

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Dissertação de Mestrado

Análise da Transição Profissional dos Egressos do
Curso de Graduação em Ciência da Computação
da Universidade Federal de Campina Grande

Felisberto Cláudio dos Santos

Campina Grande, Paraíba, Brasil

Fevereiro/2025

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Análise da Transição Profissional dos Egressos do Curso de
Graduação em Ciência da Computação da Universidade Federal
de Campina Grande

Felisberto Cláudio dos Santos

Dissertação submetida à Coordenação do Curso de Pós-Graduação em
Ciência da Computação da Universidade Federal de Campina Grande -
Campus I como parte dos requisitos necessários para obtenção do grau
de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Metodologia e Técnicas da Computação

Francisco Vilar Brasileiro (Orientador)

Campina Grande, Paraíba, Brasil

©Felisberto Cláudio dos Santos, 19/02/2025

S237a

Santos, Felisberto Cláudio dos.

Análise da transição profissional dos egressos do curso de graduação em Ciência da Computação da Universidade Federal de Campina Grande / Felisberto Cláudio dos Santos – Campina Grande, 2025.

102 f. : il. color.

Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2025.

"Orientação: Prof. Dr. Francisco Vilar Brasileiro."

Referências.

1. Ciência da Computação. 2. Empregabilidade. 3. Egressos Universitários. 4. Disciplinas Optativas. 5. Análise de Dados. 6. Transição Profissional. I. Brasileiro, Francisco Vilar. II. Título.

CDU 004(043)



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
POS-GRADUACAO EM CIENCIA DA COMPUTACAO
Rua Aprígio Veloso, 882, Edifício Telmo Silva de Araújo, Bloco CG1, - Bairro Universitário, Campina Grande/PB, CEP 58429-900
Telefone: 2101-1122 - (83) 2101-1123 - (83) 2101-1124
Site: <http://computacao.ufcg.edu.br> - E-mail: secp@computacao.ufcg.edu.br

FOLHA DE ASSINATURA PARA TESES E DISSERTAÇÕES

FELISBERTO CLÁUDIO DOS SANTOS

ANÁLISE DA TRANSIÇÃO PROFISSIONAL DOS EGRESSOS DO CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DA UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação como pré-requisito para obtenção do título de Mestre em Ciência da Computação.

Aprovada em: 24/02/2025

Prof. Dr. **Francisco Vilar Brasileiro**, UFCG, Orientador

Profa. Dra. **Lívia Maria Rodrigues Sampaio Campos**, UFCG, Examinadora Interna

Prof. Dr. **Katysco de Farias Santos**, IFPB, Examinador Externo



Documento assinado eletronicamente por **FRANCISCO VILAR BRASILEIRO, PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 25/02/2025, às 10:37, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **LIVIA MARIA RODRIGUES SAMPAIO CAMPOS, PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 25/02/2025, às 11:55, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **Katysco de Farias Santos, Usuário Externo**, em 25/02/2025, às 15:44, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



A autenticidade deste documento pode ser conferida no site <https://sei.ufcg.edu.br/autenticidade>, informando o código verificador **5259104** e o código CRC **53C300F2**.

Referência: Processo nº 23096.009445/2025-03

SEI nº 5259104



MINISTÉRIO DA EDUCAÇÃO

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

POS-GRADUACAO EM CIENCIA DA COMPUTACAO

Rua Aprígio Veloso, 882, Edifício Telmo Silva de Araújo, Bloco CG1, - Bairro Universitário, Campina Grande/PB, CEP 58429-900

Telefone: 2101-1122 - (83) 2101-1123 - (83) 2101-1124

Site: <http://computacao.ufcg.edu.br> - E-mail: secp@computacao.ufcg.edu.br

REGISTRO DE PRESENÇA E ASSINATURAS

ATA Nº 006/2025 (DISSERTAÇÃO Nº 742)

Aos vinte e quatro (24) dias do mês de fevereiro do ano de dois mil e vinte e cinco (2025), às quinze horas (15:00), no Auditório Mário Hattori, da Universidade Federal de Campina Grande - UFCG, nesta cidade, reuniu-se a Comissão Examinadora composta pelos Professores FRANCISCO VILAR BRASILEIRO, Dr., UFCG, Orientador, funcionando neste ato como Presidente, LÍVIA MARIA RODRIGUES SAMPAIO CAMPOS, Dra., UFCG, KATYUSCO DE FARIAS SANTOS, Dr., IFPB. Constituída a mencionada Comissão Examinadora pela Portaria Nº 008/2025 do Coordenador do Programa de Pós-Graduação em Ciência da Computação, tendo em vista a deliberação do Colegiado do Curso, tomada em reunião de 13 de Fevereiro de 2025 e com fundamento no Regulamento Geral dos Cursos de Pós-Graduação da Universidade Federal de Campina Grande - UFCG, juntamente com o Sr. FELISBERTO CLÁUDIO DOS SANTOS, candidato ao grau de MESTRE em Ciência da Computação, presentes ainda professores e alunos do referido centro e demais presentes. Abertos os trabalhos, o Senhor Presidente da Comissão Examinadora anunciou que a reunião tinha por finalidade a apresentação e julgamento da dissertação "ANÁLISE DA TRANSIÇÃO PROFISSIONAL DOS EGRESSOS DO CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DA UNIVERSIDADE FEDERAL DE CAMPINA GRANDE", elaborada pelo candidato acima designado, sob a orientação do(s) Professor(es) FRANCISCO VILAR BRASILEIRO, com o objetivo de atender às exigências do Regulamento Geral dos Cursos de Pós-Graduação da Universidade Federal de Campina Grande - UFCG. A seguir, concedeu a palavra ao candidato, o qual, após salientar a importância do assunto desenvolvido, defendeu o conteúdo da dissertação. Concluída a exposição e defesa do candidato, passou cada membro da Comissão Examinadora a arguir o mestrando sobre os vários aspectos que constituíram o campo de estudo tratado na referida dissertação. Terminados os trabalhos de arguição, o Senhor Presidente da Comissão Examinadora determinou a suspensão da sessão pelo tempo necessário ao julgamento da dissertação. Reunidos, em caráter secreto, no mesmo recinto, os membros da Comissão Examinadora passaram à apreciação da dissertação. Reaberta a sessão, o Presidente da Comissão Examinadora anunciou o resultado do julgamento, tendo assim, o candidato obtido o Conceito APROVADO. A seguir, foi encerrada a sessão e lavrada a presente ata, que vai assinada por mim, Paloma Nascimento Porto, pelos membros da Comissão Examinadora e pelo candidato. Campina Grande, 24 de fevereiro de 2025.



Documento assinado eletronicamente por **FRANCISCO VILAR BRASILEIRO, PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 25/02/2025, às 10:37, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **LIVIA MARIA RODRIGUES SAMPAIO CAMPOS, PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 25/02/2025, às 11:54, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **FELISBERTO CLÁUDIO DOS SANTOS, Usuário Externo**, em 25/02/2025, às 12:22, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **Katysuco de Farias Santos, Usuário Externo**, em 25/02/2025, às 15:43, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **PALOMA NASCIMENTO PORTO, ASSISTENTE EM ADMINISTRACAO**, em 25/02/2025, às 16:14, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



A autenticidade deste documento pode ser conferida no site <https://sei.ufcg.edu.br/autenticidade>, informando o código verificador **5259103** e o código CRC **02661059**.

Resumo

A inserção de egressos universitários no mundo do trabalho é um tema central na educação superior, especialmente em áreas dinâmicas como a Ciência da Computação. Este estudo investiga como as disciplinas optativas cursadas durante a graduação influenciam a inserção profissional dos egressos do curso de Ciência da Computação da Universidade Federal de Campina Grande (UFCG).

Foi realizada uma análise quantitativa dos dados acadêmicos de 352 egressos e profissionais de 263 destes. O método K-Means foi empregado para segmentação dos perfis acadêmicos e profissionais, complementado por análises de correlação entre escolhas curriculares e fatores de empregabilidade, como tempo até a primeira ocupação, modalidade de trabalho e porte empresarial.

Os resultados mostram que a clusterização acadêmica identificou três perfis principais: desenvolvimento (DEV), desenvolvimento e inteligência artificial (DEV+IA) e desenvolvimento e infraestrutura (DEV+INFRA). Embora disciplinas específicas mostrem associação com determinadas carreiras, a análise dos clusters profissionais revelou que a empregabilidade está mais relacionada a fatores externos, como modalidade de trabalho (híbrido/remoto) e tempo até a primeira ocupação, com predominância de trabalho remoto e híbrido.

A pesquisa conclui que as escolhas curriculares influenciam tendências na transição para o mundo do trabalho, mas não a determinam diretamente. A empregabilidade depende de um conjunto amplo de fatores, incluindo habilidades transversais e experiências práticas. Estes achados podem informar políticas educacionais mais alinhadas às demandas do setor tecnológico.

Palavras-chave: empregabilidade, egressos universitários, Ciência da Computação, disciplinas optativas, análise de dados, transição profissional.

