



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

EDUARDO PEREIRA DOS SANTOS

**A TRAJETÓRIA DO EVADIDO NO CURSO DE CIÊNCIA DA
COMPUTAÇÃO DA UFCG**

CAMPINA GRANDE - PB

2022

EDUARDO PEREIRA DOS SANTOS

**A TRAJETÓRIA DO EVADIDO NO CURSO DE CIÊNCIA DA
COMPUTAÇÃO DA UFCG**

**Trabalho de Conclusão Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

Orientador: Professor Dr. Francisco Vilar Brasileiro.

CAMPINA GRANDE - PB

2022

EDUARDO PEREIRA DOS SANTOS

**A TRAJETÓRIA DO EVADIDO NO CURSO DE CIÊNCIA DA
COMPUTAÇÃO DA UFCG**

**Trabalho de Conclusão Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

BANCA EXAMINADORA:

**Professor Dr.(a.) Francisco Vilar Brasileiro
Orientador – UASC/CEEI/UFCG**

**Professora Dr.(a.) Livia Sampaio Campos
Examinador – UASC/CEEI/UFCG**

**Professor Tiago Lima Massoni
Professor da Disciplina TCC – UASC/CEEI/UFCG**

Trabalho aprovado em: 30 de Março de 2022.

CAMPINA GRANDE - PB

ABSTRACT

The evasion is one of the biggest problems in Brazilian public higher education, representing a great waste of resources, as well as a great social loss. There are many studies on causes and solutions for this problem, but little is known about the trajectories followed by these students after evasion, to better understand the extent of these losses. In particular, for professions linked to the area of Information Technology, which do not have regulations in Brazil, a higher education is not necessary to enter the job market, allowing people who have not completed their courses to be absorbed. In this context, this work intended to map the trajectories followed by the dropouts from the Computer Science course at the Federal University of Campina Grande, in order to verify to what extent the invested resources were really wasted. A preliminary, quantitative analysis was presented, using the institution's database, identifying groups and their characteristics. Subsequently, professional and academic data were collected from the social network LinkedIn, adding information to the groups found. From the qualitative analysis of this information, the research showed that part of the students use the knowledge acquired in the course, either by assuming positions in the area of information technology, or in other similar undergraduate courses. These results, in a way, demystify the idea that the resources allocated to students who drop out of an undergraduate course in Computer Science are entirely lost.

Keywords: Evasion, Computer Science, Higher Education

A trajetória do evadido no curso de Ciência da Computação da UFCG

Eduardo Pereira dos Santos
eduardo.pereira.santos@ccc.ufcg.edu.br
Universidade Federal de Campina Grande
Campina Grande, Paraíba, Brasil

Francisco Vilar Brasileiro
fubica@computacao.ufcg.edu.br
Universidade Federal de Campina Grande
Campina Grande, Paraíba, Brasil

RESUMO

A evasão é um dos maiores problemas do ensino superior público brasileiro, representando um grande desperdício de recursos, como também uma grande perda social. Há muitos estudos sobre causas e soluções para este problema, porém sabe-se pouco sobre as trajetórias seguidas por esses alunos após a evasão, para entender melhor a extensão dessas perdas. Em particular, para profissões ligadas à área de Tecnologia da Informação, que não possuem regulamentação no Brasil, não é necessário uma formação superior para ingressar no mercado, possibilitando que pessoas que não finalizaram seus cursos sejam absorvidas. Nesse contexto, este trabalho pretendeu mapear as trajetórias seguidas pelos evadidos do curso de Ciência da Computação da Universidade Federal de Campina Grande, a fim de verificar em que medida os recursos investidos foram realmente desperdiçados. Foi apresentada uma análise preliminar, quantitativa, utilizando a base de dados da instituição, identificando grupos e suas características. Posteriormente, dados profissionais e acadêmicos foram coletados da rede social LinkedIn, agregando informações aos grupos encontrados. A partir da análise qualitativa dessas informações, a pesquisa mostrou que parte dos alunos utilizam os conhecimentos adquiridos no curso, seja assumindo cargos na área de tecnologia da informação, seja em outros cursos de graduação similares. Esses resultados, de certa forma, desmistificam a ideia de que os recursos alocados para alunos que evadem de um curso de graduação em Ciência da Computação são integralmente perdidos.

PALAVRAS-CHAVE

Evasão, Ciência da Computação, Ensino Superior

1 INTRODUÇÃO

A literatura retrata a evasão como um fenômeno que possui características variadas e peculiares, resultado de um cenário complexo que envolve fatores econômicos, institucionais e sociais [7]. No Brasil, a evasão é um dos maiores problemas do ensino superior público, representando um grande desperdício de recursos, como também uma grande perda social aos evadidos e ao país [14]. No setor privado, a evasão representa uma perda significativa de receita. Em ambos os casos, ela é fonte de ociosidade para professores, funcionários, equipamentos e espaços físicos e sociais [14].

Em 2007, o Instituto Lobo estimava que, todo ano, aproximadamente 4,5 bilhões de reais foram gastos em mensalidades por estudantes que não se formariam no curso em que estavam matriculados. Esse valor somado ao gasto público com alunos que abandonam seus cursos, saltava para 6 bilhões de reais anuais. No caso das Instituições de Ensino Superior (IES), a perda de receita das instituições privadas com a evasão passava dos 10 bilhões de reais

por ano, o que significava quase a metade da receita do conjunto dessas instituições [18].

Nas últimas duas décadas houve um aumento significativo de estudantes no ensino superior, motivado por programas de incentivo e democratização do acesso à educação. Com destaque ao Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI), que tinha como objetivo criar condições para a ampliação do acesso e permanência na educação superior [1], a lei de cotas que criou incentivos para que estudantes de escolas públicas, de baixa renda, negros, pardos e indígenas (PPI) e pessoas com deficiência (PcD) ingressem no ensino superior [2] e programas de acesso ao ensino superior privado, como o Fundo de Financiamento Estudantil (FIES) e o Programa Universidade para Todos (Prouni).

