



UFOP

Universidade Federal
de Ouro Preto

**Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Departamento de Computação e Sistemas**

**Um *framework* para análise e
identificação dos grupos de mulheres
da área da computação e suas
características, mediante técnicas de
mineração de dados**

Nathália Barbosa Mota Moreira

**TRABALHO DE
CONCLUSÃO DE CURSO**

**ORIENTAÇÃO:
Janniele Aparecida Soares Araújo**

**Janeiro, 2022
João Monlevade–MG**

Nathália Barbosa Mota Moreira

Um *framework* para análise e identificação dos grupos de mulheres da área da computação e suas características, mediante técnicas de mineração de dados

Orientador: Janniele Aparecida Soares Araújo

Monografia apresentada ao curso de Engenharia de Computação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

Universidade Federal de Ouro Preto

João Monlevade

Janeiro de 2022



FOLHA DE APROVAÇÃO

Nathália Barbosa Mota Moreira

Um framework para análise e identificação dos grupos de mulheres da área da computação e suas características, mediante técnicas de mineração de dados

Monografia apresentada ao Curso de Engenharia da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de bacharel em Engenharia da Computação

Aprovada em 10 de janeiro de 2022

Membros da banca

Professora Doutora - Janniele Aparecida Soares Araujo - Orientadora - Departamento de Computação e Sistemas
Professora Doutora - Gilda Aparecida de Assis - Departamento de Computação e Sistemas
Professor Doutor - Samuel Souza Brito - Departamento de Computação e Sistemas

Janniele Aparecida Soares Araujo, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 25/01/2022



Documento assinado eletronicamente por **Janniele Aparecida Soares Araujo, PROFESSOR DE MAGISTERIO SUPERIOR**, em 25/01/2022, às 14:31, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0271822** e o código CRC **A01D8B79**.

Este trabalho é dedicado a todas as mulheres da computação.

Agradecimentos

A vida acadêmica nem sempre é fácil, mas com o apoio de pessoas maravilhosas, foi possível concluir mais essa etapa.

Jamais estaria nessa posição de privilégio se não fosse pelos meus pais, Maria Aparecida e Walter. Sabemos como essa caminhada foi sacrificante. Todos os momentos que não pude estar presente porque eu precisava estudar mais uma hora, fazer mais um trabalho, ler mais um capítulo, fazer mais um exercício. A distância nos aproximou ainda mais e a saudade era apenas um lembrete da família incrível que me esperava de braço abertos sempre que eu podia voltar para casa. Com muito carinho, fui ensinada desde criança que nada cai do céu, que tudo é fruto do nosso esforço e aprendizado. Levei esse ensinamento comigo em todos os momentos de frustração, de saudade e de angústia. O apoio da minha família foi essencial para chegar até aqui.

Em João Monlevade encontrei uma segunda família cheia de irmãos maiores que eu, mesmo que só em altura. A faculdade não seria a mesma coisa sem vocês, em especial, Bruno Tomaz, Bryan Pinto, Hellrison Soares, Lucas Lírio, Lucas Pereira, Lucas Rocha, Victor Hugo Duarte e Victor Hugo Dias.

Os amigos à distância foram tão importantes nessa caminhada quantos os presenciais. Só tenho a agradecer a Joana Argenta, Júlia Azevedo e Lucas Passos pelas mensagens de carinho, os momentos de companheirismo e o apoio incondicional durante esses cinco anos. Se cheguei até aqui, saibam que a amizade de vocês foi fundamental nessa jornada.

A graduação nem sempre é o que esperamos, mas sem os professores e professoras que compartilharam comigo seu conhecimento, eu não seria a mesma profissional que termina a faculdade. Em especial, agradeço à todos os docentes que dedicaram sua atenção e respeito às minhas perguntas, dúvidas e incertezas, que foram essenciais para a conclusão bem sucedida dessa etapa. Principalmente a professora Janniele, que mesmo sem me conhecer, me acolheu e apoiou durante todo o processo de elaboração deste trabalho.

A todos que fizeram parte da minha formação acadêmica e pessoal, direta ou indiretamente, o meu muito obrigada.

*“They’d say I hustled, put in the work
They wouldn’t shake their heads and question how much of this I deserve
What I was wearing, if I was rude
Could all be separated from my good ideas and power moves”*

— Taylor Swift,
in: The Man

Resumo

Na fundamentação da computação, as mulheres representaram papel de destaque. Entretanto, atualmente, os cursos relacionados a esta área apresentam um índice cada vez menor de alunas. Sendo assim, este trabalho buscou construir um *framework* para a análise e identificação de grupos e características das mulheres pertencentes a área de Tecnologia da Informação, mediante técnicas de mineração de dados. A partir de uma base de dados fictícia, populada com a utilização de um questionário subdividido em questões de cunho pessoal, sobre a motivação para a escolha do curso, sobre situações de preconceito enfrentadas durante a graduação e questões para análise comportamental, foi possível aplicar técnicas de mineração de dados como passo do *Knowledge Discovery in Databases*, para a obtenção de conhecimento útil. Como resultado, este trabalho conclui que é possível extrair conhecimento útil e explícito que possibilita uma melhor tomada de decisão, tanto para mulheres que desejam ingressar na área, quanto para colegas dos cursos que desejam entender melhor o corpo discente feminino.

Palavras-chaves: mulheres. computação. mineração de dados.

Abstract

In the foundation of computer science, women played a prominent role. However, lately, graduations related to this field have a decreasing rate of female students. Therefore, this project sought to build a framework to analyze and identify groups and characteristics of women that belong to Information Technology, using data mining techniques. From a fictitious database, populated with a questionnaire subdivided into questions of a personal nature, about the motivation behind choosing their major, about situations of prejudice endured during graduation and questions for behavioral analysis, it was possible to apply data mining techniques as a step of Knowledge Discovery in Databases, to obtain useful knowledge. As a result, this work concluded that it is possible to extract useful and explicit knowledge that enables better decision-making, both for women who wish to enter the area, as well as, for collegiate members who wish to better understand the female student body.

Key-words: women. computer science. data mining.

Lista de ilustrações

Figura 1 – Uma visão geral dos passos que compõem o processo de <i>Knowledge Discovery in Databases</i> (KDD), retirada de Fayyad, Piatetsky-Shapiro e Smyth (1996).	21
Figura 2 – Conjunto de dados com duas classes, retirado de Fayyad, Piatetsky-Shapiro e Smyth (1996).	24
Figura 3 – Clusterização do conjunto de dados de empréstimos, retirado de Fayyad, Piatetsky-Shapiro e Smyth (1996).	25
Figura 4 – Trecho da matriz de incidência obtida.	32
Figura 5 – <i>Heatmap</i> obtido.	35
Figura 6 – Gráfico obtido a partir da utilização do <i>scree plot</i>	38
Figura 7 – Gráfico obtido a partir da utilização do método do cotovelo.	39
Figura 8 – Quantidade de observações em cada <i>cluster</i>	41
Figura 9 – Distribuição de idade em cada <i>cluster</i>	42
Figura 10 – Distribuição do ano de ingresso em cada <i>cluster</i>	42
Figura 11 – Respondentes que cursaram o ensino médio na rede particular.	42
Figura 12 – Respondentes que cursaram o ensino médio nas redes pública e particular.	43
Figura 13 – Grupos identificados dentre o que as alunas menos gostam na universidade.	43
Figura 14 – Grupos de alunas que gostam da infraestrutura da faculdade.	43
Figura 15 – Grupos de alunas que gostam dos professores.	44
Figura 16 – Grupos de alunas que trabalham na área do curso.	44
Figura 17 – Perspectivas para após a graduação.	44
Figura 18 – Grau de escolaridade do pai.	45
Figura 19 – Grupos dentre os perfis Dominance, Influence, Steadiness e Compliance (DISC) para realização de atividades em grupo.	45
Figura 20 – Grupos dentre os perfis DISC para preferência de foco.	46
Figura 21 – Média de sentimento em cada <i>cluster</i>	46

Lista de tabelas

Tabela 1 – Descrição da base de dados do questionário	31
Tabela 2 – Comparativo da análise de sentimento em português e inglês	34
Tabela 3 – Atributos mais significativos para cada um dos nove componentes.	38
Tabela 4 – Valores da média do <i>silhouette coefficient</i>	40

Lista de abreviaturas e siglas

CEP Comitê de Ética em Pesquisa

CES Censo da Educação Superior

DISC Dominance, Influence, Steadiness e Compliance

IFMG Instituto Federal de Minas Gerais

INEP Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

KDD *Knowledge Discovery in Databases*

LGPD Lei Geral de Proteção de Dados

PCA *Principal Component Analysis*

TCLE Termo de Consentimento Livre e Esclarecido

TI Tecnologia da Informação

UFPB Universidade Federal da Paraíba

WCSS within-clusters sum-of-squares

Sumário

1	INTRODUÇÃO	14
1.1	O problema de pesquisa	15
1.2	Objetivos	16
1.3	Metodologia	16
1.4	Organização do trabalho	17
2	REFERENCIAL TEÓRICO	18
2.1	Lei Geral de Proteção de Dados	18
2.2	Trabalhos relacionados	18
2.3	Metodologia DISC	20
2.4	Métodos e Ferramentas	20
2.4.1	Coleta de dados a partir de questionários	20
2.4.2	<i>Knowledge Discovery in Databases</i> e mineração de dados	21
2.4.2.1	<i>SentiStrength</i>	23
2.4.2.2	<i>Principal Component Analysis</i> ou PCA	23
2.4.2.3	Clusterização	24
3	DESENVOLVIMENTO	27
3.1	Construção do questionário e desenvolvimento da base de dados fictícia	27
3.2	Criação e pré-processamento dos dados	28
3.3	Aplicação do <i>SentiStrength</i>	32
3.4	Transformação dos dados	35
4	RESULTADOS	41
4.1	Análise dos resultados	41
5	CONCLUSÃO	47
	REFERÊNCIAS	48
	APÊNDICES	50
	APÊNDICE A – QUESTIONÁRIO ELABORADO	51
A.1	Termo de Consentimento Livre e Esclarecido (TCLE)	51

A.2	Informações pessoais	52
A.3	Motivações para a escolha do curso	55
A.4	Percepções durante o curso	56
A.5	Análise comportamental	59

1 Introdução

É notório a predominância masculina nas áreas que são englobadas pela Tecnologia da Informação (TI). Por esse motivo, através deste trabalho, busca-se criar um *framework* para identificar os perfis das mulheres de cursos relacionados à TI.

De acordo com Santos, Carvalho e Barreto (2021), em uma análise dos microdados do Censo da Educação Superior (CES) disponibilizados no portal do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), entre os anos de 2014 e 2019, é possível perceber que as discentes representam cerca de 13,8% das alunas dos cursos de tecnologia da informação e 15,2% dos formandos dos mesmos cursos. Tais dados reforçam o atual cenário da predominância masculina em cursos da área de TI no país.

Moreira, Silva e Carvalho (2018) destacam que na primeira turma do curso de Ciência da Computação da Universidade Federal da Paraíba (UFPB) em 1985, cerca de 36,5% dos discentes eram mulheres, atingindo um pico de 40,7% em 1993. Entretanto, a partir do ano 2000, houve uma queda no interesse feminino nesse curso. No período analisado pelos autores, entre 1985 e 2017, a média de mulheres matriculadas era de 21,3%, com uma diminuição anual de 0,4%. Caso a universidade não implemente ações afirmativas para reverter esse quadro, em 2050 não haverá alunas no curso de Ciência da Computação. Apesar do cenário atual, as mulheres possuíram papel de destaque como precursoras da informática.

É possível citar Ada Lovelace como exemplo amplamente difundido da participação feminina na história da computação. Nascida em dezembro de 1815 e filha do poeta Lord Byron, Lauen (2021) destaca que Ada é conhecida como a primeira programadora. Entretanto, seu pioneirismo não foi devidamente reconhecido na época. Em seu livro sobre a vida de Ada Lovelace, Baum (1986) afirma que, no dia de sua morte, em 27 de novembro de 1852, o jornal *London Times* publicou a nota de falecimento de "Augusta Ada, esposa de William Earl of Lovelace e única filha de George Gordon Noel Lord Byron, após uma longa e dolorosa doença suportada com extrema paciência e coragem. Ela deixou dois filhos e uma filha". Nota-se que nada foi dito sobre seu trabalho, sendo sua vida resumida ao papel de filha, esposa e mãe.

Segundo Hammerman e Russell (2015), a relevância de seu trabalho passou a ser reconhecida apenas na década de 1950, quase 100 anos após sua morte. B. V. Bowen incluiu em seu livro o artigo produzido por Ada sobre a Máquina Analítica de Charles Babbage, baseado no resumo em francês publicado por um jovem engenheiro que compareceu a apresentação de tal máquina. Nessa época, os computadores existentes ocupavam salas inteiras, e os programadores começaram a entender o que significava programá-los. Foi

possível relacionar as instruções escritas por Ada para a hipotética Máquina Analítica com as linhas de código utilizadas até então. Assim, suas instruções passaram a ser reconhecidas como o primeiro programa de computadores do mundo.

Apesar de seu legado sem precedentes até então, com a redescoberta de seu trabalho, a vida de Lovelace passou a ser examinada minuciosamente. [Hammerman e Russell \(2015\)](#) destacam que, sua vida e trabalho foram assunto de diversas biografias e estudos acadêmicos. Enquanto alguns autores celebraram o legado de Ada como um ícone da ciência e tecnologia, outros a retratavam de formas menos elogiosas.