Abstract

The transition of university graduates into the labor market is a central issue in higher education, particularly in dynamic fields such as Computer Science. This study investigates the extent to which elective courses taken during undergraduate studies influence the professional insertion of graduates from the Computer Science program at the Federal University of Campina Grande (UFCG), Brazil.

A quantitative analysis was conducted using academic data from 352 graduates, along with professional information from 263 of them. The K-Means algorithm was applied to segment both academic and professional profiles, followed by correlation analyses to explore the relationship between curricular choices and employability factors such as time to first employment, work modality, and company size.

The results indicate that academic clustering revealed three main profiles: development (DEV), development with artificial intelligence (DEV+AI), and development with infrastructure (DEV+INFRA). While certain electives were associated with specific career paths, the analysis of professional clusters suggests that employability is more strongly influenced by external factors, particularly work modality (remote/hybrid) and the time until first employment, with remote and hybrid work prevailing.

The study concludes that curricular choices shape tendencies in the transition to the labor market but do not directly determine professional outcomes. Employability depends on a broader set of factors, including transversal skills and practical experience. These findings may inform educational policies better aligned with the demands of the technology sector.

Keywords: employability, university graduates, Computer Science, elective courses, data analysis, career transition.

Agradecimentos

Agradeço, primeiramente, a Deus, por me conceder saúde e força ao longo desta jornada, especialmente estando longe de casa. À minha família, expresso minha mais profunda gratidão pelo apoio incondicional, pelo incentivo constante e por sempre acreditarem em mim, mesmo à distância.

Ao meu orientador, sou imensamente grato pela orientação dedicada, pelos desafios que me impulsionaram a crescer academicamente e pela paciência e incentivo ao longo deste percurso.

Aos amigos que fiz no Brasil e àqueles que deixei em minha terra natal, mas que nunca deixaram de me acompanhar, agradeço pelo companheirismo, pelas palavras de encorajamento e pelo suporte que tornaram essa caminhada mais leve e significativa.

Este trabalho é também fruto de todas essas conexões e do apoio que recebi ao longo desta trajetória. Meu sincero muito obrigado a todos.

Conteúdo

1	Introdução	1
1.1	Objetivo e Problema de Estudo	2
1.2	Contexto e Importância do Estudo	2
1.3	Estrutura da Dissertação	6
2	Fundamentação Teórica	8
2.1	Descrição e Estruturação dos Dados	9
2.1.1	Caracterização Estatística dos Dados Acadêmicos	9
2.1.2	Caracterização Estatística dos Dados Profissionais	12
2.1.3	Técnicas de Pré-processamento e Análise	15
2.2	Análise de Transições Profissionais	17
2.3	Mineração de Dados Educacionais	20
2.3.1	Abordagens Metodológicas em EDM	21
2.3.2	Métodos de Clusterização em EDM	23
2.4	Fundamentos Matemáticos da Clusterização	24
2.4.1	Normalização e Escalonamento de Dados	29
2.4.2	Justificativa para o Uso do K-Means	31
2.5	Redução de Dimensionalidade com PCA	32
2.5.1	Considerações Finais e Conexão com Trabalhos Relacionados	35
3	Trabalhos Relacionados	37
3.1	Análise das Transições Profissionais: Abordagens e Técnicas	37
3.1.1	Mineração de Dados do LinkedIn para Análise de Carreiras	38
3.1.2	Técnicas de Clusterização em Análise Educacional	38

3.1.3	Predição de Transição Profissionais com Inteligência Artificial . . .	39
3.2	Empregabilidade de Graduados em Computação	40
3.3	Satisfação dos Empregadores com Graduados em TIC	41
3.4	Síntese e Tendências Emergentes	42
3.5	Lacunas de Pesquisa e Nossa Contribuição	43
4	Metodologia	44
4.1	Ferramentas e Recursos	45
4.1.1	Fontes de Dados	45
4.1.2	Ferramentas de Análise	46
4.1.3	Plataformas de Colaboração e Documentação	46
4.1.4	Métricas Acadêmicas	47
4.1.5	Métricas Profissionais	47
4.1.6	Tratamento dos Dados Profissionais	50
4.2	Clusterização com o Algoritmo K-Means	52
4.2.1	Descrição do Algoritmo	52
4.2.2	Parâmetros de Implementação e Determinação do número de clusters K	54
4.2.3	Análise de Componentes Principais (PCA)	56
4.2.4	Explicação Técnica dos Gráficos	57
5	Análise e Discussão dos Resultados	59
5.1	Clusterização dos Dados Acadêmicos	60
5.1.1	Análise da Configuração k=6	60
5.1.2	Análise da Configuração k=5	62
5.1.3	Análise da Configuração k=4	64
5.1.4	Análise da Configuração k=3	66
5.1.5	Conclusão da Análise dos Dados Acadêmicos	72
5.2	Clusterização dos Dados Profissionais	73
5.2.1	Definição e Validação dos Clusters	73
5.2.2	Caracterização dos Perfis Profissionais	74
5.2.3	Análise Quantitativa dos Clusters	76

5.3	Discussão dos Resultados	77
5.3.1	Análise da Correlação dos Dados Profissionais com os Dados Acadêmicos dos Egressos	78
5.3.2	Resultados	84
5.3.3	Limitações e Trabalhos Futuros	89
6	Conclusões	94
A	Disciplinas optativas ofertadas por período	103
B	Primeira enquête	105
C	Segunda enquête	115

Lista de Símbolos

UFCG – Universidade Federal de Campina Grande

DEV – Desenvolvimento

IA – Inteligência Artificial

INFRA – Infraestrutura

EDM – Mineração de Dados Educacionais (Education Data Mining)

PCA – Análise de Componentes Principais (Principal Component Analysis)

DEA – Análise por Envoltória de Dados (Data Envelopment Analysis)

TIC – Tecnologias da Informação e Comunicação

KNN – K-Nearest Neighbors

DBSCAN – Density-Based Spatial Clustering of Applications with Noise

LSTM – Long Short-Term Memory

GRU – Gated Recurrent Unit

SMOTE – Synthetic Minority Over-sampling Technique

ANOVA – Análise de Variância

WCSS – Within-Cluster Sum of Squares

STFCD – Systems Theory Framework of Career Development

Lista de Figuras

4.1	Diagrama ilustrativo do processo metodológico.	45
4.2	Gráfico de calor apresentando a correlação entre os três grupos de disciplinas.	49
4.3	Quantidade de frequências de matrículas dos egressos nas disciplinas de cada grupo.	50
4.4	Gráfico de barra, distribuição por tempo, modalidade de trabalho e porte empresarial.	51
4.5	Distribuição por cargos.	51
4.6	Gráfico da matriz de correlação das variáveis dos dados profissionais.	53
4.7	Métricas de qualidade dos clusters para os dados acadêmicos.	55
4.8	Métricas de qualidade dos clusters para os dados acadêmicos.	55
4.9	Comparação entre dados originais (esquerda) e após aplicação do K-means (direita). Imagem adaptada do google.	57
5.1	Distribuição dos clusters utilizando PCA no K-means.	61
5.2	Gráfico de barras representando a média dos atributos para k=6.	61
5.3	Gráfico de barras representando a mediana dos atributos para k=6.	62
5.4	Distribuição dos clusters utilizando PCA no K-means.	63
5.5	Gráfico de barras representando a média dos atributos para k=5.	63
5.6	Gráfico de barras representando a mediana dos atributos para k=5.	64
5.7	Distribuição dos clusters utilizando PCA no K-means.	64
5.8	Gráfico de barras representando a média dos atributos para k=4.	65
5.9	Gráfico de barras representando a mediana dos atributos para k=4.	66
5.10	Distribuição dos clusters utilizando PCA no K-means.	67
5.11	Gráfico de barras representando a média dos atributos para k=3.	68

5.12	Gráfico de barras representando a mediana dos atributos para $k=3$	68
5.13	Gráfico de boxplots para $k=3$ dos dados acadêmicos.	70
5.14	Distribuição de primeiros cargos de trabalho por cluster para $k=3$	74
5.15	Distribuição da modalidade de trabalho por cluster para $K=3$	74
5.16	Distribuição de porte empresarial por cluster para $k=3$	75
5.17	Gráfico dos atributos por clusters profissionais.	76
5.18	Gráfico de boxplots dos clusters profissionais para $k = 3$	77
5.19	Gráfico de barras empilhada de distribuição dos clusters profissionais por cluster acadêmico.	79
5.20	Mapa de calor de intersecção entre clusters acadêmicos e profissionais. . . .	80
5.21	Gráfico de pizza, distribuição dos clusters profissionais em percentagens. . .	85

Lista de Tabelas

2.1	Informação estatística dos dados acadêmicos.	11
2.2	Estrutura da matriz de dados acadêmicos após tratamento.	12
2.3	Comparação entre métodos de clusterização.	23
2.4	Comparação entre métricas de clusterização.	29
2.5	Vantagens e desvantagens do PCA.	35
3.1	Comparação entre Abordagens de Análise de Trajetórias Profissionais . . .	40
5.1	Centróides dos clusters no espaço normalizado.	71
5.2	Centróides dos clusters no espaço PCA.	72
5.3	Distribuição dos clusters acadêmicos dentro de cada cluster profissional. . .	78
5.4	Percentual de cada perfil acadêmico nos diferentes clusters profissionais. . .	84

Capítulo 1

Introdução

A transição dos egressos universitários para o mundo do trabalho é um tema de grande relevância na educação superior, particularmente em áreas dinâmicas como a Ciência da Computação. Em um cenário global cada vez mais tecnológico, a empregabilidade dos recém-formados tornou-se um indicador fundamental da qualidade do ensino superior, sendo influenciada por múltiplos fatores que incluem competências técnicas, habilidades interpessoais e experiências adquiridas durante a graduação [35; 23].