Com o aumento do corpo estudantil, problemas já conhecidos, como a evasão, se tornaram objetos recorrentes de estudo. Um levantamento feito pelo G1, utilizando dados do Censo da Educação Superior, aponta uma taxa de evasão de 28.89% no ano de 2018 [16]. Há alguns estudos isolados que tentam estimar os custos e desperdícios. Segundo o cálculo do pesquisador do Instituto Lobo para o Desenvolvimento da Educação, da Ciência e da Tecnologia, Oscar Hipólito, com base nos números do Censo do Ensino Superior, as perdas financeiras com a evasão, em 2009, chegaram a 9 bilhões de reais [15]. Já um estudo realizado na Universidade Federal da Paraíba (UFPB), apontou um desperdício em torno de 600 milhões de reais entre os anos de 2007 e 2009, considerando apenas o Campus 1 daquela instituição [5]. Esse trabalho na UFPB pontua um mascaramento na evasão, pois muitos alunos abandonam o curso, mas não a instituição e são considerados novos alunos [5]. Em ambos, o aluno não é visto de forma ampla, ou seja, o pós-evasão é desconsiderado.

Nesse sentido, é plausível que esses recursos desperdiçados, mesmo que desconhecidos, estejam superestimados. Quando um aluno se transfere para o mesmo curso em outra instituição ou para um curso na mesma área de conhecimento, é comum que os créditos cursados sejam aproveitados no novo curso, dessa forma o custo dessa evasão é amortizado, principalmente no primeiro caso.

Um outro cenário está relacionado com as profissões não regulamentadas, como por exemplo aquelas ligadas à área de Tecnologia da Informação. Nessa área, não é necessário uma formação superior para ingressar no mercado de trabalho, possibilitando que pessoas que não finalizaram seus cursos sejam absorvidas. Nesse sentido, é possível que um aluno evadido consiga ingressar no mercado de trabalho mesmo sem finalizar seu curso. Dessa forma, parte da evasão não necessariamente é desperdício.

Há muitos estudos sobre causas e soluções para a evasão [9][8] [11] e alguns trabalhos que buscam explorar as trajetórias dos alunos egressos [10][13][26]. Entretanto, não existem estudos relacionados às trajetórias seguidas pelos alunos após a evasão, ou seja, o caminho após a evasão geralmente é desconhecido, por isso não se pode afirmar que o valor investido foi perdido integralmente.

Nesse contexto, este trabalho propõe um estudo de caso no curso de Ciência da Computação da Universidade Federal de Campina Grande (UFCG), a fim de mapear os caminhos dos evadidos e compreender melhor a extensão das perdas. Esse estudo teve início com uma análise quantitativa, tendo como base os dados acadêmicos desses alunos evadidos e identificou grupos com características bem distintas através da aplicação de um algoritmo de classificação. Em seguida, uma análise qualitativa foi realizada utilizando dados textuais coletados da rede social LinkedIn, agregando mais informações aos grupos identificados.

Ao final, o estudo mostrou que, alguns alunos estão utilizando os conhecimentos do curso, seja por assumirem cargos de TI ou cursarem cursos relacionados, dessa maneira foi possível traçar um perfil desse aluno que costuma seguir na área. Portanto, confrontando o pensamento de que os valores investidos em um aluno evadido são totalmente desperdiçados.

2 TRABALHOS RELACIONADOS

Nessa seção serão apresentados alguns trabalhos que buscam traçar o perfil e a trajetória dos alunos em outros contextos.

Um projeto chamado “A trajetória Acadêmica e Profissional dos Alunos da USP” deu início a uma linha permanente de estudos e análises sobre alunos e ex-alunos de graduação e pós-graduação da Universidade de São Paulo. Esse projeto se constitui em três pesquisas paralelas: um estudo longitudinal sobre a trajetória profissional dos alunos, pela aplicação de questionários ao universo de alunos ingressados na Universidade em 1991 nos cursos de graduação, um estudo sobre a vida profissional dos estudantes de graduação formados pela Universidade de São Paulo nos últimos 10 anos e um estudo sobre alunos e ex-alunos de pós-graduação da Universidade de São Paulo que iniciaram seus cursos nos últimos 10 anos. É um projeto bem extenso e infelizmente não tenho informações sobre a sua conclusão, mas inicialmente traçaram um perfil socioeconômico dos alunos com informações de formulários. É uma iniciativa bem interessante já que busca essa análise de forma longitudinal e ampla [10].

Ainda na USP, uma dissertação de mestrado com o título de “Formação e identidade profissional: a trajetória de egressos de Ciências Sociais”, teve como objetivo analisar traços que têm caracterizado a relação entre a formação e o mundo do trabalho para os ex-alunos de Ciências Sociais, comparando eventuais diferenças com outras gerações. Um questionário foi aplicado a dois mil cientistas sociais de todo o país e 35 deles, com diferentes perfis de formação e inserção profissional, participaram de uma entrevista semiestruturada. Uma das conclusões foi a percepção de que esses egressos, em quaisquer campos de atuação, estão utilizando os conhecimentos adquiridos no curso [26].

Na UFMG, um estudo no curso de Fonoaudiologia, buscou compreender aspectos da trajetória profissional e continuidade acadêmica dos egressos. Por meio de um questionário aplicado a 250

egressos, concluíram que o aumento da idade, maior tempo de graduação, jornada de trabalho acima de 20 horas, avaliação positiva sobre a formação acadêmica são aspectos que contribuíram para a independência financeira do egresso fonoaudiólogo. Os alunos que relataram menores dificuldades na inserção do mercado de trabalho avaliaram o curso de graduação mais positivamente (muito bom/ótimo) [13].

Por fim, um trabalho na UFSC, descreveu o perfil dos estudantes evadidos do curso de Gestão da Assistência Farmacêutica - EaD. Os dados utilizados foram obtidos a partir do Sistema de Controle Acadêmico da Pós-Graduação da UFSC. Como resultado, a taxa de evasão foi de 38.0%. Destacou-se o predomínio do sexo feminino, faixa etária entre 25 e 39 anos, até 10 anos de formado e vinculados aos Polos da região Nordeste ou Norte, entre outras características, sendo que algumas delas refletem o perfil geral dos estudantes matriculados no Curso. Os autores concluem que o perfil de evasão pode contribuir para a definição de estratégias que contribuam para a permanência dos estudantes em cursos a distância, especialmente aqueles de pós-graduação [12].

Nesse sentido, existe uma escassez de estudos voltados para a trajetória pós evasão. Os trabalhos existentes, geralmente, focam nas trajetórias dos egressos ou razões que levam à evasão.

