste artigo, utilizaremos uma base de dados de uma universidade federal para desenvolver um estudo semelhante aos citados acima, porém além de mensurar os alunos que evadem de seus cursos com suas informações socioeconômicas, iremos avaliar em separado os alunos que alcançam os períodos finais, para se avaliar a evasão tardia.

### **3. Métodos**

Os microdados de estudantes da Universidade Federal de Juiz de Fora são públicos e foram obtidos através de uma solicitação de informação no portal governamental FalaBR<sup>2</sup>, com o número 23546.051594/2020-15. Foram solicitados os dados de estudantes com matrículas válidas de graduação da UFJF de 2003 a 2020, na modalidade presencial. Os dados foram disponibilizados de maneira anônima para que não fossem violadas a privacidade e intimidade da comunidade acadêmica, conforme parâmetros da Lei de Acesso à Informação.

---

<sup>1</sup> Disponível em: <<https://www.semesp.org.br/mapa/edicao-11/brasil/evasao/>> Acesso em 07/7/2022.

<sup>2</sup> Disponível em: <<https://falabr.cgu.gov.br>> Acesso em 07/7/2022.

Nestes dados está representada a situação do vínculo do aluno a um curso de graduação, ou seja, a medida de evasão é similar àquela empregada pelo censo do INEP, o que abrange as evasões internas, institucionais e do sistema educacional, no que se refere à espacialidade. Em relação à temporalidade da evasão, tem-se apenas a evasão cometida após o início das atividades acadêmicas do curso. Em outras palavras, não há informações sobre a evasão precoce, porém há sobre a evasão inicial e tardia.

As variáveis disponibilizadas para o desenvolvimento deste trabalho são: **Cota:** Sistema de cotas de ingresso; **Períodos:** Número de períodos cursados; **Etnia:** Amarela; branca; indígena; parda preta; outra; **GAP:** Diferença em anos entre o fim ensino médio à entrada do ensino superior. **Grande área;** **IRA:** Índice de rendimento acadêmico, dado entre 0 e 100. **Idade de saída:** Idade ao sair do curso; **Mesma cidade:** Se o aluno faz o curso em sua cidade natal ou não; **MonTccEst:** Se o curso contém em sua grade monografia, tcc ou estágio obrigatório. **Sexo;** **Situação:** Determina se o aluno é evadido ou concluído. **Turno:** Período de realização do curso.

### 3.1. Seleção, pré-processamento e transformação

A partir desta base de dados, o projeto se inicia utilizando a linguagem de programação em Python na IDE Jupyter Notebook e a biblioteca Pandas.

É comum que *dataframes* contenham erros, como valores ausentes ou até mesmo valores errados e, portanto, é necessária uma limpeza. Desta forma, pode-se optar conforme a conveniência entre algumas possibilidades, como imputação de valores a partir de medidas de tendência central ou exclusão de registros.

Na coluna etnia, os valores ausentes foram substituídos por uma constante com valor “OUTRA”. Na coluna “GAP” que é referente ao tempo entre o término do aluno no ensino médio e a sua entrada no ensino superior em anos haviam números incoerentes, devido a este fato foi adicionado um limitador para essa coluna, no qual se houvesse um “GAP” com valores muito discrepantes seriam retirados esses registros (>60), na coluna “Idade de Saída” que é referente a idade na qual o aluno terminou o curso existiam valores negativos que também foram ignorados, por fim na coluna “Prazo Ideal” que indica o número de semestres esperado para conclusão do curso pelo aluno foram retirados registros com valores iguais a zero.

Um dos fatores centrais para a compreensão do projeto em relação a metodologia e aos resultados é a noção dos grupos de acesso prioritários utilizados no ingresso à universidade: Grupo A: baixa renda, preto, pardo ou indígena (PPI), oriundo de escola pública. Grupo B: baixa renda, oriundo de escola pública. Grupo C: ampla concorrência. Grupo D: PPI, escola pública. Grupo E: escola pública.