No campo da Ciência da Computação, essa transição assume contornos particulares devido à constante evolução das tecnologias e às exigências crescentes do setor. O mundo do trabalho de tecnologia da informação demanda que os profissionais estejam em contínuo aprendizado, adaptando-se às mudanças nas exigências do setor [48]. Esta adaptabilidade torna-se ainda mais crítica quando consideramos que os empregadores esperam dos recém-formados não apenas conhecimentos teóricos sólidos, mas também competências práticas em áreas como mineração de dados, programação e análise quantitativa [2].

Na Universidade Federal de Campina Grande (UFCG), o curso de Ciência da Computação oferece uma estrutura curricular flexível, permitindo que aproximadamente 20% dos créditos sejam compostos por disciplinas optativas. Esta flexibilidade alinha-se com tendências internacionais observadas em programas bem-sucedidos, onde a capacidade de personalização do currículo tem-se mostrado um fator importante para a empregabilidade [2]. No entanto, ainda não há um mapeamento sistemático de como essas escolhas influenciam as transições profissionais dos egressos no curso de Ciência da Computação da UFCG.

Este estudo busca preencher essa lacuna por meio da análise de dados de 353 egressos do

curso desde o período letivo 2012.2, investigando padrões nas escolhas de disciplinas optativas e sua relação com a inserção profissional desses egressos. A pesquisa combina análise de registros acadêmicos com dados sobre empregabilidade, buscando identificar padrões que possam orientar futuros estudantes e contribuir para a formulação de estratégias educacionais mais alinhadas às demandas do mundo do trabalho.

1.1 Objetivo e Problema de Estudo

O presente estudo tem como objetivo central analisar como as escolhas de disciplinas optativas influenciam a transição profissional dos egressos do curso de Ciência da Computação da UFCG. Esta investigação se fundamenta em evidências recentes que apontam um desalinhamento significativo entre a formação acadêmica e as demandas do setor tecnológico [19], particularmente em três dimensões críticas frequentemente negligenciadas nos currículos tradicionais: capacidade analítica avançada, proficiência em ferramentas computacionais específicas e habilidade de integração de diferentes tecnologias [2].

Estudos precedentes demonstram que técnicas quantitativas podem revelar padrões significativos em estruturas curriculares, identificando tanto tendências globais quanto características específicas de diferentes instituições [28]. Esta abordagem quantitativa ganha relevância especial quando consideramos evidências de que programas acadêmicos que conseguem equilibrar teoria e prática apresentam melhores indicadores de empregabilidade [2]. Rahman e Asadujjaman [36] validam que a análise de agrupamentos, quando aplicada a dados educacionais e profissionais, permite identificar perfis naturais que auxiliam na compreensão de padrões de transição entre formação e mundo do trabalho. Para abordar este objetivo central, o estudo se estrutura em três questões principais de pesquisa que serão desenvolvidas na seção seguinte.

1.2 Contexto e Importância do Estudo

A transição dos egressos universitários para o mundo do trabalho representa um desafio crítico na educação superior, especialmente em Ciência da Computação, onde as demandas do setor evoluem rapidamente. Um estudo recente analisando 38 programas acadêmicos

internacionais identificou uma lacuna significativa entre a formação tradicional e as competências valorizadas pelos empregadores, especialmente em relação às habilidades analíticas avançadas como modelagem preditiva, análise de Big Data, e proficiência em tecnologias emergentes [2].

O contexto específico da UFCG apresenta características particulares que tornam esta investigação ainda mais relevante. Com uma estrutura curricular que permite 20% de disciplinas optativas, o curso oferece aos estudantes significativa flexibilidade na construção de seu perfil profissional. Esta flexibilidade, embora alinhada com tendências internacionais bem-sucedidas, demanda uma compreensão mais profunda de como as escolhas curriculares impactam as transições profissionais dos egressos.

A importância deste estudo se manifesta em três dimensões principais:

1. **Orientação Acadêmica:** Os resultados fornecerão evidências empíricas para orientar estudantes na adequação de sua formação às aspirações profissionais. Estudos indicam que escolhas acadêmicas conscientes aumentam em 1,8 vezes as chances de uma transição bem sucedida para o mundo do trabalho [35].
2. **Desenvolvimento Curricular:** A análise sistemática das transições dos egressos permitirá identificar padrões que podem informar futuras adaptações curriculares. Como destacado por Hassan et al. [19], a compreensão da relação entre formação acadêmica e demandas profissionais é fundamental para o aprimoramento contínuo dos programas de graduação.
3. **Impacto Regional:** Considerando o papel da UFCG como centro de formação tecnológica no Nordeste brasileiro, os resultados desta pesquisa podem contribuir para o fortalecimento do ecossistema tecnológico regional, alinhando a formação acadêmica às necessidades do mundo do trabalho local e nacional.

A relevância desta investigação é reforçada por estudos recentes que demonstram como o monitoramento sistemático dos egressos pode fortalecer a vinculação institucional e aprimorar continuamente os programas acadêmicos [39]. Particularmente no contexto da Computação, onde as demandas profissionais evoluem rapidamente, este tipo de análise permite adequar a formação às necessidades emergentes do setor tecnológico.

Este estudo estrutura-se em três questões principais de pesquisa (QP), cada uma associada a objetivos específicos, metodologia e resultados esperados:

QP1: Como as escolhas de disciplinas optativas dos graduados em Ciência da Computação da UFCG se relacionam com os padrões de inserção profissional inicial?

- **Objetivo específico:** Identificar e caracterizar padrões recorrentes nas escolhas de disciplinas optativas e sua relação com a transição na primeira experiência profissional.
- **Métodos:**
 - Aplicação de técnicas de agrupamento (clusterização) para identificação de perfis acadêmicos, seguindo Matsuda et al. [28]
 - Análise em quatro dimensões proposta por Pazur et al. [32]: Empregabilidade e soft skills (embora esses não sejam o foco deste trabalho), abordagem holística para empregabilidade, métodos de ensino e adaptações curriculares.
- **Resultados esperados:** Mapeamento de perfis de escolha curricular e suas conexões com os primeiros empregos dos egressos.

QP2: Quais combinações de disciplinas optativas estão associadas a diferentes transições de inserção profissional?

- **Objetivo específico:** Analisar a correlação entre grupos de disciplinas cursadas e as áreas de atuação dos egressos.
- **Métodos:**
 - Análise de associação entre perfis curriculares e inserção profissional;
 - Técnicas estatísticas e mineração de dados [19];
 - Arcabouço de análise em quatro camadas: dados, embeddings, mapeamento e aprimoramento curricular.
- **Resultados esperados:** Identificação de combinações curriculares estratégicas associadas a diferentes carreiras no setor de tecnologia.

QP3: Como os diferentes perfis de escolhas curriculares influenciam o tempo e a área da primeira atuação profissional?

- **Objetivo específico:** Examinar o impacto temporal das escolhas acadêmicas na inserção profissional.
- **Métodos:**
 - Análise temporal das transições profissionais
 - Estudo do tempo até a primeira ocupação
 - Análise de áreas de atuação [43]
 - Técnicas de mineração de dados para identificação de padrões temporais
- **Resultados esperados:** Compreensão da relação entre perfis curriculares e tempo de inserção no mundo do trabalho, além de padrões de alocação profissional.

Esta estruturação metodológica é suportada por evidências empíricas recentes que demonstram a eficácia de combinar múltiplos métodos de análise. Um estudo abrangente com 38 programas acadêmicos [2] validou a eficácia dessa abordagem multimétodo para compreender a relação entre estrutura curricular e sucesso profissional.

A relevância desta pesquisa manifesta-se em três dimensões principais, cada uma contribuindo de forma distinta para o avanço do conhecimento e práticas no campo da educação em Computação.