3 REFERENCIAL TEÓRICO

Nesta seção definimos o conceito de agrupamento e explicamos o funcionamento do algoritmo k-means, que será utilizado na base de alunos evadidos, com intuito de extrair particularidades e entender o comportamento deles. Além disso, apresentamos a definição de *Web Scraping*, essencial para coleta dos dados textuais, *Text Mining* para processar texto e *WordCloud* para trazer uma visualização para essa análise textual. Esses métodos serão utilizados na nossa metodologia, explicada na Seção 4.

3.1 Agrupamento

Agrupamento, ou *clustering*, é o nome dado para um conjunto de técnicas computacionais cujo propósito consiste em separar objetos em grupos, baseado nas características que estes objetos possuem. A ideia consiste em colocar em um mesmo grupo, objetos que sejam similares, de acordo com algum critério pré-determinado. Considerando que o agrupamento é uma técnica de aprendizado não supervisionado, ele também pode ser usado para extrair características ocultas dos dados e fazer suposições sobre suas propriedades [20].

Um algoritmo básico de agrupamento é o k-means [23]. A abordagem, como muitos métodos de agrupamento, é altamente algorítmica (não pode ser resumida em uma fórmula) e iterativa. A ideia básica é encontrar os centróides (ponto em que as coordenadas são as médias das coordenadas dos pontos) de um número fixo de grupos de pontos em um espaço de alta dimensão.

Essa abordagem, como a maioria dos métodos de agrupamento, requer uma métrica de distância definida, um número fixo de grupos e uma estimativa inicial dos centróides do agrupamento. Não há uma abordagem definida para determinar a configuração inicial dos centróides, porém muitos algoritmos selecionam aleatoriamente os pontos do seu conjunto de dados como os centróides iniciais

Resumidamente, o algoritmo segue os seguintes passos:

- (1) Fixar o número de clusters em algum inteiro maior ou igual a 2
- (2) Iniciar os centróides de cada cluster
- (3) Atribuir os pontos ao centróide mais próximo
- (4) Recalcular a posição dos centróides e repetir o processo até a convergência ou o número de iterações máxima seja atingida

O K-means produz:

- Uma estimativa final dos centróides do cluster (ou seja, suas coordenadas)
- Uma atribuição de cada ponto ao seu respectivo cluster

A Figura 1 apresenta um exemplo de agrupamento, utilizando o k-means e considerando a tradicional base de dados nativa do RStudio *mtcars*. Os elementos foram agrupados em 4 grupos (cada um representado por uma cor), considerando variáveis como motor, peso, número de cilindros e potência.

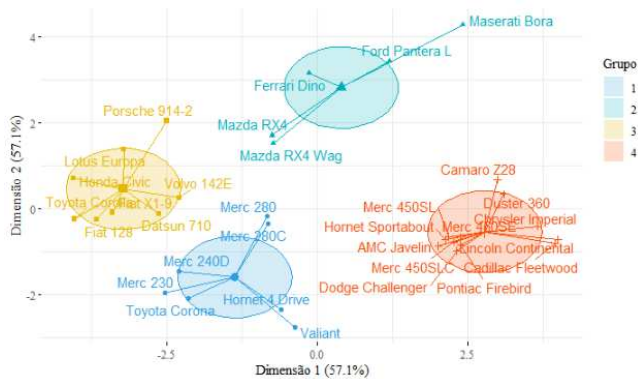


Figura 1: Exemplo de agrupamento

Um método muito utilizado para definir o número ideal de clusters (k) é o método do cotovelo. A ideia é executar o k-means com quantidades diferentes de clusters a fim de definir um número ótimo. Este método testa a variância dos dados em relação ao número de clusters. O valor ideal de K é aquele que tem a menor soma dos quadrados intra-clusters, comumente abreviado para *wcss*, do inglês *within clusters sum of squares*, e ao mesmo tempo o menor número de clusters. Chamamos de método do cotovelo, porque a partir do ponto que seria o “cotovelo” não existe uma discrepância tão significativa em termos de variância. Dessa forma, a melhor quantidade de clusters seria exatamente onde o cotovelo estaria [24]. Na Figura 2, apresentamos um exemplo da execução desse método.

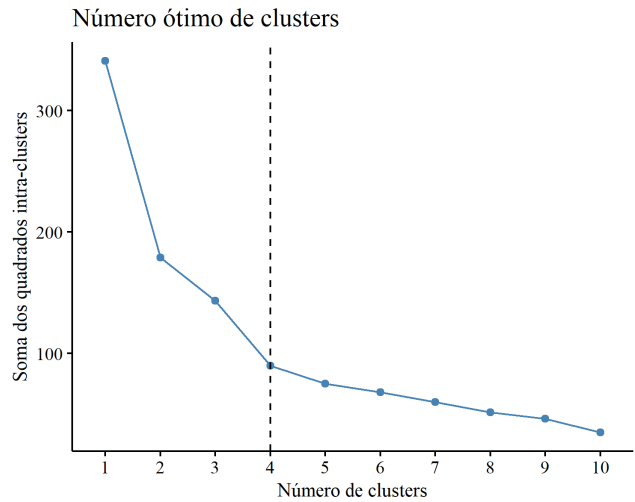


Figura 2: Exemplo do método do cotovelo

3.2 Raspagem de Rede

A raspagem de rede (tradução livre para *web scraping*), também conhecido como extração de dados da web, é o processo de coleta de dados estruturados da web de maneira automatizada. Em geral, esse método é usado por pessoas e empresas com intuito de coletar dados da web disponíveis publicamente, tipicamente para tomar decisões baseadas nesses dados [3].

O processo, geralmente, possui os seguintes passos:

- (1) Identificar o site que contém os dados de interesse
- (2) Coletar URLs das páginas de onde os dados serão extraídos
- (3) Fazer download da página no formato HTML
- (4) Usar localizadores para encontrar os dados no arquivo HTML
- (5) Salvar os dados em um arquivo JSON ou CSV ou algum outro formato estruturado

3.3 Mineração de Texto

Amplamente utilizada em organizações orientadas para o conhecimento, a mineração de texto é o processo de examinar grandes coleções de documentos para descobrir novas informações ou ajudar a responder a perguntas [21].