Com base nesses parâmetros de cotas o projeto busca uma padronização em valores ordinais para esse sistema, dessa forma desenvolveu-se um sistema como uma ordem de vulnerabilidade, na qual o maior número representa um grupo que demonstra maior vulnerabilidade social e o menor grupo a ampla concorrência, dessa forma encontra-se então: Grupo A: 5; Grupo B: 4; Grupo D: 3; Grupo E: 2; Grupo C: 2. Ao classificar os grupos desta maneira os dados se tornam então tratados para serem utilizados pelos algoritmos.

Além disto, outras colunas foram adicionadas como parte do tratamento, como a coluna “Mesma Cidade” que se refere se a cidade natal do aluno é a mesma onde é

localizada o campus da faculdade onde ele cursou sua graduação, além disso a coluna “MonTccEst” foi criada na qual seus registros iriam apresentar resultados de verdadeiro ou falso em caso de o curso apresentar monografia, trabalho de conclusão de curso ou estágio obrigatório em seu currículo. Também foi adicionada a “Percentual de Conclusão” que se refere ao percentual de períodos cursados em relação ao prazo ideal em busca de compreender se o número de períodos cursados pelo aluno causa determinado impacto na evasão e “Situação” um booleano que demonstra se o aluno em questão está em estado concluído ou evadido.

Outra coluna importante é a “IRA” que se refere ao índice de rendimento acadêmico, podendo ser descrito em 4 intervalos: Grupo 1: 0 a 25; Grupo 2: 25 a 50; Grupo 3: 50 a 75; Grupo 3: 75 a 100. Dessa forma será possível analisar se o IRA do aluno é impactante em sua decisão de evasão.

Por estar lidando com algoritmos é de suma importância transformar os dados categóricos em numéricos, desta forma é necessário tratar as variáveis categóricas nominais, como etnia e grande área, transformando os seus valores em variáveis dicotômicas, após isso temos um *dataframe* lidando exclusivamente com registros numéricos tornando viável a manipulação desses registros pelos algoritmos.

Além disto, como o foco do projeto é voltado para os alunos no final de curso que evadem, foram retirados os alunos ativos do *dataframe* tratado, pois eles não podem ser usados para a análise proposta.

Para finalizar o tratamento, o *dataframe* foi separado em dois *datasets*, um contendo todos os estudantes, chamado de geral. O outro conjunto contém estudantes que conseguiram cursar além 70% do número previsto de períodos, nomeado de finalista.

### 3.2. Correlação de Pearson

A correlação de Pearson quantifica a associação linear entre duas variáveis, ou seja, demonstra o grau e a direção de relacionamento entre elas. Assim, diz-se que duas variáveis X e Y são fortemente relacionadas se um incremento em X causa o mesmo impacto em Y. É possível mensurar essa relação de diversas maneiras. Neste trabalho foi empregada a Correlação de Pearson, que pode ser usada em diferentes aplicações de mineração de dados [ALQALLAF et al., 2002]. Ela é dada pela equação:

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

O coeficiente traz consigo duas informações: o seu sinal indica a direção do relacionamento e o módulo indica a intensidade da relação. Os coeficientes estão contidos no intervalo entre -1 e 1, sendo que se módulo igual a 1 indica a correlação perfeita. Por outro lado, o coeficiente igual a 0 indica que não há relação linear entre as variáveis. Assim, pode-se dizer que módulos abaixo de 0,5 indicam baixa correlação, e módulos mais próximos a 1 indicam relacionamentos fortes.

### 4. Resultados

O *dataset* disponibilizado possui as colunas descritas no Capítulo 3 e possui um total de 46914 registros totais após a limpeza dos dados, incluindo as situações de alunos concluídos, evadidos e ativos. Para o estudo ser realizado foi necessário filtrar o *dataset*

mantendo apenas os concluídos e evadidos totalizando 30931 registros, uma vez que os alunos ativos não seriam úteis para o estudo, vindo da necessidade de encontrar padrões entre os alunos que evadem e que concluem e as discrepâncias entre eles.