A investigação oferece contribuições práticas significativas para o processo de orientação acadêmica, fundamentadas em evidências empíricas. Esta pesquisa fornece:

- Base empírica para aconselhamento acadêmico
- Diretrizes para escolhas curriculares alinhadas às aspirações profissionais
- Subsídios para desenvolvimento de estratégias de mentoria

O estudo contribui para o avanço metodológico na análise da transição universidade-trabalho através de:

- Integração de técnicas de mineração de dados com análise de transições profissionais
- Desenvolvimento de arcabouço analítico replicável em outros contextos

- Validação de métodos quantitativos para análise de impacto curricular

Esta abordagem metodológica baseia-se em precedentes estabelecidos por Parack et al. [31] e Rane et al. [37], que demonstram a eficácia do uso de mineração de dados para a identificação de padrões em trajetórias acadêmicas.

A pesquisa apresenta relevância específica para o contexto regional de Campina Grande e do Nordeste brasileiro:

- Fornece subsídios para políticas educacionais locais;
- Contribui para o fortalecimento do pólo tecnológico regional;
- Auxilia no alinhamento entre formação acadêmica e demandas do mundo de trabalho local.

Como destacado por Zarch et al. ([52]), é fundamental considerar que mudanças curriculares podem ter impactos diferentes em distintos contextos socioeconômicos. Embora este estudo não aborde diretamente aspectos socioeconômicos, seus resultados podem informar políticas institucionais que promovam maior equidade no acesso a oportunidades profissionais.

A combinação dessas três dimensões de relevância reforça o potencial impacto desta pesquisa tanto no âmbito acadêmico quanto no desenvolvimento regional. Os resultados esperados podem contribuir não apenas para o aprimoramento das práticas educacionais na UFCG, mas também servir como referência para outras instituições que busquem fortalecer a conexão entre formação acadêmica e demandas do setor tecnológico.

1.3 Estrutura da Dissertação

Esta dissertação está estruturada em seis capítulos, abordando de maneira integrada a transição profissional dos egressos do curso de Ciência da Computação da UFCG. O Capítulo 1 (Introdução) apresenta o contexto da pesquisa, sua motivação e problematização, além dos objetivos e da justificativa do estudo, destacando a importância de compreender a relação entre a formação acadêmica e a inserção profissional. O Capítulo 2 (Fundamentação Teórica) aprofunda os conceitos centrais que sustentam a investigação, discutindo a relação

entre escolhas curriculares e empregabilidade, o impacto das disciplinas optativas na construção das trajetórias profissionais e as bases conceituais da Mineração de Dados Educacionais (EDM), além dos métodos matemáticos aplicados à análise das trajetórias acadêmicas e transições profissionais. No Capítulo 3 (Trabalhos Relacionados), são revisadas pesquisas que empregam EDM e aprendizado de máquina para análise de carreiras e empregabilidade, abordando estudos sobre mineração de dados do LinkedIn, predição de trajetórias profissionais e o alinhamento entre a formação acadêmica e as demandas do setor de tecnologia, situando esta pesquisa dentro do estado da arte. O Capítulo 4 (Metodologia) detalha os procedimentos adotados para a coleta, processamento e análise dos dados, incluindo a aplicação da clusterização via K-Means, da Análise de Componentes Principais (PCA) e de métricas estatísticas para validação dos agrupamentos, além da integração dos dados acadêmicos e profissionais para segmentação dos perfis dos egressos. O Capítulo 5 (Análise e Discussão dos Resultados) apresenta os achados da pesquisa, destacando os padrões identificados na clusterização dos dados acadêmicos e profissionais, a inserção dos egressos no mundo do trabalho e sua evolução de carreira, bem como as correlações entre escolhas curriculares e transições profissionais, avaliando o impacto das disciplinas optativas na empregabilidade. Por fim, o Capítulo 6 (Conclusão) sintetiza as principais descobertas do estudo, discutindo suas implicações para a formação acadêmica e políticas educacionais, além de apresentar as contribuições da pesquisa, suas limitações e sugestões para estudos futuros, com foco na melhoria da empregabilidade e na adaptação curricular às exigências do setor tecnológico. Com essa estrutura, a dissertação busca oferecer uma análise aprofundada e baseada em evidências sobre os fatores que influenciam a transição dos egressos para o mundo do trabalho, embora outros fatores também possam influenciar, como dito as soft skills, projetos de pesquisa e extensão, monitorias e estágios, visando fornecer subsídios para o aperfeiçoamento dos currículos acadêmicos e estratégias de formação mais alinhadas às necessidades da área de Computação.

Capítulo 2

Fundamentação Teórica

Este capítulo estabelece a base teórica da pesquisa, apresentando os conceitos fundamentais e teorias relevantes para compreender a relação entre escolhas curriculares e a transição profissional dos egressos do curso de Ciência da Computação da UFCG. A fundamentação teórica foi estruturada para examinar como as decisões acadêmicas influenciam a inserção profissional e a empregabilidade dos graduados, considerando evidências empíricas e construtos teóricos que exploram essa relação.

No contexto da análise curricular, Gannod e Henderson [17] estabelecem uma abordagem sistemática para identificação de lacunas curriculares, propondo três categorias principais: lacunas curriculares, caracterizadas pela ausência de conteúdos essenciais no currículo; lacunas de curso, evidenciadas quando disciplinas abordam conteúdos de forma desatualizada ou superficial; e lacunas de programa, manifestadas quando restrições curriculares impedem o acesso dos alunos a cursos relevantes. Esta taxonomia oferece um arcabouço metodológico para avaliar a efetividade dos programas educacionais em atender às demandas do mundo do trabalho, fundamentando ajustes curriculares estratégicos.

A construção do perfil profissional durante a graduação está intrinsecamente ligada às escolhas curriculares, particularmente na seleção de disciplinas optativas. Qenani et al. [35] demonstram que estas decisões são determinantes na formação das competências profissionais, exercendo influência direta sobre oportunidades de trabalho e desenvolvimento de carreira. Corroborando esta perspectiva, Kansal e Sadawarti [23] enfatizam que currículos flexíveis e adaptáveis são essenciais para preparar profissionais alinhados às exigências contemporâneas do mundo do trabalho.

Estudos recentes destacam que a empregabilidade é potencializada pela combinação entre especialização técnica e competências interdisciplinares. Akhtar et al. [2] demonstram que a empregabilidade é potencializada pela combinação entre especialização técnica e competências interdisciplinares. O desenvolvimento de habilidades transversais, como pensamento crítico e capacidade de resolução de problemas, mostra-se particularmente relevante em áreas emergentes como Inteligência Artificial, Engenharia de Software e Segurança da Informação, que apresentam crescente demanda no mundo do trabalho.

Evidências quantitativas suportam esta perspectiva. Costa [8] verificou que egressos que optaram por disciplinas ligadas a tecnologias emergentes demonstraram um tempo de inserção profissional mais curto e maior taxa de sucesso na transição para o mundo do trabalho. Complementarmente, Pereira et al. [33] demonstram que estudantes com formação em Inteligência Artificial e Engenharia de Software apresentam redução de até 30% no tempo médio de inserção no mundo do trabalho, comparados àqueles com formação tradicional.

A complexidade destas variáveis educacionais e seus impactos na empregabilidade demanda uma abordagem quantitativa robusta para identificar padrões nas transições profissionais dos egressos. A seção seguinte apresenta as técnicas metodológicas empregadas nesta análise de carreiras.

2.1 Descrição e Estruturação dos Dados

Esta pesquisa fundamenta-se na análise de dois conjuntos de dados complementares: os registros acadêmicos e as informações profissionais dos egressos do curso de Ciência da Computação da UFCG. A metodologia adotada para estruturação e análise desses dados foi desenvolvida visando assegurar rigor analítico e confiabilidade nos resultados obtidos.

2.1.1 Caracterização Estatística dos Dados Acadêmicos

Os dados acadêmicos foram extraídos do Sistema de Controle Acadêmico Online (SCAO/UFCG), abrangendo o histórico curricular de 352 egressos que concluíram o curso entre os períodos 2012.1 e 2024.1. A organização e preparação desses dados seguiu um processo sistemático, estruturado em três planilhas complementares:

- A primeira planilha contém os registros individuais de cada egresso em formato matricial, onde as linhas representam os estudantes (identificados por seus números de matrícula) e as colunas representam as 52 disciplinas optativas. Utilizou-se codificação binária para indicar as disciplinas cursadas (valor 1) e não cursadas (valor 0), permitindo visualizar o perfil curricular completo de cada estudante. Cada egresso deve completar um mínimo de 10 disciplinas optativas para integralização do curso;
- A segunda planilha foi dedicada à categorização das disciplinas, contendo os códigos identificadores, nomenclaturas oficiais e classificações por área temática. Esta planilha consolidou a listagem das disciplinas optativas oferecidas no período;
- A terceira planilha sintetiza os dados quantitativos, contabilizando o número de disciplinas cursadas por cada egresso em cada área temática. Esta consolidação transformou a matriz binária inicial em atributos numéricos, facilitando a aplicação de técnicas de clusterização.