A mineração de texto identifica fatos, relacionamentos e afirmações que, de outra forma, permaneceram escondidos na grande massa de dados textuais existente. Uma vez extraídas, essas informações são convertidas em um formato estruturado que pode ser analisado posteriormente ou apresentado diretamente usando tabelas, mapas mentais, gráficos e etc. A mineração de texto emprega uma variedade de metodologias para processar o texto, uma das mais importantes sendo o Processamento de Linguagem Natural (NLP, do inglês *Natural Language Processing*) [21].

O processo de mineração de texto compreende várias atividades que permitem deduzir informações de dados de texto não estruturados. Antes de aplicar diferentes técnicas de mineração de texto, é necessário o pré-processamento de texto, que é a prática de limpar e transformar dados de texto em um formato utilizável. Essa prática é um aspecto central do processamento de linguagem natural e

ano de evasão, ano de ingresso, período de evasão e período de ingresso.

- **Histórico:** número de matrícula do aluno, código da disciplina, situação da disciplina, tipo de matrícula na disciplina, período e ano da matrícula na disciplina
- **Disciplinas:** código da disciplina e quantidade de créditos

Foram registradas, nesse período, 1.224 evasões no Curso de Graduação em Ciência da Computação da UFCG (CGCC). Porém, 277 registros dizem respeito a reingresso no mesmo curso (código 12, na Tabela 1), que correspondem a alunos que fizeram reingresso no curso para “limpeza de currículo”. A limpeza do currículo é um método utilizado pelos alunos em um esforço para melhorar os baixos índices de desempenho acadêmico, ou seja, eliminar de seus currículos informações sobre reprovações ou trancamentos de matrícula em disciplinas. Nessa prática, o aluno, mesmo ainda vinculado à instituição, deve passar por uma nova seleção e ingressar no curso com nova matrícula, sendo dispensado de cursar as disciplinas nas quais já havia sido aprovado em uma matrícula anterior [22].

Além disso, identificou-se alguns alunos com duas ou mais matrículas que envolviam outras formas de evasão. Também estavam presentes alunos egressos e ativos que evadiram anteriormente. Utilizamos o CPF para identificar quais alunos tinham mais de um número de matrícula, e com essa abordagem detectamos todos os alunos que fizeram algum tipo de reingresso. Dessa forma foi possível mapear a trajetória desses alunos de forma consistente.

O primeiro passo foi remover da base todas as matrículas dos alunos egressos e ativos no curso, reduzindo o número de matrículas a 936. Em seguida, foi criado um novo registro para cada aluno, mapeando toda sua trajetória no curso: número de semestres, créditos matriculados e integralizados, mantendo a forma de evasão mais recente. Depois desse processo, foram identificados 810 alunos evadidos, distribuídos da forma apresentada na Figura 4.

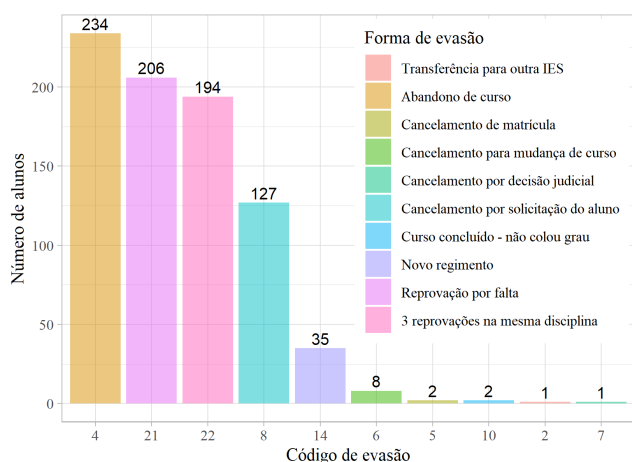


Figura 4: Distribuição dos alunos evadidos do CGCC de acordo com o código de evasão (2009-2019)

A partir dos dados apresentados na Figura 4, foram feitas algumas simplificações para facilitar a análise. Em primeiro lugar, formas de evasão similares foram colapsadas. Em particular, as formas 4 e 5, 6

e 14 e 21 e 22 foram agrupadas, conforme discutido adiante. Além disso, os dados das formas de evasão 2, 7 e 10 foram descartadas por se tratarem de um número desprezível de alunos.

Em resumo, as 936 matrículas correspondem a 810 números distintos de CPF. Após a remoção dos CPFs vinculados às matrículas com forma de evasão 2, 7 e 10, restaram 806 na base. Ao final, dois alunos foram removidos por não apresentarem histórico, totalizando 804 perfis únicos.

Por fim, as formas de evasão restantes foram agrupadas por similaridade da seguinte forma:

- **Abandono:** inclui alunos com os códigos 4 e 5. Perda de vínculo com a instituição aplicada por não efetuarem a matrícula nem o trancamento em um período letivo, ou que no período de vencimento do prazo máximo fixado para integralização curricular não solicitou prorrogação de prazo para conclusão do curso
- **Novo Curso:** inclui alunos com os códigos 6 e 14. Perda de vínculo aplicada ao aluno que realizou mudança de curso
- **Cancelamento:** inclui alunos com código 8. Perda de vínculo com a instituição aplicada ao aluno que solicitar sua desvinculação em qualquer momento do curso
- **Desempenho:** inclui alunos com os códigos 21 e 22. Perda de vínculo com a instituição aplicada ao aluno que for reprovado por falta em todas as disciplinas em um período letivo ou acumulou três reprovações na mesma disciplina.

Na Figura 5, apresentamos a distribuição dos alunos por tipo de evasão.

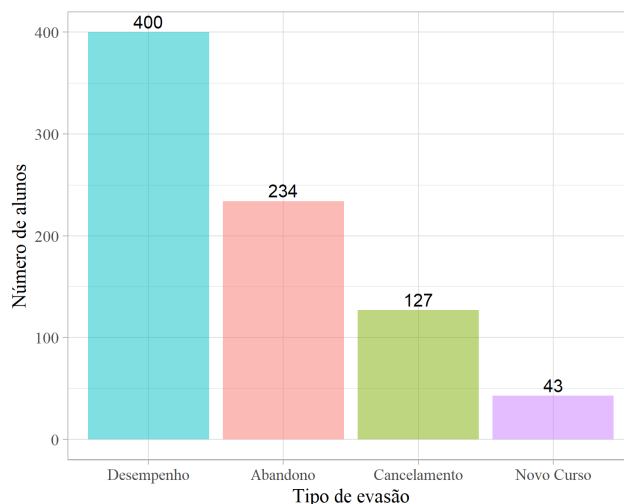


Figura 5: Distribuição de alunos por tipo de evasão

4.1.2 LinkedIn.

O processo de coleta foi dividido em duas fases. São elas:

- (1) **Busca automática de URLs:** nessa etapa os nomes dos 804 alunos foram colocados em uma planilha e processados utilizando o “LinkedIn Profile URL Finder” do Phantombuster. Após uma análise manual nos perfis encontrados, 166 correspondiam aos alunos de fato.