Para realizar uma comparação entre o perfil comum do aluno que evade e o perfil do aluno que evade no final da graduação, serão feitas duas análises. Uma geral, que abrange todos os estudantes da base (geral) e outra apenas com os considerados finalistas, que já cursaram 70% do número previsto de semestres.

Desta forma, temos o seguinte resultado exibido na Tabela 1. É possível perceber que dentre os alunos concluídos e evadidos (excluídos os alunos ativos) 37,7% dos alunos abandonaram seus cursos, a maioria desses alunos evadem nos primeiros períodos, porém alguns deles só deixaram de fazer parte da universidade no final (com mais de 70% do curso concluído), ou seja, entre os alunos finalistas, 14,8% evadiram.

**Tabela 1: Total de estudantes e percentual de evadidos.**

Geral		Finalista	
Qtde	Evasão	Qtde	Evasão
30.931	37,7%	22.174	14,8%

A partir deste ponto serão apresentadas as tabelas com as variáveis mais importantes da base de dados. Para cada possível valor de uma variável, serão exibidos as respectivas quantidades (frequência total) e o percentual de evasão, para a base geral e para a base de finalistas.

A primeira variável avaliada é o sexo (Tabela 2). Na base geral, constam 14.268 homens, dos quais 44,2% evadiram. Há ainda 16.663 mulheres das quais 32,1% desistiram do curso. Ou seja, além da quantidade de mulheres ser maior, a evasão entre elas é menor. Analisando a base de finalistas, percebe-se que esta tendência é mantida, com 12.521 mulheres com 11,4% de evasão e 9.653 homens com 19,3% de abandono.

**Tabela 2: Total de estudantes e percentual de evadidos por sexo.**

Sexo	Geral		Finalista	
	Qtde	Evasão	Qtde	Evasão
Masculino	14.268	44,2%	9.653	19,3%
Feminino	16.663	32,1%	12.521	11,4%

Na Tabela 3 são comparados o índice de rendimento acadêmico (IRA) dos alunos. Na base geral, percebe-se que a faixa com menor desempenho possui evasão total. Ou seja, entre os estudantes com rendimento inferior a 25%, todos evadem. Este fato era esperado, uma vez que um desempenho tão baixo implica em notas insuficientes para que o estudante seja aprovado nas disciplinas. Devido à retenção sofrida, o estudante tende, portanto, a abandonar o curso. Por outro lado, percebe-se, que quanto maior o rendimento, menor a evasão. Além disto, nota-se que o maior quantitativo de estudantes está na faixa superior de desempenho. Ou seja, cerca de 52% da base geral e 65% dos finalistas possui rendimento maior que 75%.

**Tabela 3: Total de estudantes e evadidos por rendimento acadêmico.**

IRA	Geral		Finalista	
	Qtde	Evasão	Qtde	Evasão
75-100	16.010	12,8%	14.327	4,9%

50-75	8.434	38,9%	6.224	18,2%
25-50	2.943	94,6%	1.277	87,5%
0-25	3.544	100,0%	346	100,0%

A Tabela 4 exibe os quantitativos e os percentuais de evasão por área do conhecimento. Percebe-se que os estudantes da área de ciências da saúde possuem a menor porcentagem de evasão em ambas as bases: 27,9% na geral e 6,3% nos finalistas. Também possuem os maiores quantitativos com 9.086 e 6.965, em torno de 27,5% e 28,8% do total. Isto é reflexo do grande número absoluto de estudantes em cursos de Saúde na UFJF, além de serem cursos com seleção competitiva, o que pode influenciar positivamente no desempenho e na permanência destes estudantes [O'NEILL, 2013].