A categorização das disciplinas foi conduzida com base em um processo empírico e colaborativo em três etapas: (1) análise das ementas oficiais, identificando objetivos de aprendizagem, conteúdos e competências técnicas; (2) validação com docentes das respectivas áreas e coordenação do curso; e (3) análise exploratória das matrículas históricas para verificar a consistência dos agrupamentos propostos. O resultado desse processo gerou quatro áreas temáticas:

O resultado desse processo é apresentado a seguir:

- **DEV (Desenvolvimento):** Inclui disciplinas voltadas ao desenvolvimento de software, abrangendo linguagens de programação, engenharia de software, testes, arquitetura de sistemas e metodologias de desenvolvimento. Exemplos: Programação Funcional, Verificação e Validação de Software, Engenharia de Software II e Programação para Web.
- **IA (Inteligência Artificial):** Reúne disciplinas com foco em técnicas de aprendizado de máquina, mineração de dados, visão computacional, Processamento de Linguagem Natural (PLN) e construção de sistemas inteligentes. Exemplos: Aprendizado de Máquina, Ciência de Dados Preditiva e Processamento de Linguagem Natural.

- **INFRA (Infraestrutura Computacional):** Contempla conteúdos relacionados a redes de computadores, segurança, computação em nuvem, sistemas operacionais e administração de sistemas. Exemplos: Administração de Sistemas, Sistemas Distribuídos e Segurança de Sistemas.
- **GERAL (Generalistas):** Abrange disciplinas com abordagem transversal, interdisciplinar ou voltada a competências complementares, como gestão, empreendedorismo, ética ou impacto social da tecnologia. Exemplos: Empreendedorismo em Software, Informática e Sociedade e Gestão de Projetos.

Essa organização em áreas não apenas reflete os eixos tradicionais de formação na Ciência da Computação, mas também se alinha às demandas contemporâneas do mundo do trabalho, permitindo identificar padrões formativos compatíveis com diferentes transições profissionais. Os agrupamentos das disciplinas optativas podem ser consultados no Apêndice A.

A Tabela 2.1 apresenta a estatística descritiva das disciplinas cursadas por área, revelando predominância da área de Desenvolvimento (média de 3,92 disciplinas por egresso), seguida pelas áreas Generalista (3,21), Inteligência Artificial (1,88) e Infraestrutura (1,84). Essa distribuição reflete a centralidade de DEV na formação, bem como a variabilidade de interesses entre os demais eixos temáticos.

Estatística	DEV	INFRA	IA	GERAL
Contagem	352	352	352	352
Média	3,92	1,84	1,88	3,21
Desvio Padrão	1,30	1,31	1,43	1,43
Mínimo	0,00	0,00	0,00	0,00
1º Quartil	3,00	1,00	1,00	2,00
Mediana	4,00	2,00	2,00	3,00
3º Quartil	5,00	3,00	3,00	4,00
Máximo	7,00	6,00	7,00	6,00

Tabela 2.1: Informação estatística dos dados acadêmicos.

No processo de tratamento dos dados, foram aplicados dois procedimentos principais:

1. Remoção da categoria GERAL das análises de clusterização, para reduzir ruídos e concentrar a análise nas áreas técnicas mais diretamente associadas à formação em Computação. As disciplinas classificadas como GERAL, embora importantes para a formação integral, apresentavam maior heterogeneidade temática e menor potencial discriminativo para identificação de perfis técnicos.
2. Eliminação de disciplinas que não apresentavam nenhuma matrícula registrada entre todos os egressos analisados (totalizando zero cursantes). Esta filtragem removeu elementos sem valor informativo para a análise de padrões, mantendo todas as disciplinas que tenham sido cursadas por pelo menos um egresso.

Após estas etapas, cada egresso foi representado por um vetor tridimensional contendo o número de disciplinas cursadas nas áreas de DEV, INFRA e IA. A Tabela 2.2 apresenta a estrutura final da matriz de dados utilizada nas etapas subsequentes de análise.

Coluna	Contagem não nula	Tipo
dev	352	int64
infra	352	int64
ia	352	int64

Tabela 2.2: Estrutura da matriz de dados acadêmicos após tratamento.

Esta estruturação garantiu que os dados estivessem adequados tanto para as análises estatísticas descritivas quanto para a aplicação dos algoritmos de agrupamento nas fases subsequentes da pesquisa.

2.1.2 Caracterização Estatística dos Dados Profissionais

A coleta dos dados profissionais dos egressos foi realizada por meio de uma abordagem triangulada, que combinou duas enquetes aplicadas em momentos distintos e a análise manual de perfis públicos no LinkedIn. O objetivo dessa estratégia foi ampliar a cobertura dos dados e oferecer uma visão mais abrangente das transições profissionais, assegurando consistência entre diferentes fontes.

Enquetes Aplicadas aos Egressos

A primeira enquete (Apêndice B), realizada na fase inicial da pesquisa, foi voltada aos egressos formados entre 2001.2 e 2022.1. Disponibilizada via Google Forms, ela foi divulgada por e-mail para os ex-alunos com apoio da coordenação do curso. Essa enquete incluiu perguntas sobre: (i) o tempo decorrido entre a graduação e o primeiro emprego na área de Computação; (ii) área de atuação inicial; (iii) modalidade de trabalho (presencial, híbrida ou remota); (iv) porte da empresa; (v) tipo de cargo e (vi) tipo de vínculo empregatício. Obteve-se um total de 296 respostas, de um universo de 1.115 egressos, representando uma taxa de retorno de 26,5%.

A segunda enquete (Apêndice C), aplicada após a qualificação da pesquisa, foi direcionada aos 352 egressos do período de 2012.1 a 2024.1, foco principal desta análise. Além das perguntas da primeira enquete, essa versão incorporou itens sobre a transição profissional posterior à primeira ocupação e sobre a percepção dos egressos quanto à contribuição do curso para sua atuação profissional. Essa etapa contou com 59 respostas (16,8% do universo), sendo utilizada prioritariamente para complementar e validar as demais fontes.

Coleta Manual e Sistematização de Dados do LinkedIn

Como complemento às enquetes, foi conduzido um processo manual de levantamento de dados a partir do LinkedIn. A coleta ocorreu de forma individualizada, sem uso de automações ou APIs, com o objetivo de preservar a qualidade das informações. Essa escolha, embora mais trabalhosa, permitiu verificar cada caso com atenção e manter a fidelidade entre os dados observados e o histórico acadêmico do egresso.

O processo seguiu os seguintes passos:

- **Identificação dos perfis:** Utilizou-se a lista oficial dos egressos fornecida pela coordenação, contendo nome completo, ano de ingresso e de conclusão do curso. A partir desses dados, buscou-se individualmente por cada egresso no LinkedIn, utilizando critérios como instituição (UFCG) e curso (Ciência da Computação) para validar a identificação;
- **Extração manual de dados:** Para cada perfil identificado, coletaram-se informações relevantes sobre a trajetória profissional, incluindo o primeiro cargo ocupado após a

graduação, nome da empresa, data de início, modalidade de trabalho, local de atuação e evolução da carreira. No total, foram analisados 252 perfis completos;

- Registro padronizado: As informações foram organizadas em planilhas estruturadas, permitindo comparações e análises posteriores. Sempre que possível, as datas de início foram cruzadas com a data de conclusão da graduação para calcular o tempo entre a formação e a inserção no mundo do trabalho;
- Categorização dos cargos: A nomenclatura original dos cargos foi mantida para registro, mas cada ocupação foi posteriormente alocada em subcategorias específicas (como “Desenvolvedor Back-end”, “Cientista de Dados” etc.) e, em seguida, agrupada em seis categorias analíticas principais, conforme descrito adiante.

A triangulação entre enquetes e LinkedIn permitiu obter dados profissionais completos para 263 dos 352 egressos com histórico acadêmico analisado (74,7% do total). A validação cruzada entre a segunda enquete e os perfis do LinkedIn indicou 92% de concordância quanto à primeira ocupação profissional. Nos casos de divergência, priorizou-se a resposta da enquete por ser autodeclarada e, em geral, mais precisa quanto a datas e modalidades de trabalho.

Para os 89 egressos restantes, adotou-se tratamento específico:

- 17 seguiram diretamente para programas de pós-graduação;
- 12 atuaram em projetos de pesquisa sem vínculo formal com pós-graduação;
- 60 não apresentaram dados verificáveis em nenhuma das fontes.

Esses casos foram excluídos da análise de transição profissional, mas mantidos na análise acadêmica quando possível.

Categorização e Transformação dos Dados Profissionais

As ocupações foram consolidadas em seis categorias principais, de modo a evitar dispersão excessiva e facilitar a análise estatística:

- Desenvolvimento e Engenharia de Software (175 egressos, 66,5%);

- Análise e Suporte em TI (36 egressos, 13,7%);
- Segurança e Infraestrutura (26 egressos, 9,9%);
- Ciência de Dados e IA (22 egressos, 8,4%);
- Gerência (3 egressos, 1,1%);
- Pesquisa e Desenvolvimento (2 egressos, 0,8%).