Sendo assim, é possível perceber que, apesar de sua essencial contribuição para o desenvolvimento e evolução das áreas relacionadas à computação, a presença feminina vem se mostrando cada vez mais escassa, fazendo-se necessário compreender o motivo e o que pode ser feito para reverter tal situação.

1.1 O problema de pesquisa

Considerando-se o meio acadêmico dos cursos relacionados à computação, nota-se a predominância dos alunos do sexo masculino. No trabalho de [Freitas \(2019\)](#), destaca-se que apenas 15% dos discentes de tais cursos, entre os anos de 2009 e 2018, eram mulheres. A autora salienta ainda que, no período analisado, houve uma diminuição de 5% no número de mulheres, enquanto o número de homens aumentou na mesma proporção.

[Gama et al. \(2021\)](#) afirmam que a desproporcionalidade dentre alunos e alunas nos cursos de graduação da área de TI tende a reforçar o estereótipo de que a tecnologia é uma área exclusivamente masculina, impedindo assim que, mulheres busquem atuar nesse segmento. As autoras destacam ainda a necessidade de que as instituições de ensino superior forneçam ferramentas para que as discentes se sintam acolhidas e representadas no meio acadêmico, como o projeto "Vai, Meninas!" proposto por esse trabalho, que tinha como objetivo criar grupos de discussão de temas e realização de atividades que visavam o encorajamento e fortalecimento das estudantes, para ajudá-las a enfrentar os desafios das vidas acadêmica e profissional.

Considerando este panorama, torna-se extremamente relevante detectar o que leva as discentes a desgostarem da graduação e do ambiente universitário, além de compreender os motivos que levam as mulheres a preterirem graduações em TI ou desistirem do curso, ainda que, segundo [Freitas \(2019\)](#) as discentes apresentem um desempenho superior aos seus colegas de classe, apesar da desmotivação para darem continuidade à graduação. Traçar perfis das mulheres dos cursos da área da computação e divulgá-los, pode contribuir para uma melhor escolha do curso de graduação dos novos ingressantes e retenção dos mesmos.

1.2 Objetivos

Tem-se como objetivo criar um *framework* para a identificação de grupos entre as mulheres da área da computação de instituições de ensino superior. Para alcançar o objetivo geral, têm-se os seguintes objetivos específicos: a partir de dados fictícios, desenvolver uma ferramenta para identificar e analisar, através de técnicas de mineração de dados, características gerais e pessoais, fatores comportamentais e conhecimento sobre a área.

1.3 Metodologia

A metodologia deste trabalho consiste no estudo e elaboração do referencial teórico, identificação de trabalhos relacionados e melhores práticas para a elaboração de questionários, seguindo as etapas do *Knowledge Discovery in Databases* (KDD) e garantindo o cumprimento da Lei Geral de Proteção de Dados (LGPD).

De acordo com Wazlawick (2009), o método de pesquisa consiste em uma sequência de passos necessários para a demonstração de que o objetivo proposto foi atingido. Assim, o método deve indicar se protótipos serão desenvolvidos, se modelos teóricos serão construídos, quais experimentos deverão ser realizados, a maneira com os dados serão organizados e comparados, entre outros passos, dependendo do tipo de pesquisa que se deseja realizar. O autor destaca ainda que, a definição do método representa um passo fundamental que deve ser executado após a definição do objetivo da pesquisa, descrevendo o caminho percorrido para atingi-lo. Para a realização deste trabalho, seguiu-se a seguinte metodologia:

- Estudo da LGPD.
- Construção da revisão bibliográfica.
- Estudo dos tipos de perfis.
- Submissão do projeto de pesquisa para o Comitê de Ética em Pesquisa (CEP).
- Criação de questionários para a coleta de dados reais em trabalhos futuros, conforme o estudo da LGPD.
- Coleta e pré-processamento dos dados fictícios.
- Estudos, análise e desenvolvimento do *framework* com a disposição de técnicas de mineração de dados.
- Análise dos resultados obtidos.

1.4 Organização do trabalho

O restante deste trabalho é organizado como se segue. O Capítulo 2 apresenta a revisão bibliográfica utilizada para o desenvolvimento, trabalhos relacionados, métodos e ferramentas utilizadas durante a análise dos dados coletados, bem como uma breve pesquisa sobre a LGPD. O Capítulo 3 descreve como o trabalho foi desenvolvido. Já o Capítulo 4 apresenta os resultados obtidos. E, por fim, o Capítulo 5 relata todas as conclusões obtidas durante o desenvolvimento deste trabalho.

2 Referencial Teórico

Este capítulo apresenta uma revisão da literatura, bem como trabalhos correlatos.

2.1 Lei Geral de Proteção de Dados

De acordo com a Lei de nº 13.709/18, de agosto de 2018, a [LGPD](#) tem como principais fundamentos o respeito à privacidade, a auto determinação informativa, a liberdade de expressão, de informação, de comunicação e de opinião, a inviolabilidade da intimidade, da honra e da imagem, entre outros.

Como citado no Art. 3º, a [LGPD](#) se aplica "a qualquer operação de tratamento realizada por pessoa natural ou por pessoa jurídica de direito público ou privado, independentemente do meio, do país de sua sede ou do país onde estejam localizados os dados, desde que: (...) III - os dados pessoais do objeto do tratamento tenham sido coletados no território nacional"(Brasil, 2018).

Nesse sentido, em cumprimento com a [LGPD](#), neste trabalho não serão coletados dados pessoais que possam identificar as respondentes, como nomes, e-mails ou números de documentos. Além disso, não serão coletados dados de crianças e adolescentes, ainda que com o consentimento dos pais ou responsáveis legais.

2.2 Trabalhos relacionados

Esta subseção apresenta trabalhos correlatos, com o objetivo de embasar teoricamente o trabalho desenvolvido e compreender o que já foi estudado sobre a presença de mulheres na área da computação.

No trabalho de [Santos, Santos e Ellwanger \(2021\)](#) foi realizado um levantamento de dados com o objetivo de mapear o interesse e conhecimento prévio de alunas do ensino médio na área da computação, e a partir desses dados, definir estratégias buscando modificar o cenário atual de predominância masculina nas graduações da área de [TI](#). Sendo assim, aplicou-se um questionário composto por 32 perguntas divididas em quatro esferas principais, sendo elas:

- mapeamento do perfil do público-alvo;
- comportamento em relação ao uso da tecnologia;
- conhecimento em relação a área de computação;

- interesse do público em relação a cursos relacionados com a área de computação.

A principal semelhança com este trabalho se encontra na aplicação de questionários com o objetivo de se traçar o perfil das mulheres de cursos de instituições de ensino superior, sendo o trabalho relacionado focado em instituições de ensino médio.

Visando auxiliar as alunas do curso de Sistemas de Informação do Instituto Federal de Minas Gerais (IFMG) – Campus de Ouro Branco, o trabalho de [Gama et al. \(2021\)](#) desenvolveu o projeto de ensino “Vai, Meninas!”. Assim como o trabalho relacionado citado anteriormente, foi aplicado um questionário com 24 perguntas, que buscava traçar o perfil das alunas do curso em relação ao conhecimento da área. Entretanto, a maior diferença entre o trabalho relacionado e este trabalho se dá pelo desenvolvimento de um projeto que tem como objetivo melhorar a participação das alunas e desenvolver um senso de comunidade entre as mesmas, a partir de grupos de discussão que visam o encorajamento e fortalecimento das discentes. Foi possível alcançar o objetivo a partir dos grupos citados, possibilitando às mesmas enfrentar os desafios da vida acadêmica e profissional da melhor forma possível.

[Cursino e Martinez \(2021\)](#) realizaram uma análise descritiva dos dados do CES dos anos de 2009 a 2018, com o objetivo de se criar métricas de apoio às comunidades para incentivarem as mulheres a buscarem cursos de graduação na área de computação. Através da análise realizada por tal trabalho, é possível perceber que apenas 15% dos discentes matriculados em cursos relacionados com a área de tecnologia da informação são mulheres, havendo uma redução de 5% do número de mulheres ao longo de dez anos. A análise estatística de dados proposta pelos autores foi desenvolvida através da linguagem de programação *Python*, com o uso de bibliotecas de suporte para visualização dos dados coletados. Neste trabalho se utilizou como ferramenta a mesma linguagem de programação e suas bibliotecas. Entretanto, os dados utilizados como base para o desenvolvimento deste trabalho foram dados fictícios, e não foram obtidos através da coleta de dados do governo no CES.

[Freitas \(2019\)](#) realiza um estudo da presença reduzida de discentes nos cursos de Ciências da Computação, Engenharia da Computação e Matemática Computacional na Universidade Federal da Paraíba (UFPB). Assim como este trabalho, foi aplicado um questionário para coletar dados sobre as alunas do Centro de Informática da UFPB. Entretanto, foi realizado um estudo de coorte retrospectivo sobre os dados obtidos, pois o mesmo se baseia em dados históricos das discentes. Apesar das semelhanças com este trabalho, os dados foram tratados e analisados se utilizando o *software* Excel e não através da linguagem de programação *Python*.

2.3 Metodologia DISC

Em seu livro *Emotions of Normal People*, Marston (1928) teoriza que existem quatro traços de personalidade, sendo eles dominância, influência, estabilidade e cautela, ou no idioma original *dominance, influence, steadiness e compliance*, palavras que dão origem à sigla DISC. Tal metodologia foi desenvolvida como um método de identificar ações observáveis e previsíveis, bem como traços de personalidade do comportamento humano no cotidiano.

A fim de identificar grupos relacionados à metodologia DISC dentre as mulheres do grupo de estudo, serão adotadas oito perguntas dentro do questionário, para que seja possível determinar qual traço de personalidade é predominante em uma determinada respondente e como essa característica se relaciona com suas outras respostas e o retorno de outras respondentes.

2.4 Métodos e Ferramentas

A seguir, são apresentados os métodos e ferramentas que serão utilizados para o desenvolvimento deste trabalho.

2.4.1 Coleta de dados a partir de questionários

Para se atingir os objetivos deste trabalho, será aplicado um questionário para a coleta de dados. Segundo Gil (2008), a coleta de dados a partir de questionários corresponde a uma técnica de investigação composta por um conjunto de questões que são submetidas à pessoas de interesse para se obter informações relevantes a respeito de seus conhecimentos, crenças, sentimentos, expectativas, entre outros. Sendo assim, o autor afirma que a construção de questionários corresponde à tradução dos objetivos da pesquisa em questões específicas, considerando-se que as respostas obtidas podem ser utilizadas para descrever características da população pesquisada, bem como testar as hipóteses que foram construídas durante o desenvolvimento e planejamento da pesquisa.

Gil (2008) cita como vantagens da aplicação de questionários de forma remota, a possibilidade de obtenção de um maior número de respostas, tendo-se em vista que a distância geográfica não representa um problema para o envio de e-mails; a manutenção do anonimato das respondentes e, considerando que as respondentes não terão contato com os autores deste trabalho, as mesmas não terão suas opiniões influenciadas.

Segundo Gil (2008), dois tipos de questões apresentam maior relevância, sendo esses, questões abertas e fechadas. As questões fechadas possibilitaram a identificação de grupos dentre as respondentes e as abertas, a aplicação do *SentiStrength*.

Além disso, Gil (2008) destaca que algumas perguntas somente devem ser disponibilizadas caso as respondentes optassem por uma determinada alternativa. Como exemplo, temos as perguntas sobre filhos da seção sobre informações pessoais deste trabalho, caso a respondente escolha a alternativa "sim" para a pergunta "Você possui filhos?", a mesma seria direcionada à outra página para que pudesse selecionar a quantidade de filhos.

Como ferramenta para a construção do questionário, será utilizada a plataforma disponibilizada de forma gratuita *Google Forms*¹. Assim, será possível coletar dados dos respondentes, e a partir dos dados obtidos, pode-se extrair informações relevantes para a pesquisa desenvolvida.

2.4.2 Knowledge Discovery in Databases e mineração de dados

De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), o conceito de *data mining* ou mineração de dados é o estudo que visa encontrar padrões úteis nos dados analisados, através da aplicação de algoritmos específicos. Tem-se que, o KDD se refere ao processo de descoberta de conhecimento relevante a partir de dados, e a mineração de dados representa um passo desse processo, conforme a Figura 1.

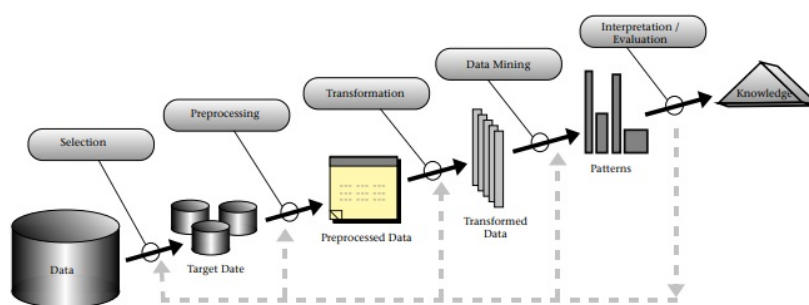


Figura 1 – Uma visão geral dos passos que compõem o processo de KDD, retirada de Fayyad, Piatetsky-Shapiro e Smyth (1996).

Fayyad, Piatetsky-Shapiro e Smyth (1996) destaca que os passos adicionais do processo de KDD, como preparação, seleção e limpeza dos dados, bem como uma interpretação adequada dos resultados obtidos são essenciais para garantir que o conhecimento adquirido seja útil aos pesquisadores. Assim, o passo de mineração de dados no processo de KDD consiste na aplicação de análise de dados e algoritmos de descoberta que produzem uma enumeração particular de padrões obtidos a partir de uma coleção de dados.