Por outro lado, é possível perceber que os alunos dos cursos da área de Engenharia têm o maior quantitativo de evasões na base de finalistas, totalizando 834 evasões, seguido das Humanidades com 738. As Ciências Exatas e da Terra possuem os maiores percentuais de evasão, porém os quantitativos são bem menores que as demais áreas.

**Tabela 4: Total de estudantes evadidos por áreas de conhecimento.**

Grande área	Geral		Finalista	
	Qtde	Evasão	Qtde	Evasão
Ciências Biológicas	534	38,2%	407	19,7%
Ciências da Saúde	9.086	27,9%	6.965	6,3%
Ciências Exatas e da Terra	1.158	73,7%	536	44,8%
Ciências Humanas	4.353	48,0%	2.794	26,4%
Ciências Sociais Aplicadas	8.511	31,3%	6.382	9,5%
Engenharias	5.278	46,6%	3.594	23,2%
Linguística, Letras e Artes	2.011	42,8%	1.496	23,5%

A distribuição de estudantes por cotas pode ser encontrada na Tabela 5. A maior frequência ocorre em ambas as bases na ampla concorrência (Grupo C) com 20.083 e 14.707 estudantes. Isto ocorre principalmente porque esta modalidade de acesso existe durante todo o período de construção da base de dados, iniciado em 2003. Os demais grupos foram criados após 2013, apenas. Além disto, ao grupo C normalmente são destinadas 50% das vagas, enquanto os demais grupos se dividem entre a outra metade das vagas. Ademais, não se percebe uma relação proporcional entre o percentual de evasão e a vulnerabilidade socioeconômica do estudante cotista. Afinal, os grupos A e B, considerados mais vulneráveis, não possuem os maiores percentuais de evasão, chegando, em alguns casos, a evadirem menos que os estudantes da ampla concorrência.

**Tabela 5: Total de estudantes e percentual de evasão por cota.**

Cotas	Geral		Finalista	
	Qtde	Evasão	Qtde	Evasão
Grupo A	2.681	45,5%	1.809	20,1%
Grupo B	5.474	31,6%	4.410	15,9%
Grupo C	20.083	35,4%	14.707	13,8%
Grupo D	1.502	61,0%	673	16,2%
Grupo E	1.153	57,6%	560	14,3%

A Tabela 6 mostra a distribuição de estudantes em relação ao seu local de moradia. Os estudantes que moram e estudam na mesma cidade são maioria, e na base geral

possuem menor percentual de evasão do que aqueles que moram e estudam em cidades diferentes. Todavia, na base de finalistas, este último grupo passa a ter menos evasão.

**Tabela 6: Total de estudantes e percentual de evasão por região de moradia.**

Mesma Cidade	Geral		Finalista	
	Qtde	Evasão	Qtde	Evasão
Mora e estuda na mesma cidade	24.841	35,5%	18.636	15,6%
Mora em cidade diferente de onde estuda	6.090	46,6%	3.538	10,8%

Para finalizar a exploração dos dados, foi realizada uma análise da Correlação de Pearson, das variáveis de entrada em relação à situação (evasão ou conclusão). Isto foi feito para os estudantes finalistas e para a base geral, com todos os estudantes. Para manter a simplicidade e relevância, foram exibidas apenas as variáveis com correlação absoluta maior que 0,1, em pelo menos uma das bases. O resultado pode ser encontrado na Tabela 7. Os dados foram tratados de forma a ser possível traçar uma correlação numérica. As variáveis categóricas foram transformadas em variáveis booleanas, de forma que cada valor seja representado por sua ausência (0) ou presença (1). A variável “Situação” teve seus valores mapeados como 0 (evadido) ou 1 (concluído). Assim, as correlações positivas indicam que a variável influencia para a conclusão, enquanto os valores negativos influenciam para a evasão.