(2) **Scrapping dos perfis:** as 166 URLs foram processadas no “LinkedIn Profile Scraper”, retornando todos os dados disponíveis do perfil - nome, emprego, empresa, educação, dados de contato (incluindo endereços de e-mail) e outros.

Em seguida, os dados semi-estruturados extraídos do LinkedIn passaram por um processo de seleção e filtragem. De início, foram selecionadas as colunas “headline”, “allSkills” e “schoolDegree” dos arquivos csv resultado do scrapping. Essas variáveis foram escolhidas por possuírem informações sobre cargos, formações acadêmicas e habilidades técnicas dos evadidos. Em seguida, foram criados datasets de documentos (chamados ‘corpus’) utilizando o pacote *tm* do R.

Após a obtenção do corpus, são aplicadas funções de limpeza para:

- Remover pontuação
- Eliminar os espaços em branco entre cada item de texto
- Remover números
- Remover palavras comuns da língua inglesa (stop words)
- Remover palavras comuns da língua portuguesa (stop words)
- Alterar as maiúsculas para minúsculas

Ao final, as frequências de palavras foram calculadas baseadas no texto limpo. Isto ranqueia as palavras mostrando os termos mais usados. Esse processo foi aplicado em todos grupos resultantes do agrupamento.

4.2 Métodos de Análise

Os dados do SCAO já filtrados foram utilizados para geração de um perfil, posteriormente utilizado no processo de *clustering*, a fim de extrair mais características dos evadidos. Além disso, é descrito todo o processo preparatório para o agrupamento.

4.2.1 Perfil do Evadido.

De posse dos dados acadêmicos, foi possível traçar um perfil para cada aluno evadido. Utilizando a biblioteca *tidyquery*³ para extrair essas variáveis.

Esse perfil é formado pelas seguintes variáveis: forma de evasão, créditos matriculados, créditos integralizados, taxa de sucesso, número de semestres, carga média e idade do aluno no ano da evasão. A escolha dessas variáveis se deu pelo fato de serem, na nossa opinião, as que melhor caracterizam individualmente cada aluno e sua trajetória no curso.

Logo abaixo, as variáveis extraídas são descritas:

- **Créditos matriculados:** soma de todos os créditos matriculados
- **Créditos integralizados:** soma de todos os créditos onde o aluno obteve a aprovação
- **Semestres:** número de semestres cursados
- **Taxa de sucesso:** calculada a partir da divisão entre os créditos integralizados e matriculados. Representa o nível de aproveitamento do curso. Um número entre 0 e 1, onde 0 representa nenhum aproveitamento e 1 aproveitamento total
- **Carga média:** calculada a partir da divisão entre créditos matriculados e o número de semestres cursados. Representa

o passo de cada aluno no curso, ou seja quantos créditos, em média, cada aluno se matriculou

- **Idade:** calculada a partir da diferença entre o ano de evasão mais recente e o ano de nascimento

4.2.2 Preparação para o agrupamento e definição do número de clusters.

Nesse processo, as variáveis: número de semestres, idade do aluno no ano da evasão, taxa de sucesso e carga média, pertencentes ao perfil, foram selecionadas e normalizadas com a função *scale* do R. Em seguida, o método do cotovelo foi aplicado a fim de auxiliar na escolha do número de *clusters* (k) ideal. Como observado na Figura 6, acima de 6 clusters se tem pouco ganho de homogeneidade. Examinando os resultados com 4 e 5 clusters, foi possível perceber grupos mais descritivos com 4 clusters, por conta disso, arbitramos 4 grupos.

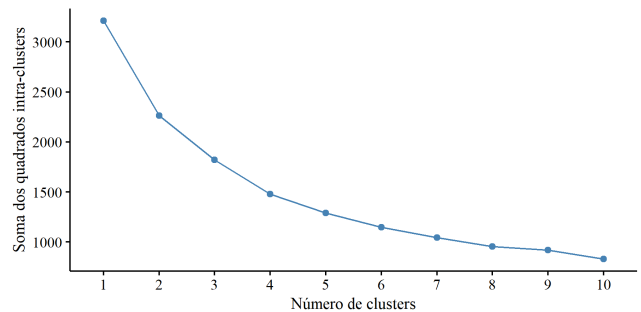


Figura 6: Aplicação do método do cotovelo

5 RESULTADOS

5.1 Agrupamento

Como resultado do agrupamento, temos 4 grupos de alunos evadidos. Na Figura 7, é possível observar a média das variáveis em cada *cluster*, assim é possível extrair características específicas de cada um deles.

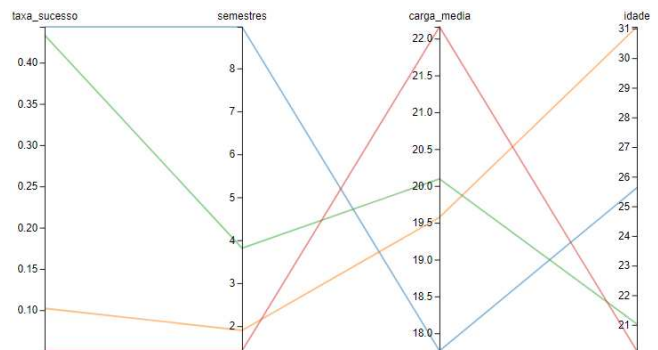


Figura 7: Média das variáveis em cada cluster

Nesse sentido, podemos descrever os clusters da seguinte forma:

- **Evadido mediano:** aquele aluno que obteve êxito em torno de metade das disciplinas e cursou em torno de 4 semestres

³<https://github.com/ianmcook/tidyquery>

- **Evadido persistente:** aquele aluno que possui uma quantidade de semestres acima do permitido, indicando possível reingresso no curso
- **Evadido experiente:** aquele aluno mais velho do que a maioria dos alunos
- **Evadido habitual:** aquele aluno típico que evade no primeiro semestre do curso e praticamente não conclui nenhuma disciplina, explicado pela taxa de sucesso próxima a 0

A distribuição deles é mostrada na Figura 8.