Ainda segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), o processo de aplicação do KDD é interativo e iterativo, envolvendo diversos passos e decisões tomadas pelo usuário. Sendo assim, os passos mais básicos são:

¹ <https://docs.google.com/forms/u/0/>

- O entendimento do domínio da aplicação e o conhecimento anterior relevante, bem como a identificação do objetivo da aplicação do **KDD** do ponto de vista do cliente.
- A criação de um conjunto de dados alvo, a partir da seleção dos dados que se deseja obter ou foco em um subconjunto de variáveis ou amostra de dados, onde a descoberta será aplicada.
- Limpeza e pré-processamento dos dados, a partir da remoção de ruído, coleta de informações necessárias para modelar e quantificar o ruído, decisão de como lidar com campos incompletos e contabilização de informações de sequência de tempo, bem como mudanças conhecidas.
- A redução e projeção dos dados são atingidas a partir da descoberta de recursos úteis para representar os dados, dependendo do objetivo da análise. Com a redução dimensional ou métodos de transformação, o número de variáveis efetivas a serem consideradas pode ser reduzido, ou se pode encontrar uma representação invariável dos dados.
- Os objetivos do processo de **KDD** devem ser combinados a um método específico de mineração de dados.
- A análise exploratória e a seleção de modelo e hipótese para a descoberta de padrões dentre os dados fornecidos.
- A mineração de dados consiste na busca de padrões de interesse em uma forma particular de representação ou um conjunto de representações, incluindo regras ou árvores de classificação, regressão e clusterização.
- Os padrões que foram minerados são interpretados, sendo possível retornar a qualquer um dos passos listados anteriormente. Esse passo também pode envolver a visualização dos padrões extraídos e modelos, bem como a visualização dos dados a partir dos modelos extraídos.
- Por fim, a partir dos dados obtidos, é possível utilizar o conhecimento adquirido de forma direta, incorporá-lo a outros sistemas ou simplesmente documentar as descobertas obtidas. Esse processo também pode incluir a comprovação ou resolução de conflitos com conhecimentos previamente obtidos.

Os passos listados por **Fayyad, Piatetsky-Shapiro e Smyth (1996)** podem ainda envolver iterações entre si, bem como laços entre dois passos específicos, dependendo do objetivo da pesquisa em questão, sendo o fluxo básico descrito pela Figura 1.

2.4.2.1 *SentiStrength*

Proveniente do artigo [Thelwall et al. \(2010\)](#), o *SentiStrength* é um software utilizado para a extração simultânea da força de sentimentos positivos e negativos em textos curtos e informais, utilizando um dicionário de palavras associadas à medição de força de sentimentos. Assim, segundo os autores, o *SentiStrength* possui saídas possíveis para a classificação, sendo elas:

- Binário: classificação positiva ou negativa, associando-se valores 1 e -1, respectivamente.
- Trinário: além das classificações pertencentes a saída binária, possui ainda a classificação neutra.
- Em escala: classificação que varia entre -4 e 4.
- Dual: objeto que possui dois valores, sendo o primeiro deles uma escala entre 1 e 5 para o sentimento positivo e o segundo uma escala entre -1 e -5 para o sentimento negativo.

Além de possuir arquivos para a identificação de *emojis*, ironias e gírias, tal *software* também possui arquivos utilizados para a realização da análise do sentimento contido em um texto curto como *EmotionLookupTable*, que relaciona palavras a um valor de classificação de sentimento; *NegativeWordList*, que possui palavras negativas, utilizadas para a inversão do sentimento de uma frase e *BoosterWordList*, que possui palavras que intensificam o sentimento de uma frase.

2.4.2.2 *Principal Component Analysis* ou PCA

De acordo com [Jolliffe e Cadima \(2016\)](#), conjuntos de dados que possuem alta dimensionalidade podem dificultar sua interpretação. Sendo assim, o *Principal Component Analysis (PCA)* surge como uma técnica de redução da dimensionalidade desses conjuntos, mantendo o máximo de variabilidade, ou informação estatística, possível. Os autores afirmam ainda que manter o máximo de variabilidade possível significa encontrar novas variáveis que são funções lineares daquelas no conjunto de dados original, que maximizavam sucessivamente a variância e não estavam correlacionadas umas as outras. Encontrar essas novas variáveis, ou os componentes principais, resume-se a resolver um problema de autovalor/autovetor.

[Abdi e Williams \(2010\)](#) afirmam que os objetivos da aplicação do *PCA* são extrair as informações mais relevantes de um conjunto de dados; diminuir o tamanho de um conjunto de dados, mantendo apenas informações importantes; simplificar a descrição de um conjunto de dados e analisar a estrutura das observações e das variáveis.

Para se atingir esses objetivos, [Abdi e Williams \(2010\)](#) destacam que, o **PCA** calcula os componentes principais, que são obtidos como combinações lineares das variáveis originais. O primeiro componente principal deve ter a maior variância possível. Já o segundo componente principal deve ser ortogonal ao primeiro componente obtido, além de possuir a maior variância possível. Os outros componentes são obtidos de maneira semelhante. Os valores dessas novas variáveis são chamados de *factor scores* e podem ser interpretados geometricamente como projeções das novas variáveis obtidas no componente principal.

Nesse sentido, a partir da aplicação do **PCA**, obtém-se uma nova matriz que organiza os componentes principais em ordem decrescente de variância. Assim, de posse dessa matriz, pode-se obter informações relevantes sobre o conjunto de dados, como *outliers* e variáveis de maior expressividade para os dados originais.

2.4.2.3 Clusterização

Segundo [Fayyad, Piatetsky-Shapiro e Smyth \(1996\)](#), a clusterização é uma tarefa descritiva que busca identificar um grupo finito de categorias ou *clusters* para descrever dados. Essas categorias podem ser mutuamente exclusivas e exaustivas ou categorias hierárquicas ou sobrepostas.

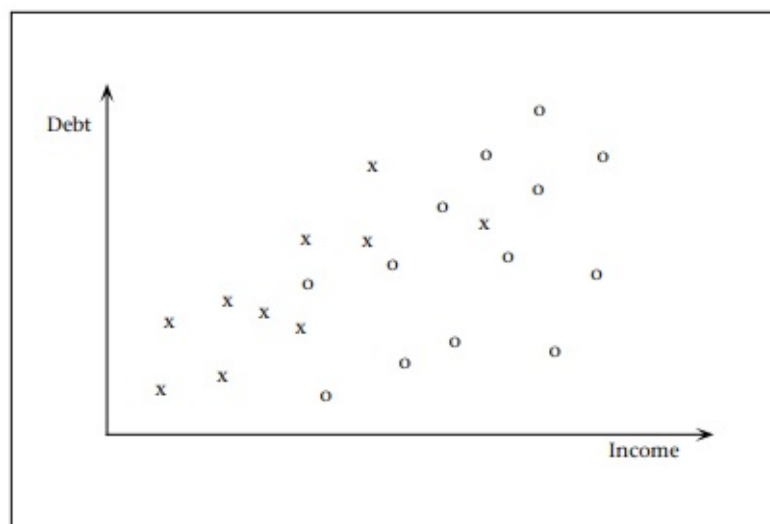


Figura 2 – Conjunto de dados com duas classes, retirado de [Fayyad, Piatetsky-Shapiro e Smyth \(1996\)](#).

A Figura 2 representa um conjunto de dados fictício de duas dimensões com 23 casos. Cada ponto no gráfico representa uma pessoa que recebeu um empréstimo de um determinado banco no passado. O eixo horizontal representa a renda dessa pessoa e o eixo vertical, o total da dívida. Os dados foram representados em duas classes, sendo que a letra *x* representa pessoas que não puderam pagar suas dívidas e a letra *o* representam pessoas com bom crédito no banco. Já a Figura 3, exemplifica uma possível clusterização desse

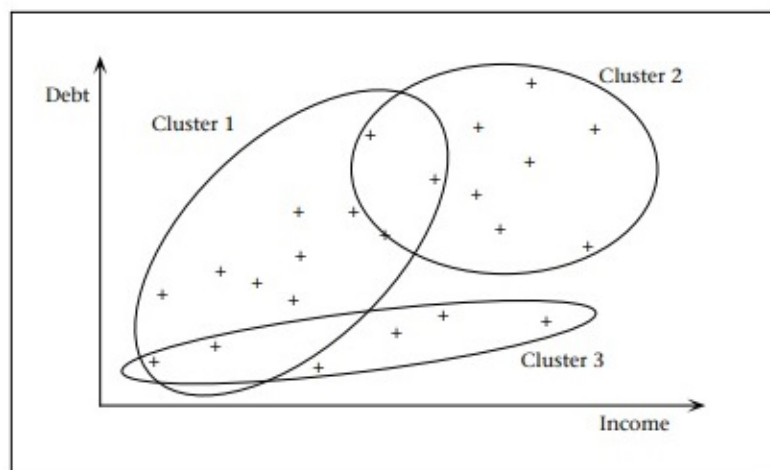


Figura 3 – Clusterização do conjunto de dados de empréstimos, retirado de [Fayyad, Piatetsky-Shapiro e Smyth \(1996\)](#).

conjunto de dados, em três grupos. Percebe-se que os *clusters* se sobrepõem, permitindo que dados pertençam a diferentes grupos simultaneamente. Nesta figura, os símbolos foram substituídos por +, o que indica que os dados não podem mais ser identificados por suas classes.

lscape

3 Desenvolvimento

Este capítulo descreve o desenvolvimento do trabalho, contendo a criação do questionário, a construção da base de dados fictícia, bem como a criação de um *framework* com as técnicas de mineração de dados relevantes.

3.1 Construção do questionário e desenvolvimento da base de dados fictícia

O questionário desenvolvido para a coleta de dados que serão utilizados para a criação do *framework*, foi subdividido em cinco seções. A primeira seção contém o Termo de Consentimento Livre e Esclarecido (TCLE), que deve ser aceito para que seja possível prosseguir com o preenchimento do questionário. Caso a respondente não aceite tal termo, o questionário é encerrado, não sendo necessário preencher nenhuma outra pergunta.

A segunda seção contém questões de cunho pessoal, como idade, cor/etnia, local de nascimento, curso de graduação escolhido, ano e período de ingresso, previsão de formatura, onde a respondente cursou o ensino médio, se a respondente realizou curso técnico na área da graduação, estado civil, se a respondente possui filhos - e em caso afirmativo, a quantidade -, forma de ingresso no curso, grau de escolaridade dos pais, quantos irmãos a respondente possui - sendo possível responder que a mesma não possui irmãos - e se a respondente trabalha.

A terceira seção compreende as questões sobre motivações para escolha do curso, como se o curso escolhido foi a primeira opção - e em caso negativo, qual era sua primeira opção de graduação -, motivação para ingresso no curso, contato com algo/alguém da área antes da graduação.

A quarta seção possui questões sobre percepções das respondentes durante o curso como quantidade de mulheres que ingressaram e permaneceram na graduação na mesma turma da respondente, o motivo para as mulheres da turma terem se separado, relação com colegas mulheres e homens, relação com professores e professoras, participação em atividades extracurriculares, descrição de situações de preconceito e discriminação que foram presenciadas pela respondente, o que a respondente mais e menos gosta no curso, se a respondente já pensou em desistir - e em caso positivo, qual o motivo -, se a respondente conhece mulheres que contribuíram para a computação - e em caso positivo, qual/quais mulher(es) -, opiniões sobre a baixa presença de mulheres na área, obstáculos enfrentados no mercado de trabalho e disponibilidade de oportunidades para mulheres.

Por fim, a quinta seção contém questões que visam analisar o comportamento das respondentes, com base nos perfis comportamentais destacados por [Marston \(1928\)](#).

Todas as questões desse questionário são obrigatórias, podendo ser de múltipla escolha, onde algumas permitem a seleção de mais de uma alternativa ou acréscimo de uma resposta personalizada, a partir da marcação da alternativa "outros" ou questões abertas, que permitem a elaboração de uma resposta curta para a questão proposta.

O questionário desenvolvido foi ilustrado no Apêndice [A](#), sendo elaborado na plataforma *Google Forms*.

3.2 Criação e pré-processamento dos dados

Para o levantamento dos dados necessários, o questionário foi preenchido 20 vezes pelos autores deste trabalho, buscando-se a maior diversidade de respostas possível.

A base de dados desenvolvida possui uma estrutura semelhante a [Tabela 1](#). Na coluna Atributo, utilizou-se a nomenclatura adotada durante o pré-processamento dos dados.