Esses agrupamentos foram definidos com base na frequência das denominações e similaridade funcional entre os cargos. Por exemplo, na categoria de Desenvolvimento e Engenharia de Software, foram incluídas funções como Desenvolvedor Front-End, Full-Stack e Engenheiro de Software em diversos níveis.

As variáveis utilizadas na análise foram tratadas conforme as exigências do algoritmo K-means: as ordinais receberam codificação numérica hierárquica (ex: 0 para "até 6 meses"), enquanto as nominais foram transformadas via *One-Hot Encoding*. Os dados foram normalizados com *StandardScaler*, garantindo contribuição equivalente de todas as variáveis na formação dos clusters. Registros incompletos ou categorias com poucos dados foram tratados adequadamente para manter a coerência da base.

2.1.3 Técnicas de Pré-processamento e Análise

A escolha por centrar a análise nas disciplinas optativas como principal variável acadêmica está fundamentada em quatro aspectos metodológicos. Em primeiro lugar, essas disciplinas representam escolhas intencionais dos estudantes, expressando preferências individuais, interesses e, muitas vezes, estratégias de formação voltadas a áreas específicas da Computação. Ao contrário das disciplinas obrigatórias, que são comuns a todos os alunos, as optativas introduzem variabilidade nos percursos formativos, tornando-se uma oportunidade analítica para identificar possíveis associações entre decisões curriculares e transições profissionais. Naturalmente, essas escolhas estão limitadas pela oferta de disciplinas optativas em cada período letivo. O Apêndice A mostra os períodos de oferta das várias disciplinas optativas cursadas pelos egressos estudados.

Em segundo lugar, o bloco de disciplinas optativas corresponde a uma parcela considerável da carga horária total do curso, aproximadamente 20%. Essa margem de flexibilidade

permite que cada aluno componha um percurso alinhado às suas habilidades e aspirações, dentro dos limites curriculares definidos. Como todos os estudantes têm acesso ao mesmo conjunto de disciplinas optativas (respeitando pré-requisitos e disponibilidade por semestre), essa variável se mostra adequada para comparações entre perfis de formação.

Outro fator relevante é a disponibilidade e padronização dos dados, as informações sobre disciplinas optativas estão registradas de forma sistemática no histórico acadêmico, o que garante completude e consistência em toda a amostra de egressos, mesmo em análises retrospectivas que abrangem mais de uma década. Esse nível de padronização é uma condição favorável para a aplicação de técnicas estatísticas e de aprendizado de máquina.

Por fim, observa-se que as disciplinas optativas frequentemente refletem áreas técnicas complementares ao núcleo obrigatório do curso. Elas funcionam, assim, como um possível marcador de especialização e podem contribuir para o desenvolvimento de competências associadas a setores específicos do mundo do trabalho, sendo relevantes para análises sobre inserção profissional.

Embora outros fatores como participação em projetos de pesquisa, estágios e monitorias também possam influenciar a transição profissional, estes foram excluídos devido a limitações metodológicas, incluindo a heterogeneidade de formatos, fragmentação dos registros, lacunas históricas na documentação e critérios de acesso não uniformes que comprometeriam a consistência da análise:

- Variedade de formatos e durações: Diferente das disciplinas optativas, essas atividades não seguem um padrão uniforme quanto à carga horária, duração, ou objetivos, o que dificulta sua comparação direta ou quantificação padronizada;
- Fragmentação e acesso restrito aos registros: As bases de dados que concentram informações sobre estágios, projetos e intercâmbios são descentralizadas e variam em estrutura e completude, dificultando sua consolidação em uma base única e consistente para todos os egressos;
- Lacunas históricas nos registros: Para egressos formados em anos anteriores, especialmente entre 2012 e 2016, constatou-se uma escassez de registros digitalizados, o que tornaria a amostra parcial e potencialmente enviesada;

- Critérios de acesso restritivos: Atividades como intercâmbio acadêmico, pesquisa científica ou monitorias exigem seleção prévia, disponibilidade de vagas e, por vezes, apoio institucional ou financeiro. Isso significa que nem todos os estudantes têm as mesmas oportunidades de participação, o que poderia introduzir distorções ao considerar essas variáveis como fatores comparáveis.

Diante dessas limitações, optou-se por focar nas disciplinas optativas, que se mostraram mais adequadas do ponto de vista metodológico. Reconhece-se, contudo, que a ausência de dados sobre atividades extracurriculares pode restringir a compreensão de certos aspectos das transições profissionais, especialmente aqueles relacionados à vivência prática ou à formação complementar. Ainda assim, entende-se que a abordagem adotada é válida e justificada diante das condições de coleta e da necessidade de manter consistência nos dados analisados.

Para garantir a qualidade da análise multivariada, seguindo Camacho et al. [7], aplicamos técnicas específicas de pré-processamento, incluindo identificação de outliers pelo método IQR. Estas etapas preparatórias viabilizaram a aplicação de técnicas de redução de dimensionalidade (PCA) e agrupamento (K-means), facilitando a identificação de padrões significativos nos dados acadêmicos e profissionais.

Em consonância com os princípios de ciência aberta, todos os dados utilizados na pesquisa, devidamente anonimizados, estão disponíveis para consulta e replicação no repositório GitHub do projeto ¹, junto aos notebooks de análise e documentação técnica.

2.2 Análise de Transições Profissionais

A transição dos egressos de Ciência da Computação para o mundo do trabalho constitui um processo multifacetado, influenciado por diversos fatores que incluem o conjunto de disciplinas cursadas, experiências de estágio, projetos de pesquisa e extensão, intercâmbios, certificações complementares e redes de relacionamento profissional desenvolvidas durante a graduação [19]. Esta complexa interação de elementos demanda uma análise sistemática para compreender como diferentes aspectos da formação acadêmica impactam a transição profissional.

¹https://github.com/felis-cc/dissertacao_ppgcc_ufcg.git

A literatura contemporânea enfatiza a necessidade crítica de alinhamento entre formação acadêmica e demandas do setor produtivo como condição fundamental para uma inserção profissional efetiva [48]. Particularmente relevante é a observação de que profissionais que desenvolvem um perfil híbrido, combinando competências técnicas robustas com habilidades de comunicação e gestão, tendem a ascender mais rapidamente a posições de maior responsabilidade e remuneração.

Este cenário é enriquecido pela diversidade de transições possíveis no campo da Tecnologia da Informação (TI), que abrange desde posições em grandes corporações até oportunidades em *startups*, iniciativas empreendedoras e carreiras acadêmicas. Neste contexto, a flexibilidade curricular emerge como um facilitador crucial, permitindo que os estudantes moldem seu perfil profissional de acordo com suas aspirações de carreira.

A complexidade destas transições tem motivado o desenvolvimento de abordagens quantitativas mais sofisticadas para análise de transições profissionais. Marques e Silva [27] demonstraram a eficácia da aplicação de técnicas de Mineração de Dados Educacionais (MDE), particularmente o algoritmo *K-Means*, para identificar padrões de empregabilidade. Esse estudo, que considerou variáveis como tempo de inserção no mundo do trabalho, setor de atuação e faixa salarial, estabeleceu bases metodológicas para a identificação de perfis profissionais recorrentes.

Complementarmente, Santos et al. [42] introduziram uma abordagem inovadora combinando Análise por Envoltória de Dados (DEA) com K-Means para avaliar a eficiência educacional. Seus resultados estabeleceram padrões replicáveis de excelência institucional, contribuindo para a compreensão quantitativa das trajetórias acadêmicas.

No âmbito da segmentação de perfis profissionais, Gupta e Grover [49] identificaram uma dicotomia significativa entre estudantes de engenharia: os "*Practicable Students*", que demonstram forte alinhamento com as demandas do mundo do trabalho, e os "*Impracticable Students*", que enfrentam dificuldades de inserção profissional devido a lacunas em sua preparação. Esta classificação ressalta a importância de fatores demográficos, como idade e gênero, no desenvolvimento de competências profissionais.

Gannod et al. [17] aprofundam esta discussão ao demonstrar que a mera inclusão de novos componentes curriculares não garante o desenvolvimento das competências essenciais para a empregabilidade. Os autores enfatizam a necessidade de reformulações estruturais

que transcendam a simples adição de disciplinas, destacando a importância de uma revisão sistêmica dos cursos obrigatórios para assegurar sua relevância frente às exigências do mundo do trabalho.

Avanços recentes em aprendizado de máquina têm possibilitado abordagens mais sofisticadas para a análise de transições e trajetórias profissionais. Al-Dossari et al. [3] desenvolveram o *CareerRec*, um sistema preditivo que utiliza diversos algoritmos (KNN, Decision Trees, Gradient Boosting e XGBoost) para recomendar trajetórias profissionais. Baseado na análise de 2.255 profissionais de TI, o sistema demonstrou eficácia na previsão de caminhos profissionais entre as carreiras de desenvolvedor, analista e engenheiro.