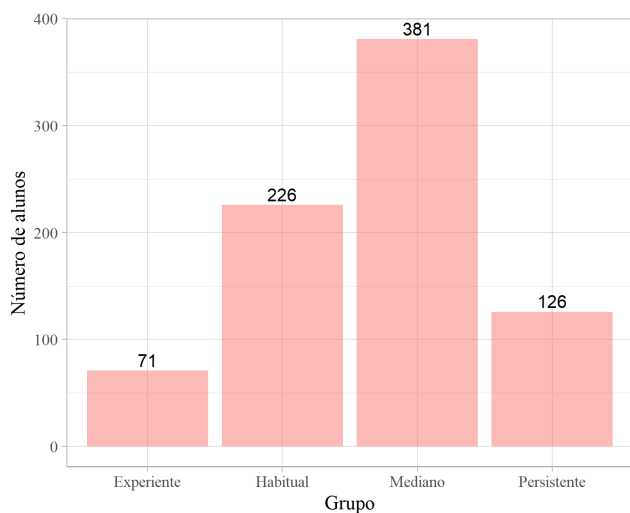


Figura 8: Distribuição dos evadidos por grupo/cluster

Na Figura 9, apresentamos a distribuição dos membros de cada grupo em relação aos tipos de evasão. Observamos que em todos os grupos o tipo de evasão mais comum é o *Desempenho*. O *Cancelamento* e *Novo Curso* apresentam uma participação mais relevante nos grupos *Habitual* e *Mediano*. Já o *Abandono* figura em todos os grupos.

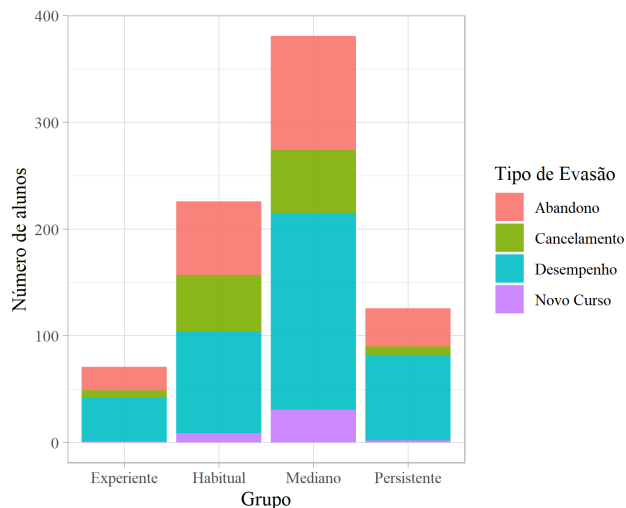


Figura 9: Distribuição dos evadidos por grupo e tipo de evasão

5.2 Análise do LinkedIn

Para cada um dos quatro grupos foi gerado uma nuvem de palavras, com o intuito de destacar as profissões, formações e habilidades e verificar o que é mais recorrente em cada grupo, ou seja trazendo informações que sintetizam a trajetória desses alunos após a evasão. Na Figura 10, temos a distribuição dos perfis encontrados no LinkedIn separados por grupo.

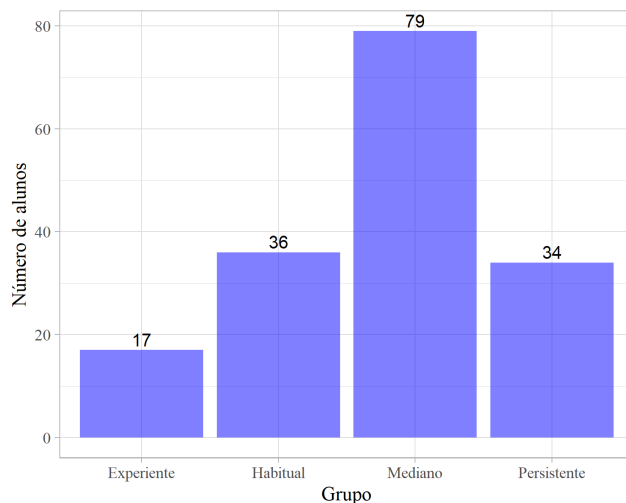


Figura 10: Distribuição dos perfis com LinkedIn por grupo

5.2.1 Evadido Experiente.

Neste grupo foram encontrados e coletados 24% do total de perfis. Há uma diversidade de profissões nesse grupo, entre elas: Corretor de seguros, Instrutor de língua estrangeira e contador. A nuvem de palavras apresentada na Figura 11, indica que as principais habilidades desses profissionais são bem distantes de TI, como já era esperado.

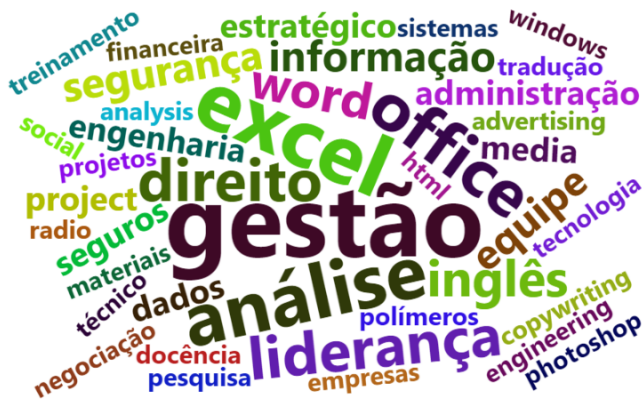


Figura 11: Nuvem de palavras (Evadido Experiente)

5.2.2 Evadido Habitual.

Para esse grupo, foram localizados 16% do total de perfis. Uma diversidade de carreiras está presente, entre elas: Medicina, engenharia civil, direito, economia, geografia e ciência política. É válido ressaltar, a presença de alunos que continuaram na área de TI, cursando alguns cursos ofertados por universidades e faculdades da região. Ciência da Computação - UFPB, Análise e Desenvolvimento de Sistemas - Uninassau e Sistema de Informação - Facisa, figuram como opções para esses evadidos. A frequência de alguns termos ligados a TI, na nuvem de palavras apresentada na Figura 12, como css, java, sql e git reforçam isso.