Atributo	Descrição	Exemplo de resposta
Idade	Qual a sua idade?	24
Etnia	Qual a sua cor/etnia?	Branca.
CidadeEstado	Qual seu local de nascimento? (Cidade-Estado)	Ouro Preto - MG
Curso	Qual o seu curso?	Engenharia de Computação.
AnoPeriodoIngresso	Qual ano/período do seu ingresso no curso?	2017.1
PrevisaoFormatura	Qual ano/período da sua formatura ou previsão da formatura?	2021.2
EnsinoMedio	Onde você cursou o ensino médio?	Somente em escola particular.
CursoTecnico	Você realizou curso técnico na área?	Não.
EstadoCivil	Qual o seu estado civil?	Solteira.
PossuiFilhos	Você possui filhos?	Não.
QuantosFilhos	Quantos filhos?	0

Continua na próxima página

Tabela 1 – Continuação da página anterior

Atributo	Descrição	Exemplo de resposta
FormaIngresso	Qual sua forma de ingresso no curso?	Ampla concorrência.
EscolaridadeMae	Qual o grau de escolaridade da sua mãe?	Ensino médio completo.
EscolaridadePai	Qual o grau de escolaridade do seu pai?	Ensino médio completo.
QuantosIrmaos	Você possui quantos irmãos/irmãs?	0
Trabalha	Você trabalha?	Sim, na área do curso.
PrimeiraOpcaoCurso	O seu curso foi sua primeira opção?	Não.
QualPrimeiraOpcao	Neste caso, qual foi sua primeira opção de graduação?	Engenharia Elétrica.
PorqueCurso	Por que ingressou no curso?	Motivação pessoal; Incentivo familiar.
ContatoArea	Antes de ingressar na graduação, você teve contato com algo/alguém da área?	Sim, na escola.
IngressoMulheresTurma	Quantas mulheres ingressaram na sua turma, além de você?	5
MulheresPermaneceram	Quantas destas permaneceram no curso, além de você?	2
MulheresSepararam	Na sua opinião, por que as mulheres da turma se separaram?	Suas colegas desistiram; Suas colegas não estão regulares.
ColegasMulheres	Como é sua relação com as colegas mulheres da faculdade?	Próxima; Sinto-me à vontade com elas.
ColegasHomens	Como é sua relação com os colegas homens da faculdade?	Próxima; Tenho amigos na turma.
Continua na próxima página		

Tabela 1 – Continuação da página anterior

Atributo	Descrição	Exemplo de resposta
RelacaoProfessoras	Como é sua relação com as professoras?	Sou tratada de forma igualitária; Minhas observações e perguntas são consideradas durante as aulas..
RelacaoProfessores	Como é sua relação com os professores?	Tenho uma relação mais próxima com as professoras; Sou tratada de forma igualitária.
AtividadeExtracurricular	Você participa/já participou de alguma atividade extracurricular, como projetos de extensão, iniciação científica, entre outros?	Sim.
SituacaoPreconceito	Você já sofreu e/ou presenciou alguma cena de preconceito/discriminação contra mulheres no curso? Se sim, relate brevemente.	Sim, uma colega foi descreditada durante uma aula.
MaisGostaCurso	O que você mais gosta no curso?	A ementa do curso; Aulas práticas de software.
MenosGostaCurso	O que você menos gosta no curso?	Localização do campus; Aulas práticas de hardware.
DesistirCurso	Você já pensou em desistir do curso?	Não.
MotivosDesistir	Por quais motivos você pensou em desistir do curso?	-
PerspectivasConclusao	Quais suas perspectivas para depois da conclusão do curso?	Procurar emprego na área; Fazer uma especialização.
ContribuicaoMulheres	Você conhece mulheres que contribuem/contribuíram para a evolução da computação?	Sim.
MulheresConhece	Qual/Quais mulher(es) você conhece?	Grace Hopper.

Continua na próxima página

Tabela 1 – Continuação da página anterior

Atributo	Descrição	Exemplo de resposta
BaixaPresenca	Na sua opinião, qual a explicação para a baixa presença de mulheres na computação?	Falta de conhecimento sobre curso na área; Falta de incentivo de familiares/amigos/professores, entre outros.
ObstaculoTrabalho	Na sua opinião, as mulheres enfrentam algum tipo de obstáculo para prosseguir na computação? Se sim, qual/quais?	Sim. As profissões ainda têm uma predominância masculina e as mulheres podem se sentir hostilizadas.
OportunidadesTrabalho	Na sua opinião, as mulheres têm as mesmas oportunidades que os homens na computação? Por quê?	Não, pois pode existir preconceito da parte dos recrutadores caso a empresa não possua políticas para a contratação igualitária.
Disc1	Eu sou:	Carismática e otimista.
Disc2	Prefiro focar:	No planejamento.
Disc3	Eu gosto de:	Fazer junto.
Disc4	Eu sou uma pessoa:	Compreensiva.
Disc5	Quando tenho uma ideia:	Planejo os próximos passos.
Disc6	Eu acredito que:	Paciência e um ambiente harmônico são a chave para o sucesso.
Disc7	Eu penso que:	Suas cabeças pensam melhor do que uma.
Disc8	Minha opinião sobre mudanças é:	Aceito, desde que sejam bem estruturadas e sem riscos.

Tabela 1 – Descrição da base de dados do questionário

Como primeiro passo do pré-processamento dos dados coletados, verificou-se a existência de erros no conjunto de dados, para que os mesmos pudessem ser corrigidos. Nota-se que, como a base de dados foi criada com o intuito de se desenvolver um *framework* para análise de dados, não existem dados duplicados ou respostas faltantes, fator que deve ser observado quando são utilizados dados coletados de forma orgânica.

Diversas perguntas do questionário possibilitavam a marcação de uma ou mais

alternativas. Sendo assim, cada alternativa desse tipo de questão foi transformada em um atributo para a base de dados, a fim de se construir uma matriz de incidência a partir da estratégia *bag-of-words*. Modificou-se tais atributos, para que alternativas como *Motivação pessoal* fossem substituídas por *motivação pessoal*. Em seguida, utilizando-se o método *CountVectorizer*¹ obteve-se uma matriz com 142 colunas que correspondem a cada alternativa das questões de múltipla escolha, onde o atributo possuía valor 1 caso tenha sido marcado e 0 caso contrário, como demonstrado na Figura 4.

	amarela	branca	indígena	parda	preta	ciênciasda computação	engenhariade computação	sistemasde informação
0	0	1	0	0	0	0	1	0
1	0	0	0	0	1	0	0	1
2	0	0	0	1	0	0	1	0
3	1	0	0	0	0	0	0	1
4	0	1	0	0	0	0	1	0

Figura 4 – Trecho da matriz de incidência obtida.

De forma a tratar os dados em linguagem natural presentes, retirou-se todos os caracteres não ASCII e todo o texto foi convertido para caixa baixa. Para a remoção dos caracteres não pertencentes a língua portuguesa, utilizou-se o módulo *re*² nativo da linguagem de programação *Python*, utilizado para o tratamento de expressões regulares. O método *word_tokenize* foi utilizado para realizar a tokenização das frases. Além disso, removeu-se todas as *stop words*, palavras sem relevância para o tratamento do texto, bem como as pontuações.

Destaca-se também, que, o pré-processamento desenvolvido para este trabalho se encontra pronto para a aplicação de dados verdadeiros.

3.3 Aplicação do *SentiStrength*

A fim de extrair informações relevantes a respeito do sentimento demonstrado nas respostas atribuídas para a questão "Você já sofreu e/ou presenciou alguma cena de preconceito/discriminação contra mulheres no curso? Se sim, relate brevemente.", utilizou-se o *SentiStrength*. A análise de sentimento para o português ainda se encontra em fase de testes, sendo utilizada em combinação com a análise de sentimento para o inglês.

Dessa forma, os textos obtidos foram atribuídos à variável *textos_preconceito* e, utilizando-se a biblioteca *GoogleTrans* do *Python*, realizou-se a tradução dos mesmos, sendo necessário inspecionar o arquivo obtido, a fim de verificar erros de tradução e/ou digitação. Os arquivos obtidos foram submetidos à análise de sentimento do *SentiStrength*³ para

¹ <https://bit.ly/3dQdQBp>

² <https://docs.python.org/3/library/re.html>

³ <http://sentistrength.wlv.ac.uk/>

ambos os idiomas considerados. Optou-se pelo acréscimo manual dos valores fornecidos aos arquivos de texto.

É importante destacar que, a utilização da análise de sentimento em português não é tão precisa como a análise em inglês, tendo-se em vista que os valores obtidos para a mesma frase nem sempre foram os mesmos, como observado na Tabela 2.

Frase em PT-BR	Análise em PT-BR	Frase em EN	Análise em EN
Não.	[1, -3]	No.	[1, -1]
Sim, uma colega foi descreditada durante uma aula.	[1, -1]	Yes, a classmate was discredited during a class.	[1, -2]
Sim, eu sofri uma situação de preconceito durante uma aula, onde minhas perguntas não eram consideradas, mas se os colegas homens faziam perguntas iguais e/ou parecidas, suas perguntas eram respondidas sem nenhum comentário adicional.	[1, -3]	Yes, I suffered a prejudiced situation during a class, where my questions were not considered, but if male colleagues asked the same and/or similar questions, their questions were answered without further comment.	[1, -3]
Não.	[1, -3]	No.	[1, -1]
Sim, uma colega foi ignorada completamente durante uma aula onde a mesma desejava fazer perguntas.	[1, -1]	Yes, a classmate was completely ignored during a class where she wanted to ask questions.	[1, -1]
Sim, uma colega foi ignorada durante as práticas de uma matéria por ser mulher.	[1, -1]	Yes, a colleague was ignored during the practices of a subject because she was a woman.	[1, -1]

Continua na próxima página

Tabela 2 – Continuação da página anterior

Frase em PT-BR	Análise em PT-BR	Frase em EN	Análise em EN
Sim, fui hostilizada durante uma aula por fazer perguntas que meus colegas homens fizeram e não aconteceu nada com eles.	[1, -3]	Yes, I was harassed during a class for asking questions my male colleagues asked and nothing happened to them.	[1, -3]
Sim, já senti que os professores homens olhavam diferente para mim durante as aulas.	[1, -1]	Yes, I felt that male teachers looked at me differently during classes.	[1, -1]
Não.	[1, -3]	No.	[1, -1]
Não.	[1, -3]	No.	[1, -1]
Não.	[1, -3]	No.	[1, -1]
Não.	[1, -3]	No.	[1, -1]
Não.	[1, -3]	No.	[1, -1]
Sim, fui descredita- tada durante uma aula prática, devido as minhas dúvidas.	[1, -1]	Yes, I was discredited during a practical class, due to my doubts.	[1, -2]
Sim, uma colega foi observada de forma constrangedora durante várias aulas do mesmo professor.	[1, -1]	Yes, a classmate was observed in an embarrassing way during several classes by the same teacher.	[1, -3]
Não.	[1, -3]	No.	[1, -1]
Não.	[1, -3]	No.	[1, -1]

Tabela 2 – Comparativo da análise de sentimento em português e inglês

De acordo com [Thelwall et al. \(2010\)](#), a classificação atribuída aos textos retirados do questionário, representam uma avaliação dupla, sendo o primeiro número a análise do sentimento positivo e o segundo a análise do sentimento negativo. Para o sentimento positivo, 1 representa um sentimento *não positivo* e 5, um sentimento *extremamente positivo*. Da mesma forma, para o sentimento negativo, -1 representa um sentimento *não*

negativo e -5, um sentimento *extremamente negativo*.

3.4 Transformação dos dados

Com o término do pré-processamento dos dados, iniciou-se o processo de aplicação da transformação e mineração de dados, a partir dos métodos definidos por [Fayyad, Piatetsky-Shapiro e Smyth \(1996\)](#). Com a aplicação de algoritmos de agrupamento, classificação, sumarização e regressão, é possível encontrar grupos e padrões dentro dos dados coletados.

A partir do relacionamento das variáveis restantes na base de dados, pode-se extrair conhecimento útil, aplicando-se a análise de correlação através do método de *spearman*. Devido à grande quantidade de atributos obtidos, considerando-se o número de variáveis retiradas das questões de múltipla escolha, encontrou-se um *heatmap* bastante extenso, onde cada linha e coluna nos eixos representam um dos atributos identificados, apresentado na Figura 5.

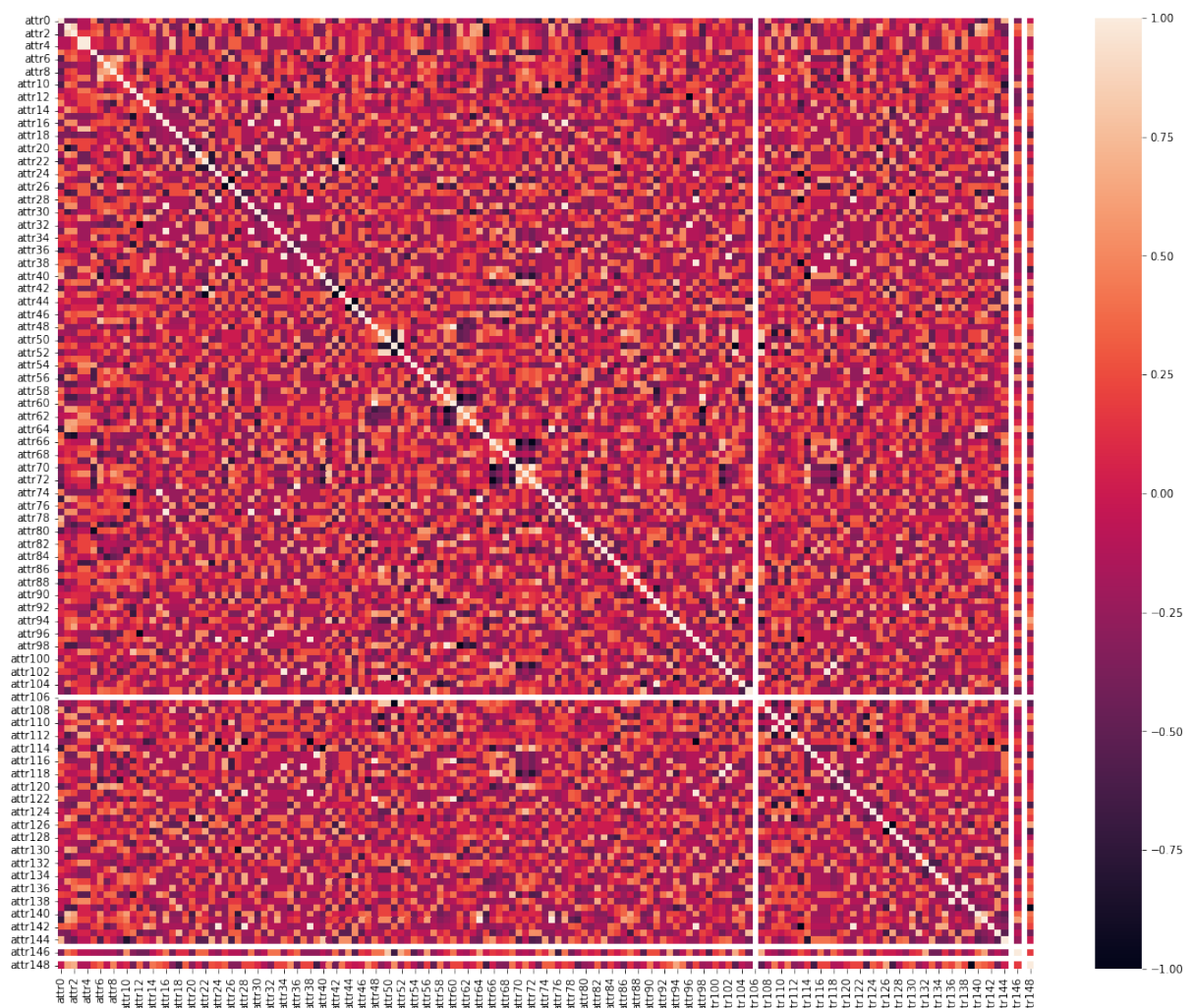


Figura 5 – Heatmap obtido.