Os desafios contemporâneos da transição profissional são extensivamente analisados por Safavi, Davoodi e Koutra. [41], que identificam fatores críticos como a prevalência de contratos temporários, períodos de desemprego intermitente e a necessidade constante de atualização profissional. Esta análise é complementada por Rätty, et al. [38], que introduzem o conceito de "*ability self*", a autopercepção dos estudantes sobre suas competências e adequação ao mundo do trabalho.

Blokker et al. [6] expandem esta perspectiva ao investigar a relação entre planejamento de carreira, proatividade e percepção de empregabilidade. Seus achados indicam que, mesmo conscientes dos desafios profissionais, muitos estudantes não adotam posturas proativas para desenvolver suas competências, evidenciando a necessidade de políticas institucionais que fomentem o planejamento de carreira e o engajamento em atividades extracurriculares.

Contribuições recentes têm explorado o potencial de sistemas baseados em aprendizado de máquina para orientação profissional. Kamal et al. [22] desenvolveram um sistema de recomendação para auxiliar decisões acadêmico-profissionais, enquanto Bhaskaran et al. [5] apresentaram um modelo preditivo de carreiras baseado em múltiplas variáveis acadêmicas e extracurriculares, utilizando algoritmos como Naïve Bayes e KNN.

No estudo de Williams et al. [13] são oferecidas perspectivas complementares sobre a empregabilidade em Tecnologias da Informação e Comunicação (TIC), destacando tanto lacunas entre formação técnica e demandas práticas quanto a importância do capital humano e da autogestão profissional. Maree et al. [26] contribuem com uma dimensão adicional ao enfatizar o papel da adaptabilidade de carreira como mediador entre competências e empregabilidade.

A análise quantitativa das trajetórias profissionais fundamenta-se em indicadores mensuráveis que permitem avaliar o impacto das escolhas acadêmicas na carreira:

- Tempo de inserção no mundo do trabalho: mensura o intervalo entre conclusão do curso e primeira atuação profissional;
- Evolução salarial: acompanha a progressão da remuneração ao longo da carreira;
- Estabilidade no emprego: avalia o tempo médio de permanência em posições profissionais;
- Mobilidade profissional: analisa a frequência e natureza das transições entre diferentes áreas do setor de TI.

A objetividade na análise de carreiras é assegurada pela aplicação de técnicas quantitativas robustas:

- Clusterização com K-Means: permite a identificação de padrões e agrupamentos de perfis similares;
- Análise de Regressão Logística: possibilita previsões de empregabilidade baseadas em múltiplas variáveis;
- Redes Bayesianas: modelam probabilidades de transições entre estágios da carreira.

A eficácia da clusterização via K-Means na identificação de padrões de empregabilidade é corroborada por estudos recentes [27];[42], especialmente quando complementada por técnicas estatísticas adicionais como regressão logística e ANOVA.

2.3 Mineração de Dados Educacionais

A necessidade de compreender padrões na inserção profissional dos egressos impulsionou o desenvolvimento de abordagens analíticas robustas, sendo a Mineração de Dados Educacionais (EDM, do inglês *Education Data Mining*) uma metodologia amplamente utilizada. Gupta e Grover [49] demonstram que a aplicação de técnicas de clusterização aos perfis de empregabilidade permite uma segmentação eficaz dos estudantes conforme suas perspectivas

profissionais e nível de preparação para o mundo do trabalho, estabelecendo bases quantitativas para o desenvolvimento de estratégias acadêmicas mais efetivas.

A validação curricular, conforme argumentam Gannod et al. [17], constitui um componente essencial da EDM, transcendendo a simples avaliação do conteúdo ministrado para examinar sua adequação às demandas do mundo do trabalho. Esta abordagem possibilita a identificação sistemática de lacunas curriculares e a proposição de intervenções fundamentadas em evidências quantitativas.

No escopo específico desta pesquisa, a EDM desempenha papel crucial na análise da correlação entre disciplinas cursadas e sucesso profissional dos egressos. Santos et al. [42] demonstram como a integração entre clusterização e Análise por Envoltória de Dados (DEA, do inglês *Data Envelopment Analysis*) permite avaliar a eficiência de unidades educacionais, estabelecendo métricas objetivas para validação de padrões de sucesso entre egressos.

No estudo de Vista, Chicon e Mozzaquatro [50] é expandida esta perspectiva ao demonstrar a aplicabilidade da clusterização na segmentação de perfis decisórios em ambientes corporativos. Esta metodologia, quando adaptada ao contexto educacional, oferece percepções valiosas sobre a relação entre formação acadêmica e inserção profissional. Seus resultados, baseados em questionários estruturados, evidenciam a necessidade de métodos quantitativos robustos para a análise de transições acadêmico-profissionais.

A relevância da EDM na identificação de discrepâncias entre formação acadêmica e empregabilidade é corroborada por Al-Dossari et al. (2020), que empregam técnicas avançadas para análise preditiva de carreiras em TI. Complementarmente, Zubair e Moazzam [54] destacam como a exposição a trajetórias profissionais reais influencia positivamente o processo decisório dos estudantes, reforçando a importância de iniciativas que aproximem a academia e o mundo do trabalho.

2.3.1 Abordagens Metodológicas em EDM

A aplicação de métodos avançados de aprendizado de máquina tem ampliado o potencial analítico da EDM. Kamal et al. [22] demonstram a eficácia de algoritmos como Random Forest e XGBoost na previsão de trajetórias profissionais, estabelecendo correlações robustas entre competências acadêmicas e demandas do mundo do trabalho. Esta perspectiva é complementada por Gao et al. [18], que evidenciam o potencial da EDM na avaliação do

impacto de programas acadêmicos sobre a empregabilidade.

Williams et al. [13] enriquecem esta discussão ao examinar como variáveis demográficas e acadêmicas influenciam a percepção de empregabilidade, fornecendo subsídios para uma segmentação mais precisa e modelagem preditiva mais acurada. Bhaskaran et al. [5] reforçam esta abordagem ao demonstrar como a mineração de dados pode otimizar o alinhamento entre escolhas curriculares e oportunidades profissionais.

A análise dos clusters não apenas identifica padrões, mas também serve como ferramenta para tomada de decisão acadêmica e institucional. Algumas das aplicações incluem:

- **Aconselhamento Acadêmico Personalizado:** Estudantes podem receber recomendações baseadas em padrões de sucesso identificados em alunos anteriores.
- **Apoio na Formulação de Políticas Educacionais:** Os dados podem ser usados por gestores acadêmicos para ajustar currículos e definir novas estratégias de ensino.
- **Monitoramento Contínuo de Desempenho:** A análise periódica dos clusters permite acompanhar o impacto das mudanças curriculares ao longo do tempo.

Por exemplo, um estudo de Santos et al. [42] demonstrou como a segmentação de estudantes com base em sua trajetória acadêmica permitiu a reformulação de currículos em universidades brasileiras, resultando em uma melhoria de 15% na taxa de empregabilidade dos egressos.

Dessa forma, a mineração de dados educacionais se torna uma ferramenta essencial para compreender e aperfeiçoar a trajetória acadêmica e a transição profissional dos estudantes. A próxima seção explora os fundamentos matemáticos da clusterização, detalhando como os algoritmos aplicados no estudo processam e organizam os dados educacionais.

A mineração de dados educacionais é uma área em crescimento que busca extrair padrões e percepções a partir de grandes volumes de dados acadêmicos, auxiliando na compreensão de tendências e na tomada de decisões para otimização curricular e empregabilidade. No contexto desta pesquisa, as técnicas de mineração de dados são fundamentais para analisar a relação entre as disciplinas cursadas e a transição profissional dos egressos.

2.3.2 Métodos de Clusterização em EDM

A implementação efetiva da EDM requer uma seleção criteriosa dos métodos de clusterização. A análise comparativa dos principais métodos revela características distintivas que influenciam sua aplicabilidade em diferentes contextos, como mostrado na tabela 2.3:

Método	Vantagens	Limitações
K-Means	Simplicidade e eficiência com grandes volumes de dados; Interpretabilidade dos resultados.	Sensibilidade a outliers; Necessidade de definição prévia do número de clusters.
DBSCAN	Identificação de clusters com formas arbitrárias; Detecção natural de outliers.	Desempenho limitado em alta dimensionalidade.
Hierarchical Clustering	Flexibilidade na definição do número de clusters; Visualização hierárquica clara.	Alto custo computacional para grandes conjuntos de dados.

Tabela 2.3: Comparação entre métodos de clusterização.

Segundo Brum, Mozzaquatro e Zanatta [10], seus estudos fornecem evidência empírica da relação entre a estrutura dos dados e a eficácia dos algoritmos de clusterização. Seus resultados confirmam a superioridade do K-Means para a análise de grandes volumes de dados educacionais.

A escolha da métrica de distância apropriada representa outro aspecto crucial da clusterização. Vista, Chicon e Mozzaquatro [50], em sua análise dos microdados do ENADE, exploram sistematicamente as implicações de diferentes métricas: Euclidiana, Manhattan, Minkowski e Chebyshev. Demonstrando seu impacto na qualidade dos agrupamentos resultantes.