Figura 12: Nuvem de palavras (Evadido Habitual)

5.2.3 Evadido Mediano.

Esse grupo é o que concentra a maior parte dos evadidos, foram coletados 20.7% do total. Isso colabora para uma diversidade de carreiras e cursos. É exatamente o que foi encontrado; pessoas cursando ou graduadas em Direito, Arte e Mídia, Publicidade, Administração, Medicina, Design ou em alguma Engenharia, por exemplo. É notável a existência de vários termos relacionados com cursos de TI, como mostrado na nuvem de palavras apresentada na Figura 13. Ciência da Computação na UFPB e UFPE, Sistemas da Informação - Facisa e Engenharia da Computação-IFPB foram apenas alguns dos cursos de tecnologia da informação encontrados.



Figura 13: Nuvem de palavras (Evadido Mediano)

5.2.4 Evadido Persistente.

Foram encontrados e coletados 26.5% dos perfis pertencentes a esse grupo, a partir deles foi possível verificar que essas pessoas, em grande maioria, assumem cargos de TI, entre eles: Software Engineer, Data Scientist, Analista, Quality Assurance e Technical Lead. É válido ressaltar também alguns cargos de liderança. No âmbito acadêmico, percebe-se a transferência desses alunos para outras universidades públicas e privadas, como: Facisa, Uninassau, UEPB, IFPB, Universidade Paulista. Dentre os cursos, foi comum encontrar Ciência da Computação, Jogos Digitais, Telemática e Sistemas da Informação. Outras áreas como engenharia figuram, porém em uma frequência bem baixa.

Como observado na nuvem de palavras da Figura 14, os termos de TI são os mais frequentes e aparecem com mais relevância.



Figura 14: Nuvem de palavras (Evadido Persistente)

6 DISCUSSÕES

Nossos resultados apontaram que, de fato, uma parcela considerável dos alunos evadidos, permanecem e atuam na área. É necessário ressaltar que não é todo tipo de perfil que segue essa ideia. Constatamos que o evadido persistente, ou seja aquele que passou mais tempo no curso, teve uma tendência maior de permanecer na área. Algo a pontuar é que um aluno que permanece tanto tempo em

curso provavelmente possui alguma afinidade com o mesmo, então um resultado como esse é bem plausível.

No grupo dos evadidos medianos, temos mais dados, pois foi o que conseguimos encontrar mais perfis no LinkedIn em números absolutos. Nele, a presença de profissionais de TI foi bem notável, apesar de outras profissões também figurarem com frequência.

Em alguns grupos, a frequência de termos em TI é bem menor. Os evadidos experientes, eram estudantes que em média passavam 2 semestres no curso, ou seja, faz sentido pensar que há uma tendência maior em seguir carreiras diferentes de Ciência da Computação e relacionados, pois é possível que o curso tenha sido uma segunda carreira, por exemplo. O mesmo acontece com os evadidos habituais, pois o senso comum é acreditar que um aluno que evade, geralmente, no primeiro semestre do curso, provavelmente não se adaptou ou simplesmente não gostou. Claro que não é tão simples assim, há muitos fatores que podem influenciar a evasão e isto não é objeto desse estudo.

O último ponto interessante observado na análise é que os alunos que continuaram em TI, buscaram formação em outras instituições, principalmente em universidades e faculdades na Paraíba. Saber as razões dessas migrações trariam informações complementares e de grande relevância para esse estudo.

7 CONCLUSÕES E TRABALHOS FUTUROS

Esse estudo buscou mapear as trajetórias dos alunos evadidos do curso de Ciência da Computação da Universidade Federal de Campina Grande, com intuito de demonstrar que parte desses alunos estão utilizando os conhecimentos do curso, ou seja, o dinheiro investido em um aluno evadido nem sempre é totalmente perdido, dessa forma o custo pode estar superestimado. A partir da análise dos dados acadêmicos e com auxílio de informações do LinkedIn, concluímos que muitos alunos seguiram carreiras em TI, inclusive concluindo cursos na região. É importante frisar que uma amostra pequena (20%) dos alunos foram mapeados, por isso considerar outras fontes de dados agregaria muito à pesquisa.

Em uma breve pesquisa, algumas bases se mostraram bem promissoras. Na Plataforma Lattes, encontramos alguns perfis com apenas uma busca manual. A plataforma facilita a busca pois os currículos apresentam nome completo, caso a pessoa tenha vínculo com algum curso de pós-graduação a chance de encontrar aumenta bastante.

O Github, Gitlab e BitBucket poderiam ser usados para evidenciar vínculo com a área de TI. Apesar do uso de nomes personalizados ser um fator que dificulta a pesquisa, foi possível encontrar alguns evadidos pelo nome completo. Os dados do Sisu e Prouni podem indicar mudança de curso, porém não encontramos uma base única com as informações. Explorar os dados divulgados pelas instituições tornaria essa procura inviável, no entanto uma alternativa de mitigação é utilizar o próprio mecanismo de busca do Google. Outra fonte considerada foi o Portal do Empreendedor, porém não encontramos uma forma de pesquisar na plataforma.

Essa pesquisa pode ser expandida também utilizando dados coletados via formulários enviados a esses alunos, como também realizando entrevistas. Por fim, seria interessante e enriquecedor saber o que motiva as pessoas a trocarem o curso de Ciência da Computação na UFCG por outros cursos próximos, como: Ciência

da Computação na UFPB e UEPB, Sistemas da Informação na Facisa e Engenharia da Computação no IFPB. O resultado poderia ser considerado em políticas para reter esse alunos e evitar que a evasão aconteça

AGRADECIMENTOS

Agradeço aos meus pais pelo suporte durante toda a graduação e por estarem sempre ao meu lado. Quero agradecer a todos os professores por essa formação de excelente qualidade, especialmente ao meu orientador, Francisco Brasileiro por todo apoio durante a preparação e execução desse trabalho. Também agradeço a professora Lívia Sampaio pela ajuda na escolha do tema e no acesso aos dados do controle acadêmico, e a Matheus Silva pela colaboração neste trabalho. Por fim, agradeço aos meus amigos por estarem comigo nos momentos mais difíceis do curso.