Sendo assim, dos valores obtidos, é possível destacar as correlações positivas e negativas para alguns dos seguintes atributos:

- Correlações positivas:
 - Alunas que participam de atividade extracurriculares consideram que a baixa presença de mulheres na área de TI se deve pela falta de conhecimento sobre cursos na área.
 - Mulheres de etnia indígena ingressaram na universidade através de cotas étnico raciais e sua mãe concluiu sua educação no ensino fundamental completo.
 - Aluna estudou nas redes pública e particular durante o ensino médio e a mesma trabalha, seja na área do curso ou não.
 - Aluna estudou somente na escola pública durante o ensino médio e aluna trabalha, o pai cursou alguma pós-graduação e mesma prefere focar no planejamento.
 - Aluna ingressou na universidade a partir de cotas para rede pública e pai possui pós graduação, ela prefere focar no planejamento e aceita mudanças desde que sejam bem estruturadas e sem riscos.
 - Aluna ingressou na universidade a partir de cotas étnico-raciais e sua mãe possui apenas o ensino fundamental completo.
 - Mãe da aluna possui apenas ensino fundamental incompleto e a aluna se considera ousada e direta.
 - Pai da aluna possui apenas ensino fundamental incompleto e a aluna se considera ousada e direta.
 - Aluna teve contato com a área de TI e se considera detalhista e organizada, prefere focar no planejamento e dá preferência para atividades onde pode agir individualmente.
 - Aluna se considera detalhista e organizada e prefere agir individualmente.
 - Aluna prefere focar no planejamento e aceita mudanças apenas quando as mesmas são bem estruturadas e sem riscos.
- Correlações negativas:
 - Existe uma baixa correlação entre alunas que acreditam que mulheres encontram obstáculos no mercado de trabalho e suas mães estudaram até o ensino superior incompleto.
 - Aluna cursou o ensino médio nas redes pública e particular e trabalha durante a graduação, seja na área do curso ou não.

- Aluna se encontra em uma união estável e acredita que o baixo interesse das mulheres pelos cursos da área de TI se deve pelo preconceito ou aluna sonha alto a respeito de seu futuro.
- Aluna ingressou na universidade através de cotas para a rede pública e acredita que o baixo interesse das mulheres pelos cursos de TI ocorre pelo preconceito.
- Aluna trabalha na área do curso e se considera uma pessoa acolhedora e paciente.
- Aluna trabalha durante a graduação e escolheu tal área devido à boas perspectivas para o mercado de trabalho.
- Aluna escolheu sua graduação por incentivo familiar e escolar.
- Aluna acredita que o baixo interesse das mulheres pelos cursos da área de TI se deve pelo preconceito, prefere focar no planejamento e aceita mudanças desde que sejam bem estruturadas e sem riscos.

É importante destacar que, como a base de dados utilizadas foi criada a partir de dados fictícios, as correlações obtidas não refletem as opiniões de discentes de uma universidade ou campus específico. Tais correlações apenas exemplificam os valores que podem ser obtidos a partir da aplicação das técnicas de mineração utilizadas para a construção deste trabalho.

Considerando a alta dimensionalidade do conjunto de dados obtido, se faz necessário a aplicação a técnica de PCA, com objetivo de se reduzir a quantidade de variáveis analisadas, bem como a identificação das variáveis que melhor representam a variabilidade dos dados. A fim de se escolher a quantidade de componentes principais que representarão os dados, utilizou-se a métrica *scree plot* que, de acordo com Kanyongo (2005), plota autovalores em ordem decrescente e examina a ocorrência de quebra de nivelamento na linha tracejada, sendo o número ideal obtido através da contagem de componentes antes do ponto de quebra. Obteve-se o gráfico da Figura 6 com a aplicação do *scree plot*. Tal gráfico ilustra a quantidade de componentes principais em relação ao auto valores. Optou-se pela utilização de 9 componentes principais, já que existe um desnível observável para esta quantidade de componentes na Figura 6.

Para a visualização dos padrões encontrados nas respostas do questionário após a redução de dimensionalidade, os componentes principais obtidos foram aplicados no algoritmo de agrupamento *k-means*⁴. Para a utilização de tal algoritmo, é necessário escolher previamente a quantidade ideal de *clusters* que melhor definem o conjunto de dados.

Primeiramente, utilizou-se o *elbow method* ou método do cotovelo para avaliar a qualidade da clusterização. Bholowalia e Kumar (2014) afirmam que o método do cotovelo

⁴ <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

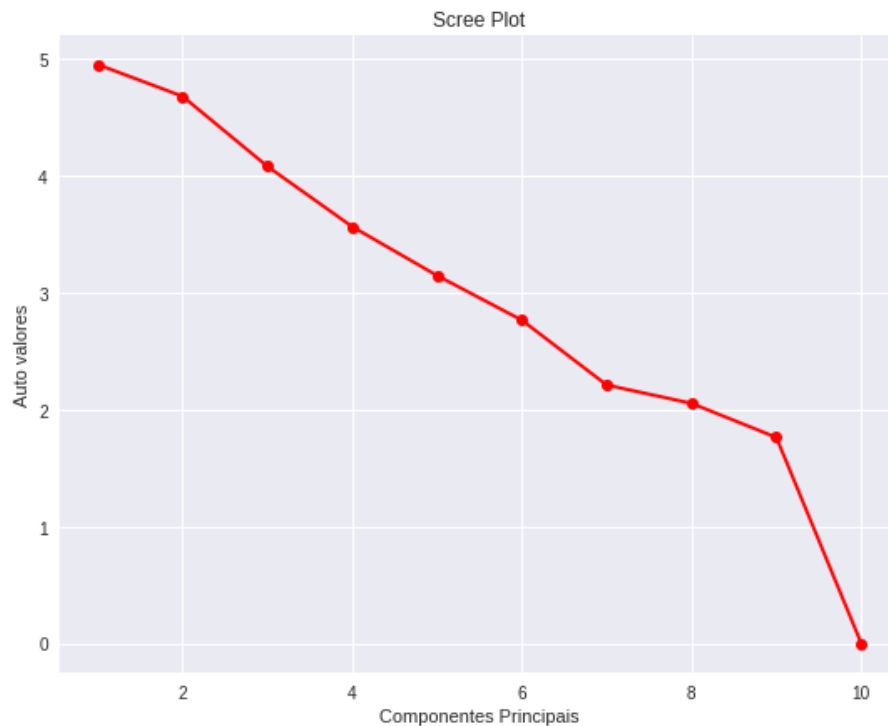


Figura 6 – Gráfico obtido a partir da utilização do *scree plot*.

se baseia na análise da porcentagem da variância explicada através de uma função do número de clusters escolhido. Assim, o número de *clusters* ideal é encontrado quando, ao se acrescentar mais um *cluster*, não é possível obter uma melhor modelagem da base de dados.

Destaca-se que, no *framework* desenvolvido, é possível visualizar a contribuição de cada um dos atributos encontrados para os componentes principais, a partir da sobreposição da seta do *mouse* sobre o atributo desejado. Como se possui um número bastante elevado de atributos, a análise da contribuição dos mesmos para cada componente é trabalhosa. Dessa forma, para cada um dos nove componentes principais, as Tabela 3 ilustram os cinco atributos mais significativos:

	1º atributo	2º atributo	3º atributo	4º atributo	5º atributo
1	sãonecessárias	somenteescolaparticular	procurarumempregonaárea	narepúblicaeparticular	não
2	professores	naáreadocurso	não	seguircarreiraacadêmica	emresultados
3	ensinomédiocompleto	engenhariadecomputação	tomodecisõesimpulsivas	tercontrole	incentivofamiliar
4	delegar	possuifilhos	quantosfilhos	meatentoaosdetalhesajocomcautela	suascolegasnãoestãoregulares
5	naspressoasomeuredor	ousadaedireta	localizaçãodocampus	motivaçãopessoal	atividadextracurricular
6	transferênciainternaexterna	entreoutros	aulapráticascomsoftware	localizaçãodocampus	professores
7	compreensiva	questionadora	amarela	duascabeçaspensammelhordoqueuma	parda
8	atividadesextracurriculares	colegas	fazerumaespecialização	naspressoasomeuredor	baixaperspectivasalarial
9	acolhedoraepaciente	feitoémelhorqueperfeito	negative_sent_pt	aulapráticascomsoftware	ingressomulheresturma

Tabela 3 – Atributos mais significativos para cada um dos nove componentes.

Inicia-se a análise para uma quantidade de *clusters* de $k = 2$, onde são realizadas iterações para um k crescente em uma unidade até que o ganho de informação, representado pela soma dos quadrados intra-clusters ou *within-clusters sum-of-squares* (*WCSS*), sofra uma queda abrupta, representada por um ângulo acentuado no gráfico. O resultado da

aplicação do método do cotovelo é ilustrado pela Figura 7.

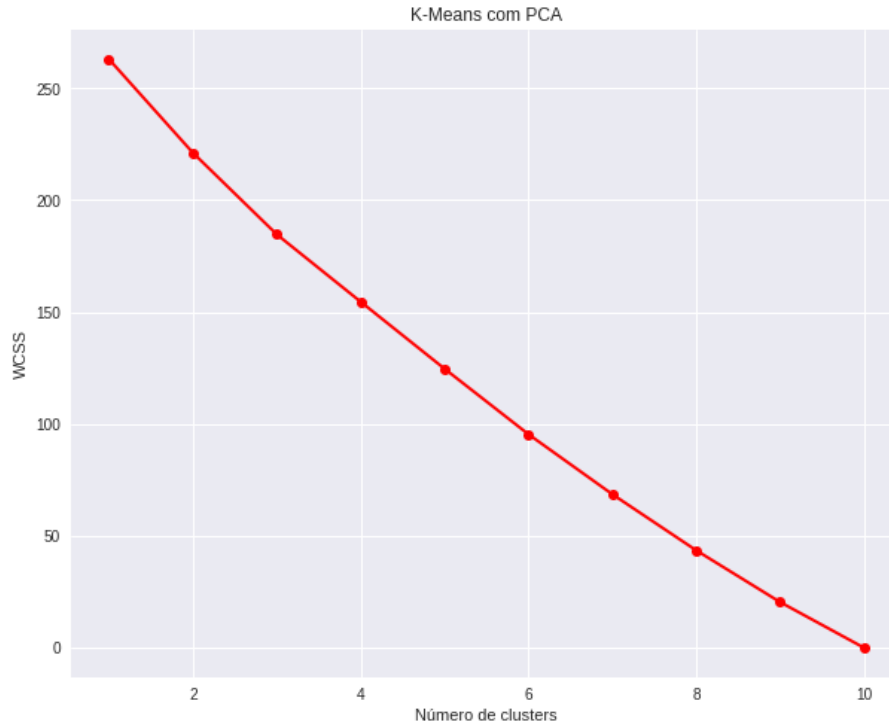


Figura 7 – Gráfico obtido a partir da utilização do método do cotovelo.

Em seguida, utilizou-se o *silhouette coeficient*, a fim de se obter uma melhor análise e compará-la ao método do cotovelo aplicado anteriormente. De acordo com [Struyf, Hubert e Rousseeuw \(1997\)](#), o *silhouette plot* permite a seleção do número de *clusters*. Para cada objeto i existe um *cluster* que o contém, que pode ser calculado por $a(i)$, onde tal valor é igual a dissimilaridade média de i para todos os outros objetos de A :

$$a(i) := \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j)$$

Da mesma forma, considerando um determinado *cluster* C diferente de A , temos $d(i, C)$ sendo igual dissimilaridade média de i para todos os objetos de C :

$$d(i, C) := \frac{1}{|C|} \sum_{j \in C} d(i, j)$$

Após realizarmos o cálculo de $d(i, C)$ para todos os *clusters* $C \neq A$, adotamos o menor *cluster* desse subconjunto, a partir de:

$$b(i) := \min_{C \neq A} d(i, C)$$

Tal valor é representado pelo *cluster* B , sendo chamado de *vizinho* do objeto i . Esse é o segundo melhor *cluster* para o objeto i .