É importante ressaltar que essa análise é uma correlação meramente linear, o que não é uma relação direta de causa ou consequência. Analisando os dados obtidos fica evidente que o IRA (Índice de Rendimento Acadêmico) é o critério que possui a maior correlação com a situação do aluno, isso significa que quanto maior o IRA pode existir uma probabilidade maior do aluno concluir o curso, isso é observado para os dois grupos de estudantes (finalista e geral).

**Tabela 7: Coeficiente de correlação das variáveis de entrada com a situação.**

Variável	Geral	Finalista
IRA	+0,62	+0,48
Grande Área (Ciências da Saúde)	+0,15	+0,19
Grande Área (Ciências Exatas e da Terra)	-0,14	-0,13
Turno (Noturno)	-0,12	-0,11
Sexo (Feminino)	+0,11	0,08
Grande Área (Ciências Humanas)	-0,11	-0,18
Turno (Integral)	+0,10	+0,10
Idade de Saída	+0,06	-0,17
MonTccEst	+0,06	+0,14

Todas as demais variáveis apresentaram índices de correlação bem inferiores ao da variável do IRA, o que indica menores influências na conclusão ou na evasão dos estudantes. De qualquer maneira, pode-se dizer que a segunda variável em termos de influência é a Grande Área. Nota-se que que cursos de Ciências da Saúde de fato influenciam para a conclusão nas bases geral e finalista. Por outro lado, percebe-se que os cursos de exatas e da Terra influenciam para a evasão. As demais áreas também aparecem na tabela, porém com menores impactos na situação final do estudante. Em seguida, consta o sexo, uma vez que as mulheres têm um percentual de conclusão maior

que os homens. O turno também aparece, em que o integral influencia positivamente e o noturno, negativamente.

## **5. Considerações finais**

Este trabalho realizou uma análise exploratória na base de dados de estudantes de graduação presencial de uma universidade federal. Foi considerado como evasão o abandono do curso, ignorando-se a destinação do aluno. Ou seja, adotou-se um critério rigoroso, uma vez que é muito comum não considerar como evasão a troca de curso dentro da própria instituição, ou em alguns casos até mesmo desconsidera-se a remoção do aluno de uma instituição para outra. A abordagem micro foi escolhida por permitir uma análise mais detalhada para a instituição, além de ser mais adequada para avaliação interna dos cursos, afinal os planos pedagógicos de cursos e a execução deles são passíveis de críticas e reformulações. Pretende-se, portanto, que os métodos aplicados neste artigo sejam disponibilizados à instituição que forneceu os dados, caso queiram utilizá-los para fomentar debates e melhorias em suas políticas de permanência e gestão de recursos.

Em contrapartida, algumas análises consideram apenas os níveis macro de evasão. Isto acontece porque em certas situações é mais importante saber se um estudante se graduou ou não. Em outras palavras, não importa se o estudante trocou de curso ou se foi transferido. Além disto, na instituição estudada neste artigo, existem Bacharelados Interdisciplinares, em que o estudante ingressa em uma das três áreas (Exatas, Humanas ou Artes), e após o ciclo básico de três anos, o estudante escolhe um bacharelado para se especializar. Um estudante de Exatas poderia escolher Matemática, Química, Física, Ciência da Computação ou algumas engenharias, por exemplo. E é permitido que o estudante troque de curso quantas vezes queira. Assim, ocorre uma evasão de curso, que é considerada neste artigo, independente do estudante finalizar sua nova graduação ou não. Em outras avaliações, consideraria-se evasão apenas se o estudante não terminasse nenhuma graduação. Desta forma, não é possível realizar uma comparação direta dos percentuais de evasão apresentados neste artigo com os dados do Censo e com outros artigos. Para que a comparação seja efetiva e justa, é necessário que o leitor se atente ao tipo de evasão que está sendo retratada, destacando-se mais uma vez que este trabalho considerou a evasão de curso.