REFERÊNCIAS

- [1] Brasil. 2007. Decreto nº 6.096, de 24 de abril de 2007. Institui o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais - REUNI. Brasília, DF: Presidência da República. Retrieved September 22, 2021 from http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2007/Decreto/D6096.htm
- [2] Brasil. 2012. Decreto nº 12.711, de 29 de agosto de 2012. Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências. Brasília, DF: Presidência da República. Retrieved September 29, 2021 from http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/112711.htm
- [3] Kenny Colm. 2020. What is web scraping. Retrieved March 09, 2022 from <https://www.zyte.com/learn/what-is-web-scraping>
- [4] John W Creswell. 2010. *Projeto de pesquisa: métodos qualitativo, quantitativo e misto* (3 ed.).
- [5] Emília Maria da T. Prestes e Marília G. Duarte Fialho. 2018. Evasão na educação superior e gestão institucional: o caso da Universidade Federal da Paraíba. *Ensaio: Avaliação e Políticas Públicas em Educação* 26 (2018), 869–889.
- [6] Universidade Federal de Campina Grande. 2007. RESOLUÇÃO Nº 26/2007 Homologa o Regulamento do Ensino de Graduação da Universidade Federal de Campina Grande. Retrieved March 10, 2022 from http://www.cfp.ufcg.edu.br/geo/RESOLUCAO_26_2007.pdf
- [7] Elizabeth Mercuri e Camila Alves Fior. 2017. Análise dos fatores preditivos da evasão em uma universidade confessional. *Congresso Latinoamericano sobre el Abandono de la Educación Superior (CLABES) II* (out 2017).
- [8] Marucia Patta Bordagi e Cláudio Simon Hutz. 2009. Não havia outra saída: percepções de alunos evadidos sobre o abandono do curso superior. *Psico-USF* 14 (2009), 95–105.
- [9] Nádia Velleda Caldas e Flávio Sacco dos Anjos. 2021. Enfrentando a evasão universitária: o caso do projeto tutorias no curso de Zootecnia. *Revista Desenvolvimento Socioeconômico em Debate* 7, 1 (2021), 153–166.
- [10] Simon Schwartzman e Maria H. de M. Castro. 1991. A trajetória acadêmica e profissional dos alunos da USP. *Documento de trabalho* 2 (1991), 91–91.
- [11] Maria A. S. Lopes Ellen Christine M. Dias, Carlos R. Theóphilo e LOPES. 2010. Evasão no ensino superior: estudo dos fatores causadores da evasão no curso de Ciências Contábeis da Universidade Estadual de Montes Claros–Unimontes-MG. *Congresso USP de Iniciação Científica em Contabilidade* (2010), 1–16.
- [12] Adriana Cristiane Longo et al. 2018. Evasão em um curso de pós-graduação a distância: perfil dos estudantes do curso de gestão da assistência farmacêutica. *ANAIS CIAED* (2018).
- [13] Leticia Caldas Teixeira et al. 2013. Trajetória profissional de egressos em Fonoaudiologia. *Revista CEFAC* 15 (2013), 1591–1600.
- [14] Roberto Leal Silva Filho Lobo et al. 2007. A evasão no ensino superior brasileiro. *Cadernos de pesquisa* 37 (set/dez 2007), 641–659.
- [15] G1. 2011. País perde R\$9 bilhões com evasão no ensino superior, diz pesquisador. Retrieved September 23, 2021 from <http://g1.globo.com/educacao/noticia/2011/02/pais-perde-r-9-bilhoes-com-evasao-no-ensino-superior-diz-pesquisador.html>
- [16] G1. 2020. N° de alunos que abandonam faculdade deve subir após a pandemia, e setores poderão enfrentar falta de mão de obra. Retrieved September 09, 2021 from <https://g1.globo.com/educacao/noticia/2020/09/13/no-de-alunos-que-abandonam-faculdade-deve-subir-apos-a-pandemia-e-setores-poderao-enfrentar-falta-de-mao-de-obra.ghtml>
- [17] The R Graph Gallery. 2011. The Wordcloud2 library. Retrieved March 12, 2022 from <https://www.r-graph-gallery.com/196-the-wordcloud2-library.html>
- [18] Terceiro Grau. 6. O Informativo do Gestor Universitário. *Lobo & Associados - Uma Referência Nacional de Consultoria em Educação* 12 (6). <https://www.institutolobo.com>

- org.br/core/uploads/artigos/anexo_cea8c46cc07eea1f0ec9eb2e987ef9d4.pdf
- [19] IBM. [n. d.]. What is Text Mining? Retrieved March 10, 2022 from <https://www.ibm.com/cloud/learn/text-mining>
- [20] Ricardo Linden. 2009. Técnicas de agrupamento. *Revista de Sistemas de Informação da FSMA* 4 (2009), 18–36.
- [21] Linguamatics. [n. d.]. What is Text Mining, Text Analytics and Natural Language Processing? Retrieved March 10, 2022 from <https://www.linguamatics.com/what-text-mining-text-analytics-and-natural-language-processing>
- [22] Francinaldo Carlos Nunes. 2020. Estudo exploratório sobre a evasão no Curso de Computação da UFCG: um olhar sobre a disciplina Cálculo I. *Dissertação de Mestrado. Universidade Federal de Campina Grande* (2020).
- [23] Roger D. Peng. 2020. K-Means Clustering | Exploratory Data Analysis with R. Retrieved January 20, 2022 from <https://bookdown.org/rdpeng/exdata/k-means-clustering.html>
- [24] Erivelton Pires. 2019. Clusterização (k-means). Retrieved March 15, 2022 from <https://www.kaggle.com/code/eriveltonguedes/7-clusteriza-o-k-means-erivelton/notebook>
- [25] PRE. 2016. Procedimentos Operacionais. Retrieved September 22, 2021 from <https://pre.ufcg.edu.br/pre/procedimentos-operacionais-pre>
- [26] Danilo Martins Torini. 2012. Formação e identidade profissional: a trajetória de egressos de Ciências Sociais. *Dissertação de Mestrado. Universidade de São Paulo* (2012).