O *silhouette value* $s(i)$ do objeto i é definido como:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

É possível perceber que o valor de $s(i)$ está sempre entre -1 e 1, sendo seu valor interpretado da seguinte forma:

- $s(i) \approx 1 \Rightarrow$ o objeto i foi bem classificado em A .
- $s(i) \approx 0 \Rightarrow$ o objeto i se encontra entre os *clusters* A e B .
- $s(i) \approx -1 \Rightarrow$ o objeto i foi mal classificado em A , estando mais próximo de B .

Assim, a partir das informações apresentadas anteriormente, tem-se que a média do *silhouette coefficient* entre 2 e 9 *clusters* é ilustrada pela Tabela 4:

Quantidade de <i>clusters</i>	Média do <i>silhouette coefficient</i>
2	0.0562
3	0.0353
4	0.0433
5	0.0441
6	0.0426
7	0.0399
8	0.0305
9	0.0194

Tabela 4 – Valores da média do *silhouette coefficient*.

Nesse sentido, de acordo com as duas métricas utilizadas, optou-se pela adoção de 2 *clusters*. É prudente destacar que, toda metodologia para a escolha dos *clusters* e dos componentes principais são apresentados no *framework* desenvolvido, sendo possível aplicar uma base de dados real.

4 Resultados

Este capítulo descreve os resultados obtidos através da mineração da base de dados fictícia inserida no *framework* desenvolvido, que pode ser encontrado nos seguintes links^{1,2}.

4.1 Análise dos resultados

Ao fim da etapa de mineração de dados, com a obtenção e rotulação de cada *cluster* identificado, iniciou-se a interpretação dos resultados.

Como ilustrado na Figura 8, 70% das respondentes foram agrupadas no *cluster* A, o que corresponde a 14 observações e 30% das respondentes foram agrupadas no *cluster* B, o que corresponde a 6 observações.

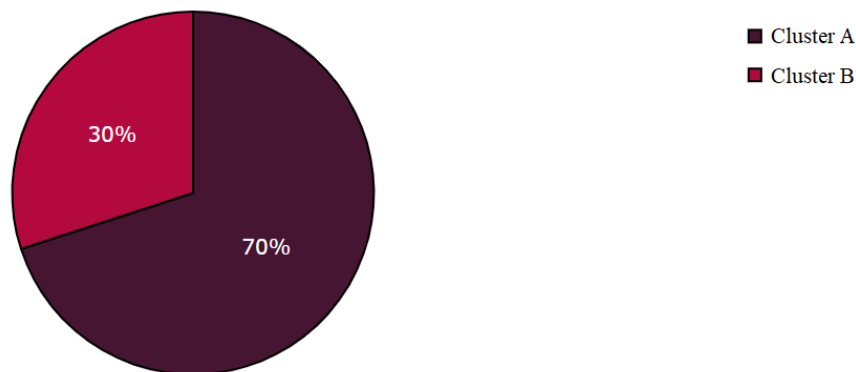


Figura 8 – Quantidade de observações em cada *cluster*.

Para a Figura 9, é possível perceber que o *cluster* que contém o maior número de observações, *cluster* A, possui uma menor dispersão de idades, com alta concentração de alunas entre 25 e 30 anos, sendo 28 anos a mediana das idades. Já o *cluster* B, possui uma maior dispersão, com maior concentração entre 20 e 25 anos, com uma mediana de 22 anos.

Em uma análise do atributo ano de ingresso, na Figura 10, nota-se que o *cluster* A possui uma maior concentração de alunas que ingressaram na faculdade entre os anos de 2014 e 2018, sendo a mediana o ano de 2016. O *cluster* B, entretanto, possui uma maior concentração de respondentes entre os anos de 2018 e 2022, sendo a mediana deste *cluster* o ano de 2018.

A partir dos gráficos obtidos para a concentração de alunos em determinados *clusters* de acordo com algumas características, levanta-se as seguintes observações, sendo prudente

¹ <https://bit.ly/3FXM6at>

² <https://bit.ly/3f0ku8E>



Figura 9 – Distribuição de idade em cada *cluster*.



Figura 10 – Distribuição do ano de ingresso em cada *cluster*.

destacar que todos os dados obtidos não refletem a distribuição das discentes de uma determinada universidade ou campus, pois a base de dados aplicada no processo descrito por este trabalho foi construída de forma fictícia. É possível perceber que na Figura 11, no *cluster* A, 100% das respondentes cursaram o ensino médio apenas na rede particular. Entretanto, esse índice representa apenas 42.9% das discentes no *cluster* B. Além disso, como ilustrado na Figura 12, no *cluster* A, 100% das alunas cursaram o ensino médio nas redes pública e particular. Já no *cluster* B, esse valor corresponde a 42.9% das respostas.

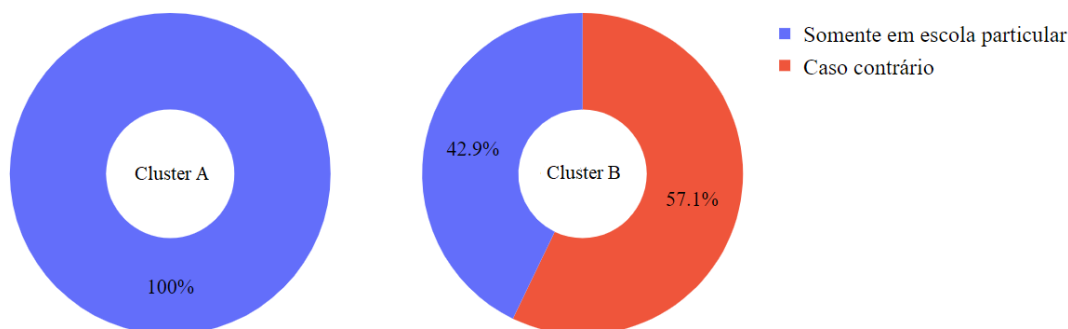


Figura 11 – Respondentes que cursaram o ensino médio na rede particular.

Sobre as percepções das respondentes sobre a universidade, como estrutura, profes-

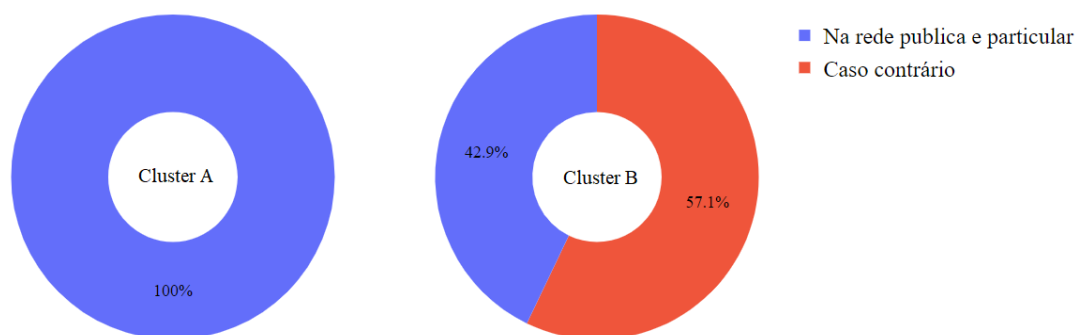


Figura 12 – Respondentes que cursaram o ensino médio nas redes pública e particular.

sores, presença de atividades extracurriculares, localização do campus, entre outros, na Figura 13, no *cluster A*, 100% das respondentes afirmam não gostarem de participar de atividades extracurricular, sendo esse valor igual a 28.6% para o *cluster B*.

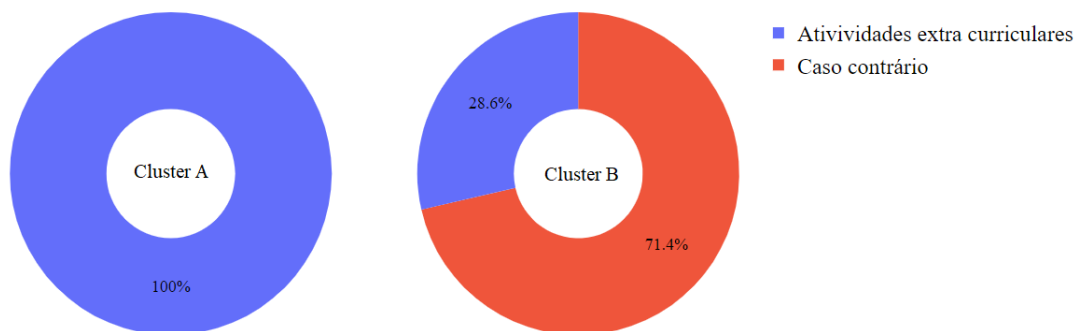


Figura 13 – Grupos identificados dentre o que as alunas menos gostam na universidade.

Ainda sobre percepções, para os dois *clusters* identificados nas Figuras 14 e 15, é possível perceber que para o *cluster A*, 100% das respondentes gostam da infraestrutura do campus e 42.9% gostam dos professores que lecionam disciplinas do curso. Da mesma forma, para o *cluster B*, apenas 14.3% das discentes gostam da infraestrutura da universidade e 57.1% dizem gostar dos professores da faculdade.

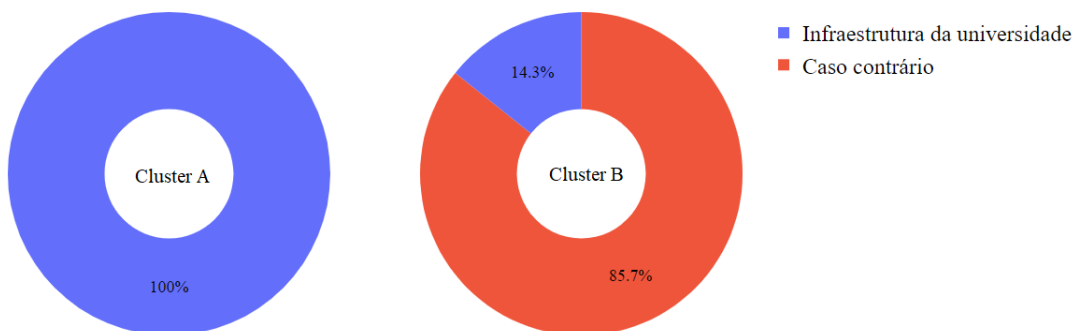


Figura 14 – Grupos de alunas que gostam da infraestrutura da faculdade.

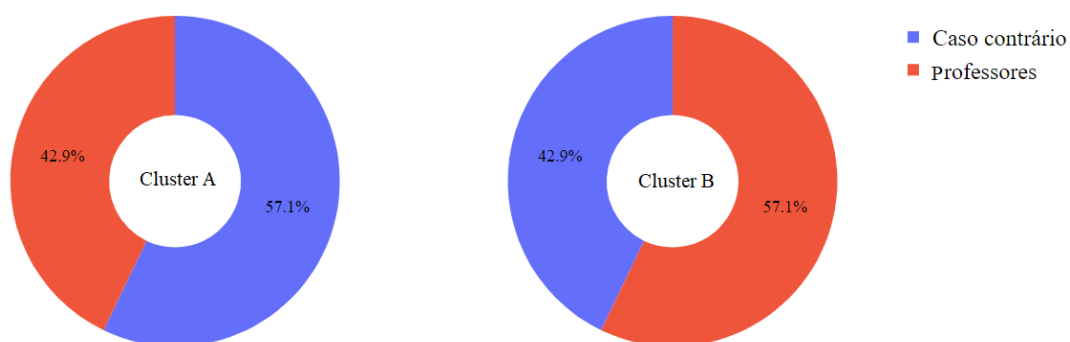


Figura 15 – Grupos de alunas que gostam dos professores.

No agrupamento de alunas que trabalham ou não durante a graduação, na Figura 16, para o *cluster A*, 100% das respondentes afirmam não trabalhar ou trabalhar em uma área que não se relaciona à graduação de alguma forma. Já para o *cluster B*, apenas 42.9% das respondentes disseram o mesmo.

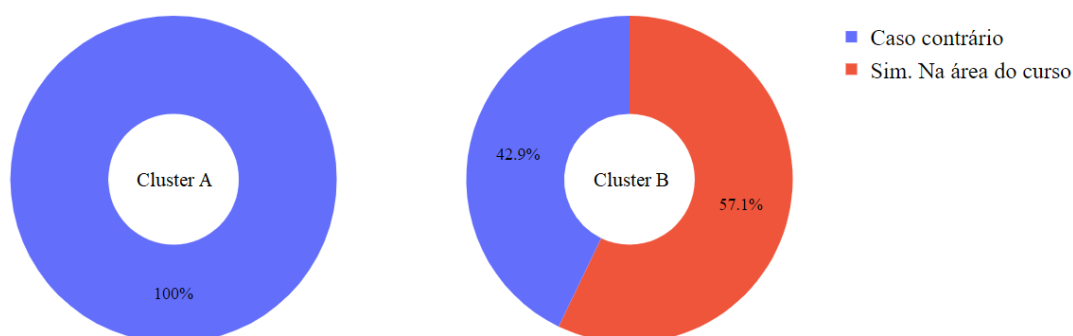


Figura 16 – Grupos de alunas que trabalham na área do curso.

É notório também, que, na Figura 17, a maioria das respondentes consideram seguir a carreira acadêmica para ambos os *clusters* identificados. No *cluster A*, 66.7% considera continuar sua educação a fim de continuarem no meio acadêmico, sendo esse valor igual a 57.1% no *cluster B*.