A análise exploratória exibiu diversas tabelas com o total de estudantes distribuídos em cada valor das variáveis, com o seu respectivo percentual de evasão. Notou-se que, como era de se esperar, a evasão nos períodos finais é menor em proporção e em número absoluto. Também se notou que o número de mulheres é maior que o de homens, além delas evadirem menos proporcionalmente. A variável mais impactante para a evasão ou conclusão, no entanto, é o desempenho, medido através do índice de rendimento acadêmico. Foi observado que a maior parte dos estudantes se encontra na faixa superior, com rendimento acima de 75%. Além disto, também como esperado, percebeu-se que os estudantes com desempenho mais alto tendem a evadir bem menos, o que indica que baixas notas e retenções são fatores importantes na desistência. A grande área possui percentual de evasão bem diferentes entre si. As Ciências Exatas demonstraram o menor percentual conclusão, enquanto as Ciências da Saúde obtiveram o melhor resultado. Em todos os casos, a evasão dos finalistas é menor que a base geral.

Em relação as cotas, elas refletem as políticas de ações afirmativas no ingresso dos estudantes. Percebe-se que há um alto percentual de evasão nos grupos D e E na base

geral. Porém, ao se analisar as cotas A e B, que são as mais vulneráveis, elas possuem níveis menores de desistência. O grupo B alcança a menor evasão entre todos os grupos, inclusive menor que a ampla concorrência (Grupo C). Na base de finalistas, os percentuais de evasão atingiram níveis semelhantes entre si. Portanto, pode-se concluir que em termos de evasão escolar os estudantes cotistas têm obtido desempenho semelhante aos não-cotistas, o que indica o sucesso de tais políticas.

É importante ressaltar que esta pesquisa faz uma avaliação meramente quantitativa das informações constantes no sistema, sendo limitada a estas variáveis. Assim, em trabalhos futuros, pretende-se aprofundar a pesquisa sobre evasão tardia, através abordagens qualitativas. Ao se tentar diferentes formas de coleta de dados, como entrevistas, pretende-se captar informações mais subjetivas e amplas acerca do problema.

## Referências

- 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)]
- Brito Júnior, I. (2018). Uso de mineração de dados educacionais para a classificação e identificação de perfis de evasão de graduandos em Sistemas de Informação (Monografia, Universidade Federal do Rio Grande do Norte).
- Couto, D., & Santana, A. (2017). Mineração de dados educacionais aplicada à identificação de variáveis associadas à evasão e retenção. II Congresso sobre Tecnologia na Educação (pp. 333-344).
- de Oliveira Júnior, J. G.; Noronha, R V.; Kaestner, C. A. A. (2017) Método de seleção de atributos aplicados na previsão da evasão de cursos de graduação. Revista de Informática Aplicada, v. 13, n. 2. [DOI: [10.13037/ria.vol13n2.206](https://doi.org/10.13037/ria.vol13n2.206)]
- O'Neill, L., Vonsild, M. C., Wallstedt, B., & Dornan, T. (2013). Admission criteria and diversity in medical school. *Medical education*, 47(6), 557-561.
- Pereira, F. C. B. et al. (2003) Determinantes da evasão de alunos e os custos ocultos para as instituições de ensino superior: uma aplicação na Universidade do Extremo Sul Catarinense. Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia de Produção.
- Prestes, E. M. D. T.; Fialho, M. G. D.; Pfeiffer, D. K. (2014). A evasão no ensino superior globalizado e suas repercussões na gestão universitária. 6º Encontro Internacional da Sociedade Brasileira de Educação Comparada.
- Santana, A. P.; Perrosso, J. E. C.; Macedo, K. L. O.; Farias, S. P. D. (1996) Evasão escolar em escolas públicas municipais rurais localizadas em Montes Claros. Trabalho de Conclusão de Curso. Universidade Estadual de Montes Claros.
- Vasquez, J.; Castaño, E.; Gallón, S.; Gomez, K. (2003). Determinantes de la deserción estudiantil en la Universidad de Antioquia. Borradores del Cie, (04), 1-38.