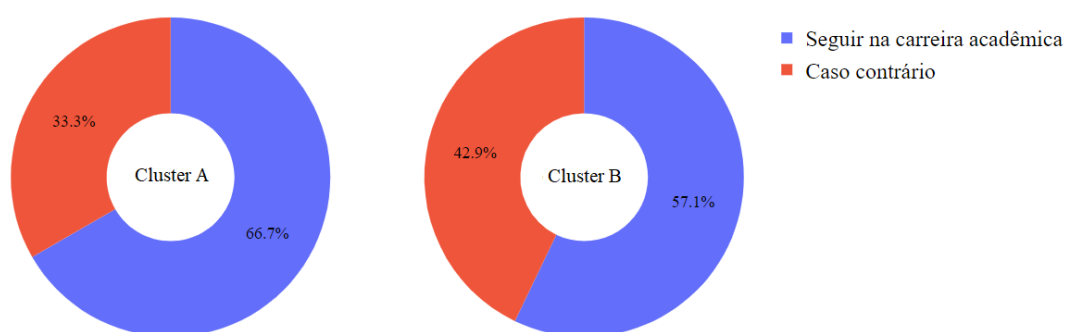


Figura 17 – Perspectivas para após a graduação.

Por fim, dentre os grupos de respondentes para características pessoais, na Figura 18, no *cluster* A, 33.3% das alunas disseram que seus pais possuem grau de escolaridade superior completo. Entretanto, para o *cluster* B, esse valor corresponde a 28.6% das respostas.

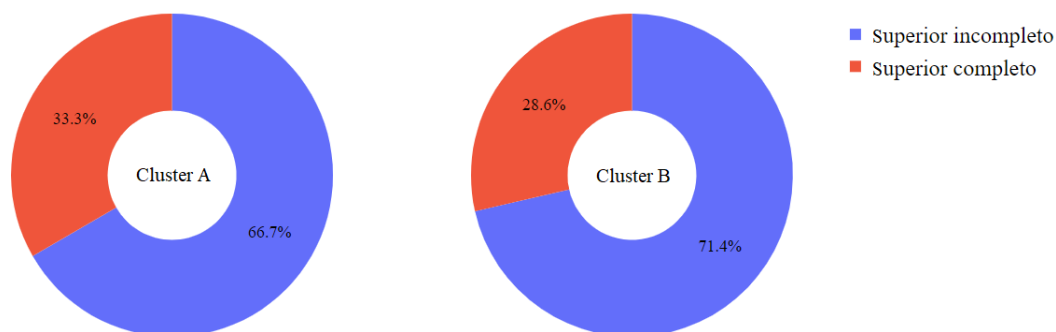


Figura 18 – Grau de escolaridade do pai.

Sobre os perfis identificados por Marston (1928), para a primeira pergunta analisada, ilustrada pela Figura 19, sobre preferências a respeito da realização de atividades, no *cluster* A, 100% das respondentes disseram que preferem realizar atividades em conjunto, o que, de acordo com o autor, corresponderia ao perfil *S ou estabilidade*. Já para o *cluster* B, apenas 28.6% das respondentes assinalaram a mesma alternativa.

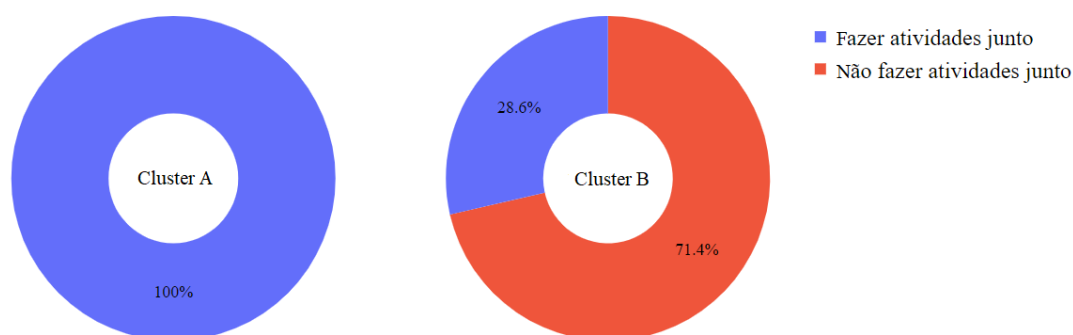


Figura 19 – Grupos dentre os perfis DISC para realização de atividades em grupo.

Da mesma forma, para a segunda questão analisada, representada pela Figura 20, sobre onde as respondentes preferem focar sua atenção, para o *cluster* A, 100% das alunas disseram que preferem focar em resultados, planejamentos ou nas pessoas ao seu redor. Já no *cluster* B, 57.1% das discentes disseram que preferem focar nas tarefas que precisam ser realizadas, o que, de acordo com Marston (1928), corresponde ao perfil *C ou conformidade*.

Finalmente, para a média de sentimento dos textos sobre situações de preconceito enfrentadas ou presenciadas pelas respondentes, realizou-se uma análise de acordo com a média dos sentimento em cada *cluster*, de acordo com a Figura 21. Para o *cluster* A, possui uma média de sentimentos negativos para ambas as línguas analisadas durante

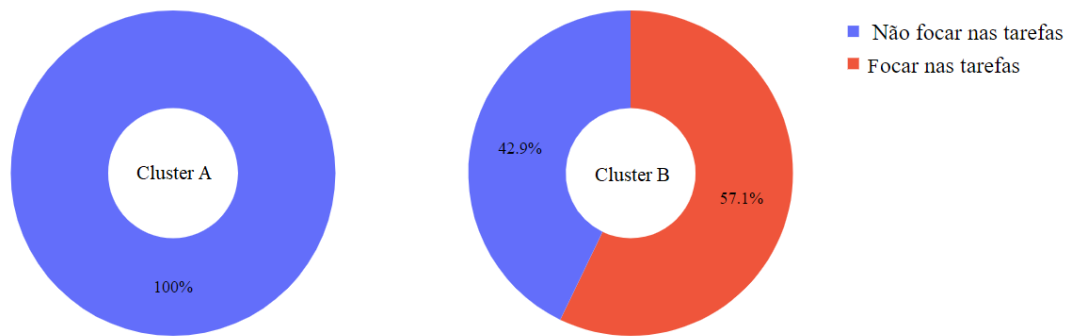


Figura 20 – Grupos dentre os perfis DISC para preferência de foco.

este trabalho, e o *cluster* B possui uma média de sentimentos negativos para a língua portuguesa. Já para os sentimentos negativos, em ambos os *clusters*, a média de sentimentos positivos para o português e inglês se mantém estável. Vale desatacar que, como ilustrado na Tabela 2, a análise de sentimentos para o português e inglês apresentou-se diferenças significativas quando a respondente disse nunca ter sofrido ou presenciado situações de preconceito, sendo o índice de sentimento negativo para a língua portuguesa igual a -3 e para a língua inglesa, igual a -1.

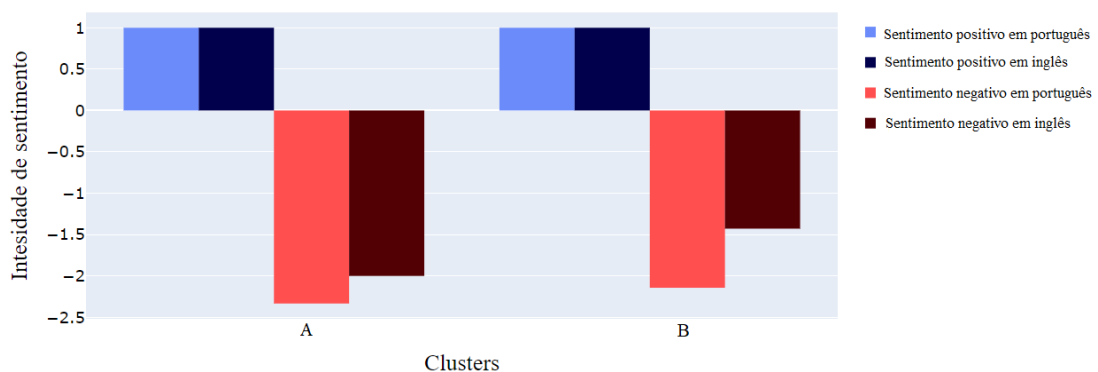


Figura 21 – Média de sentimento em cada *cluster*.

5 Conclusão

Neste capítulo são apresentadas as considerações finais sobre o trabalho desenvolvido, bem como os resultados obtidos. São descritas também, possibilidades para trabalhos futuros, identificados durante a construção deste trabalho.

Este trabalho surgiu da necessidade de compreender os motivos para a diminuição das discentes nas graduações relacionadas a áreas de TI. Apesar da presença expressiva durante o início da computação, é notório que o interesse das mulheres por cursos relacionados a essa área é decrescente. Neste contexto, a partir das pesquisas existentes a cerca da participação de mulheres na computação, em especial, trabalhos que propuseram questionários para coletar respostas de alunas, foi possível levantar todas as metodologias necessárias para realizar as análises propostas.

Assim, a partir de uma base de dados fictícia, foram aplicadas técnicas de mineração de dados, a fim de se obter conhecimentos relevantes acerca dos dados. Como resultado, foi possível perceber como diferentes respondentes podem ser agrupadas em *clusters* de acordo com o que consideram mais ou menos importante, sendo prudente destacar a análise de sentimento de determinados textos obtidos através de questões abertas.

Como continuidade deste trabalho, com base no questionário formulado contido no Apêndice A, será possível coletar dados verdadeiros, visando a proteção das respondentes de acordo com a LGPD. Assim, o pré-processamento, análise de compreensão dos resultados obtidos, torna-se ainda mais relevante, principalmente para a identificação de grupos dentre as discentes de determinadas instituições de ensino, bem como criação de metodologias para tornar o meio acadêmico o mais agradável possível para esses grupos.

Referências

- ABDI, H.; WILLIAMS, L. J. Principal component analysis. *WIREs Computational Statistics*, v. 2, n. 4, p. 433–459, 2010. Disponível em: <<https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wics.101>>. Citado 2 vezes nas páginas 23 e 24.
- BAUM, J. *The Calculating Passion of Ada Byron*. [S.l.]: Archon Books, 1986. Citado na página 14.
- BHLOWALIA, P.; KUMAR, A. Ebc-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications*, v. 105, n. 9, p. 17–24, 2014. Citado na página 37.
- CURSINO, A. R.; MARTINEZ, J. F. P. Análise estatística descritiva e regressão da inserção das mulheres nos cursos de ti nos anos de 2009 a 2018. 2021. Disponível em: <<https://sol.sbc.org.br/index.php/wit/article/view/15838>>. Acesso em: 10 ago. 2021. Citado na página 19.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. 1996. Disponível em: <<https://ojs.aaai.org/index.php/aimagazine/article/view/1230>>. Citado 6 vezes nas páginas 9, 21, 22, 24, 25 e 35.
- FREITAS, M. J. T. *Mulheres na computação: experiências, trajetórias e perspectivas de estudantes universitárias*. Dissertação (Mestrado) — Universidade Federal da Paraíba - UFPB, 2019. Citado 2 vezes nas páginas 15 e 19.
- GAMA, R. L. S. et al. “vai, meninas!” - uma iniciativa para fomentar a participação das mulheres nos cursos da área de tecnologia da informação. 2021. Disponível em: <<https://sol.sbc.org.br/index.php/wit/article/view/15883>>. Acesso em: 10 ago. 2021. Citado 2 vezes nas páginas 15 e 19.
- GIL, A. C. *Métodos e Técnicas de Pesquisa Social*. [S.l.]: Editora Atlas S.A., 2008. Citado 2 vezes nas páginas 20 e 21.
- HAMMERMAN, R.; RUSSELL, A. L. *Ada’s Legacy: Cultures of Computing from the Victorian to the Digital Age*. [S.l.]: Morgan Claypool Publishers - ACM Books, 2015. Citado 2 vezes nas páginas 14 e 15.
- JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 374, n. 2065, p. 20150202, 2016. Disponível em: <<https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2015.0202>>. Citado na página 23.
- KANYONGO, G. Determining the correct number of components to extract from a principal components analysis: A monte carlo study of the accuracy of the scree plot. *Journal of Modern Applied Statistical Methods*, v. 4, p. 120–133, 05 2005. Citado na página 37.

- LAUEN, A. *The Idolization of Ada Lovelace and Its Necessity*. 2021. Disponível em: <<https://scholarworks.uni.edu/cgi/viewcontent.cgi?article=1475&context=hpt>>. Citado na página 14.
- MARSTON, W. M. *Emotions of Normal People*. [S.l.]: London, K. Paul, Trench, Trubner & Co. Ltd.; New York, Harcourt, Brace and Company, 1928. Citado 3 vezes nas páginas 20, 28 e 45.
- MOREIRA, J. A.; SILVA, R. M.; CARVALHO, M. E. P. Cenários prospectivos: Uma visão do futuro da presença feminina em cursos de ciência da computação de uma instituição de ensino superior. In: *Anais do XXVI Workshop sobre Educação em Computação*. Porto Alegre, RS, Brasil: SBC, 2018. ISSN 2595-6175. Disponível em: <<https://sol.sbc.org.br/index.php/wei/article/view/3519>>. Citado na página 14.
- SANTOS, M. P.; SANTOS, C. P.; ELLWANGER, C. A computação sob a ótica de meninas do ensino médio. 2021. Disponível em: <<https://sol.sbc.org.br/index.php/wit/article/view/3404>>. Acesso em: 10 ago. 2021. Citado na página 18.
- SANTOS, V. L. A.; CARVALHO, T. F. M.; BARRETO, M. d. S. V. Mulheres na tecnologia da informação: Histórico e cenário atual nos cursos superiores. 2021. Disponível em: <<https://sol.sbc.org.br/index.php/wit/article/view/15847>>. Acesso em: 24 jul. 2021. Citado na página 14.
- STRUYF, A.; HUBERT, M.; ROUSSEEUW, P. Clustering in an object-oriented environment. *Journal of Statistical Software*, v. 1, n. 4, p. 1–30, 1997. Citado na página 39.
- THELWALL, M. et al. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, v. 61, p. 2544–2558, 2010. Citado 2 vezes nas páginas 23 e 34.
- WAZLAWICK, R. S. *Metodologia de Pesquisa para Ciência da Computação*. [S.l.]: Editora GEN LTC, 2009. Citado na página 16.

Apêndices

APÊNDICE A – Questionário elaborado

As questões desenvolvidas para o questionário foram subdivididas da seguinte forma:

A.1 TCLE

Prezada participante,

Você está sendo convidada para participar como voluntária do projeto de pesquisa intitulado “Análise e identificação dos grupos de mulheres da área da computação da UFOP e suas características, mediante técnicas de mineração de dados”, que pretende identificar os perfis das mulheres da Universidade Federal de Ouro Preto dos cursos de Ciência da Computação, Engenharia da Computação e Sistemas de Informação aplicando técnicas de mineração de dados às características gerais, pessoais, fatores comportamentais e de conhecimento sobre a área. Esta pesquisa é vinculada ao Instituto de Ciências Exatas e Aplicadas da Universidade Federal de Ouro Preto (ICEA/UFOP). A pesquisadora responsável por este projeto de pesquisa é Janniele Aparecida Soares Araújo, que poderá ser contatada a qualquer tempo através do e-mail janniele@ufop.edu.br.

Sua participação é possível pois você atende aos critérios de inclusão previstos na pesquisa, os quais são: ser discente ou docente do sexo feminino vinculada aos cursos da área de computação da UFOP. Sua participação consiste em responder o questionário online disponibilizado através dos e-mails e canais oficiais de comunicação da UFOP, que ficará disponível até dezembro de 2021, podendo ser estendido o período.

Nessa condição, é possível que alguns desconfortos aconteçam, como por exemplo a exposição a perguntas que remetem a algum momento desagradável durante o período da sua permanência dentro da universidade. Estes desconfortos, se ocorrerem, serão minimizados da seguinte forma: o participante pode a qualquer momento deixar de responder às perguntas. Existem riscos relacionados à coleta dos dados no ambiente virtual. Estes riscos, se ocorrerem, serão minimizados da seguinte forma: não são coletados dados sensíveis de acordo com a LGPD e a elaboração do questionário atende às regras dos Procedimentos que devem ser adotados para pesquisas em ambientes virtuais da UFOP (https://comitedeetica.ufop.br/sites/default/files/comitedeetica/files/orientacoes_para_procedimentos_em_pesquisas_com_qualquer_etapa_em_ambiente_virtual.pdf?m=1614288626). A sua participação trará benefícios, como contribuir para uma melhor escolha do curso de graduação das novas ingressantes e retenção das mesmas nos cursos, uma vez que as novas ingressantes poderão se identificar com algum perfil

traçado nesta pesquisa.

Para sua participação nessa pesquisa você não terá nenhuma despesa com transporte, alimentação, exames, materiais a serem utilizados ou despesas de qualquer natureza. A atual pesquisa não arcará com o custo de acesso à internet para que a voluntária responda ao questionário. Ao final da pesquisa você terá acesso aos resultados através dos canais oficiais da ufop e através da divulgação do trabalho de conclusão de curso da discente envolvida.

Pelo presente Termo de Consentimento Livre e Esclarecido eu declaro que autorizo a minha participação neste projeto de pesquisa, pois fui informado/a, de forma clara e detalhada, livre de qualquer forma de constrangimento e coerção, dos objetivos, da justificativa e dos procedimentos que serei submetido, dos riscos, desconfortos e benefícios, assim como das alternativas às quais poderia ser submetido, todos acima listados.

Fui, igualmente, informado/a:

a) da garantia de receber resposta a qualquer pergunta ou esclarecimento a qualquer dúvida acerca dos procedimentos, riscos, benefícios e outros assuntos relacionados com a pesquisa;

b) da liberdade de retirar meu consentimento, a qualquer momento, e deixar de participar do estudo, sem que isto traga prejuízo à continuação de meu cuidado e tratamento;

c) da garantia de que não serei identificado quando da divulgação dos resultados e que as informações obtidas serão utilizadas apenas para fins científicos vinculados ao presente projeto de pesquisa;

d) do compromisso de proporcionar informação atualizada obtida durante o estudo; ainda que esta possa afetar a minha vontade em continuar participando.

O Comitê de Ética em Pesquisa responsável pela apreciação do projeto pode ser consultado, para fins de esclarecimento.

1. De acordo com o termo de consentimento apresentado:

- a) Estou de acordo com o termo de consentimento, e desejo prosseguir com o questionário.
- b) Não estou de acordo com o termo de consentimento, e não desejo responder ao questionário.

A.2 Informações pessoais

1. Qual a sua idade?

-
2. Qual a sua cor/etnia?
 - a) Branca.
 - b) Parda.
 - c) Preta.
 - d) Amarela.
 - e) Indígena.
 - f) Outros.
 3. Qual seu local de nascimento? (Cidade-Estado)
 4. Qual o seu curso?
 - a) Engenharia de Computação.
 - b) Sistemas de Informação.
 - c) Ciências da Computação.
 5. Qual ano/período do seu ingresso no curso?
 6. Qual ano/período da sua formatura ou previsão da formatura?
 7. Onde você cursou o ensino médio?
 - a) Somente em escola pública.
 - b) Somente em escola particular.
 - c) Na rede pública e particular.
 8. Você realizou curso técnico na área?
 - a) Sim.
 - b) Não.
 9. Qual o seu estado civil?
 - a) Solteira.
 - b) Casada.
 - c) Divorciada.
 - d) Em união estável.
 - e) Viúva.
 10. Você possui filhos?
 - a) Sim.

b) Não.

11. Quantos filhos?

a) 1.

b) 2.

c) 3.

d) 4.

e) 5 ou mais.

12. Qual sua forma de ingresso no curso?

a) Ampla concorrência.

b) Cota étnico racial.

c) Cota para a rede pública.

d) Cota para baixa renda e rede pública.

e) Cota para pessoas com deficiência.

f) Transferência.

g) Reingresso.

h) Outros.

13. Qual o grau de escolaridade da sua mãe?

a) Não sabe.

b) Não frequentou a escola.

c) Fundamental incompleto.

d) Fundamental completo.

e) Ensino médio incompleto.

f) Ensino médio completo.

g) Ensino superior incompleto.

h) Ensino superior completo.

i) Pós-graduação.

14. Qual o grau de escolaridade do seu pai?

a) Não sabe.

b) Não frequentou a escola.

c) Fundamental incompleto.

- d) Fundamental completo.
 - e) Ensino médio incompleto.
 - f) Ensino médio completo.
 - g) Ensino superior incompleto.
 - h) Ensino superior completo.
 - i) Pós-graduação.
15. Você possui quantos irmãos/irmãs?
16. Você trabalha?
- a) Não.
 - b) Sim, na área do curso.
 - c) Sim, mas não na área do curso.

A.3 Motivações para a escolha do curso

1. O seu curso foi sua primeira opção?
 - a) Sim.
 - b) Não.
2. Neste caso, qual foi sua primeira opção de graduação?
3. Por que ingressou no curso?
 - a) Motivação pessoal.
 - b) Boas perspectivas para o mercado de trabalho.
 - c) Incentivo familiar.
 - d) Incentivo escolar.
 - e) Outros.
4. Antes de ingressar na graduação, você teve contato com algo/alguém da área?
 - a) Não.
 - b) Sim, em casa.
 - c) Sim, em cursos.
 - d) Sim, na escola.
 - e) Sim, em outro(s) ambiente(s).

A.4 Percepções durante o curso

1. Quantas mulheres ingressaram na sua turma, além de você?
2. Quantas destas permaneceram no curso, além de você?
3. Na sua opinião, por que as mulheres da turma se separaram?
 - a) Suas colegas desistiram.
 - b) Suas colegas não estão regulares.
 - c) Você não está regular.
 - d) Suas colegas foram jubiladas.
 - e) Você foi jubilada.
 - f) Transferência interna/externa.
 - g) Outros.
4. Como é sua relação com as colegas mulheres da faculdade?
 - a) Próxima.
 - b) Distante.
 - c) Formal.
 - d) Tenho amigas na turma.
 - e) Não tenho amigas na turma.
 - f) Tenho colegas com as quais às vezes realizo atividades.
 - g) Tenho colegas com as quais sempre realizo atividades.
 - h) Sinto-me à vontade com elas.
 - i) Já sofri algum tipo de assédio.
 - j) Já me senti hostilizada.
 - k) Outros.
5. Como é sua relação com os colegas homens da faculdade?
 - a) Próxima.
 - b) Distante.
 - c) Formal.
 - d) Tenho amigos na turma.
 - e) Não tenho amigos na turma.

- f) Tenho colegas com os quais às vezes realizo atividades.
 - g) Tenho colegas com os quais sempre realizo atividades.
 - h) Sinto-me à vontade com eles.
 - i) Já sofri algum tipo de assédio.
 - j) Já me senti hostilizada.
 - k) Outros.
6. Como é sua relação com as professoras?
- a) Tenho uma relação mais próxima com os professores.
 - b) Sou tratada de forma igualitária.
 - c) Minhas observações e perguntas são consideradas durante as aulas.
 - d) Sinto-me acolhida e estimulada a continuar no curso/na área.
 - e) Não me sinto à vontade com elas.
 - f) Já senti alguma manifestação de descrédito.
 - g) Outros.
7. Como é sua relação com os professores?
- a) Tenho uma relação mais próxima com as professoras.
 - b) Sou tratada de forma igualitária.
 - c) Minhas observações e perguntas são consideradas durante as aulas.
 - d) Sinto-me acolhida e estimulada a continuar no curso/na área.
 - e) Não me sinto à vontade com eles.
 - f) Já senti alguma manifestação de descrédito.
 - g) Outros.
8. Você participa/já participou de alguma atividade extracurricular, como projetos de extensão, iniciação científica, entre outros?
- a) Sim.
 - b) Não.
9. Você já sofreu e/ou presenciou alguma cena de preconceito/discriminação contra mulheres no curso? Se sim, relate brevemente.
10. O que você mais gosta no curso?
- a) A ementa do curso.

- b) Aulas práticas com hardware.
 - c) Aulas práticas com software.
 - d) Professores.
 - e) Atividades extracurriculares.
 - f) Colegas.
 - g) Localização do campus.
 - h) Infraestrutura da universidade.
 - i) Outros.
11. O que você menos gosta no curso?
- a) A ementa do curso.
 - b) Aulas práticas com hardware.
 - c) Aulas práticas com software.
 - d) Professores.
 - e) Atividades extracurriculares.
 - f) Colegas.
 - g) Localização do campus.
 - h) Infraestrutura da universidade.
 - i) Outros.
12. Você já pensou em desistir do curso?
- a) Sim.
 - b) Não.
13. Por quais motivos você pensou em desistir do curso?
14. Quais suas perspectivas para depois da conclusão do curso?
- a) Procurar um emprego na área.
 - b) Procurar um emprego em outra área.
 - c) Fazer uma especialização.
 - d) Seguir carreira acadêmica, fazendo mestrado e doutorado.
 - e) Continuar no emprego que estou.
 - f) Outros.
15. Você conhece mulheres que contribuem/contribuíram para a evolução da computação?

- a) Sim.
 - b) Não.
16. Qual/Quais mulher(es) você conhece?
17. Na sua opinião, qual a explicação para a baixa presença de mulheres na computação?
- a) Falta de interesse.
 - b) Falta de conhecimento sobre cursos na área.
 - c) Falta de incentivo de familiares/amigos/professores, entre outros.
 - d) Preconceito.
 - e) Baixa perspectiva salarial.
 - f) Outros.
18. Na sua opinião, as mulheres enfrentam algum tipo de obstáculo para prosseguir na computação? Se sim, qual/quais?
19. Na sua opinião, as mulheres têm as mesmas oportunidades que os homens na computação? Por quê?

A.5 Análise comportamental

1. Eu sou:
 - a) Carismática e otimista.
 - b) Acolhedora e paciente.
 - c) Detalhista e organizada.
 - d) Ousada e direta.
2. Prefiro focar:
 - a) Em resultados.
 - b) No planejamento.
 - c) Nas pessoas ao meu redor.
 - d) Nas tarefas que precisam ser feitas.
3. Eu gosto de:
 - a) Delegar.
 - b) Ter controle.

- c) Fazer junto.
 - d) Agir individualmente.
4. Eu sou uma pessoa:
- a) Questionadora.
 - b) Compreensiva.
 - c) Persuasiva.
 - d) Autodisciplinada.
5. Quando tenho uma ideia:
- a) Me atento aos detalhes e ajo com cautela.
 - b) Sonho alto.
 - c) Tomo decisões impulsivas.
 - d) Planejo os próximos passos.
6. Eu acredito que:
- a) Feito é melhor que perfeito.
 - b) Tudo deve ser feito com a melhor qualidade.
 - c) Está tudo bem fazer mudanças no processo.
 - d) Paciência e um ambiente harmônico são a chave para o sucesso.
7. Eu penso que:
- a) É melhor prudência do que arrependimento.
 - b) Duas cabeças pensam melhor do que uma.
 - c) Se você não tem condições de competir, não compita.
 - d) Não é fácil ficar encurralado.
8. Minha opinião sobre mudanças é:
- a) São necessárias.
 - b) Prefiro manter a rotina.
 - c) Aceito, desde que não impeçam meu progresso.
 - d) Aceito, desde que sejam bem estruturadas e sem riscos.