A adoção do K-Means nesta pesquisa fundamenta-se em sua comprovada eficiência na análise de grandes conjuntos de dados educacionais, permitindo agrupamentos consistentes baseados em padrões de formação e empregabilidade [51].

A confiabilidade dos resultados da clusterização depende de um rigoroso processo de pré-processamento dos dados educacionais, que compreende: Normalização e Escalonamento

- Padronização de atributos com diferentes escalas;
- Eliminação de viés dimensional na clusterização.

Tratamento de *Outliers*

- Identificação de valores atípicos por métodos estatísticos;
- Mitigação de distorções nos agrupamentos.

Seleção de Características

- Remoção de atributos redundantes ou irrelevantes;
- Otimização da eficiência computacional.

Transformação de Dados

- Codificação de variáveis categóricas;
- Preparação para aplicação de algoritmos de aprendizado.

Santos et al. [42] demonstram em seu estudo como a integração entre K-Means e Análise por DEA pode potencializar a análise de eficiência educacional, reforçando a importância de abordagens híbridas no processamento de dados educacionais.

Esta estrutura metodológica fornece uma base sólida para a análise quantitativa das relações entre formação acadêmica e a transição profissional, permitindo identificar padrões significativos e orientar decisões curriculares fundamentadas em evidências.

2.4 Fundamentos Matemáticos da Clusterização

A clusterização, como técnica de aprendizado de máquina não supervisionado, constitui-se em uma ferramenta fundamental para identificar padrões latentes em grandes volumes de dados. No contexto específico desta pesquisa, sua aplicação permite analisar como as escolhas acadêmicas das disciplinas optativas dos egressos do curso de Ciência da Computação da UFCG influenciam suas transições profissionais, através da segmentação baseada em características compartilhadas.

Esta abordagem fundamenta-se na capacidade dos algoritmos de clusterização de agrupar observações similares, simultaneamente reduzindo a variabilidade intragrupo e maximizando as diferenças intergrupos. O K-Means foi selecionado por sua eficiência computacional, sua robustez na segmentação de grandes conjuntos de dados e pela capacidade de gerar agrupamentos interpretáveis. Embora existam alternativas, como DBSCAN e Hierarchical Clustering, estudos como o de Gupta e Grover [49] demonstram que o K-Means apresenta desempenho superior para conjuntos de dados de alta dimensionalidade que se considera em dados com mais de 10 a 15 dimensões, no entanto, esse valor não é rigorosamente definido e pode variar dependendo do contexto específico de como os utilizados durante esta pesquisa, pois o K-Means enfrenta desafios significativos com alta dimensionalidade devido ao fenômeno conhecido como a "maldição da dimensionalidade". É altamente recomendável aplicar técnicas de redução de dimensionalidade como PCA (Análise de Componentes Principais) ou t-SNE antes de executar o K-Means para obter resultados mais confiáveis e interpretáveis (inserir referência).

No estudo de Santos et al. [42], estes validam uma abordagem integrada, demonstrando como a aplicação da Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*) previamente ao K-Means e à DEA aprimora significativamente a interpretabilidade dos agrupamentos educacionais. Seus resultados confirmam a eficácia da PCA na redução dimensional sem comprometimento significativo da informação original.

Porém, Gupta e Grover [49] aprofundam esta discussão ao identificar fatores críticos para uma segmentação eficiente, destacando particularmente a importância da escolha adequada da métrica de distância e do número de clusters. Sua aplicação do K-Means para a classificação de estudantes fornece evidências empíricas da capacidade do algoritmo em revelar padrões significativos em trajetórias acadêmicas.

A robustez desta abordagem é reforçada por Kamal et al. [22], que introduzem o uso de SMOTE para balanceamento de dados, enquanto Bhaskaran et al. [5] oferecem uma análise comparativa entre K-Means e técnicas supervisionadas como KNN e Naïve Bayes. Esta comparação evidencia a adequação do K-Means para identificação de padrões emergentes em dados não rotulados, contexto típico da análise de trajetórias educacionais.

Santos e Oliveira [11] exemplificam a versatilidade do K-Means na segmentação de perfis estratégicos, demonstrando sua eficácia na categorização de trajetórias acadêmicas e escolhas

curriculares. Esta aplicabilidade é corroborada por Gao et al. [18], que evidenciam como a segmentação de perfis de graduados revela relações significativas entre formação acadêmica e transições profissionais.

Algoritmo K-Means

O algoritmo K-Means representa uma abordagem iterativa para particionamento de dados em k grupos distintos, fundamentando-se no princípio de minimização da variabilidade intracluster. Sua implementação segue um processo sistemático estruturado em quatro etapas fundamentais ²:

1. **Inicialização:** Seleção de k centróides iniciais.
2. **Atribuição:** Cada ponto de dado é atribuído ao cluster cujo centróide está mais próximo, com base em uma métrica de distância.
3. **Atualização:** Recalcula-se o centróide de cada cluster com base nos pontos atribuídos.
4. **Convergência:** Repete-se os passos de atribuição e atualização até que os centróides não mudem significativamente.

A métrica de distância mais comumente utilizada é a **distância euclidiana**, definida como:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.1)$$

onde x e y representam vetores de atributos no espaço de características, constituem a base para o cálculo de similaridade entre observações. No entanto, a seleção da métrica apropriada deve considerar as características específicas dos dados em análise.

A definição do número ideal de clusters é um desafio central na clusterização, pois um número inadequado pode comprometer a segmentação e prejudicar a análise dos padrões extraídos. Para garantir a melhor separação dos grupos, duas métricas amplamente utilizadas são o Índice de Silhueta e o Método do Cotovelo.

O Índice de Silhueta estabelece-se como métrica fundamental para validação quantitativa dos agrupamentos gerados por algoritmos de clusterização, esta medida avalia dois aspectos

²https://docs.google.com/document/d/1co7vtVdMGV4lmQqG5FXM3H2pr46_DDQY/edit.

críticos da qualidade dos clusters: a coesão interna, determinada pela proximidade entre elementos do mesmo grupo, e a separação externa, definida pelo distanciamento entre clusters distintos.

A pesquisa de Zhang et al. [53] implementou o Índice de Silhueta para avaliar a qualidade dos clusters formados em um modelo de recomendação personalizada de exercícios para estudantes. O estudo demonstrou que valores mais altos do índice indicam grupos mais coesos e bem segmentados, enquanto valores baixos sugerem a necessidade de ajuste no número de clusters ou no método de agrupamento utilizado.

Matematicamente, o Índice de Silhueta para um ponto é definido como:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2.2)$$

onde:

- $a(i)$ representa a dissimilaridade média intracluster: distância média entre o ponto i e todos os outros pontos em seu próprio cluster;
- $b(i)$ denota a dissimilaridade mínima intercluster: menor distância média entre o ponto i e os pontos de qualquer outro cluster.

O índice resultante varia no intervalo $[-1, 1]$. Esta métrica assegura que os agrupamentos gerados pelo K-Means apresentem significância estatística e reflitam efetivamente os padrões de diferenciação entre trajetórias acadêmicas e profissionais. Valores próximos a 1 indicam clusters bem separados e compactos, sugerindo uma boa segmentação, enquanto que valores próximos a 0 indicam sobreposição entre clusters, apontando a necessidade de reavaliação do número de grupos; por fim, os valores negativos sugerem que o ponto pode estar alocado no cluster errado, indicando uma má segmentação.

O Método do Cotovelo é outra abordagem amplamente utilizada para determinar o número ideal de clusters. Ele se baseia na análise da inércia intra-cluster, definida como a soma das distâncias quadráticas entre cada ponto e seu centróide. O objetivo é identificar o ponto de inflexão no gráfico, onde o acréscimo de novos clusters deixa de trazer benefícios significativos. Estudos como o de Macêdo et al. [25] aplicaram o Método do Cotovelo para validar a segmentação de perfis de engajamento em ambientes virtuais de aprendizagem. A pesquisa demonstrou que a escolha adequada do número de clusters é essencial para garantir agrupamentos coerentes e representativos dos padrões de comportamento estudantil.

A aplicação do Método do Cotovelo segue os seguintes passos:

1. Para avaliar a compactação dos clusters formados, a inércia é calculada para diferentes valores de k , variando, por exemplo, de 1 a 10 clusters.;
2. **Calcular a inércia** para cada valor de k , onde:

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2.3)$$

sendo:

- $WCSS$ (Within-Cluster Sum of Squares) representa a soma das distâncias quadráticas entre os pontos e seus respectivos centróides;
 - C_i é um cluster específico;
 - μ_i é o centróide do cluster i ;
 - x representa um ponto de dado pertencente ao cluster C_i ;
 - $\|x - \mu_i\|^2$ denota a distância euclidiana ao quadrado entre cada ponto x e seu centróide μ_i .
3. Plotar um gráfico de WCSS em função do número de clusters k ;
 4. Identificar o ponto de cotovelo, ou seja, onde a redução da inércia começa a se estabilizar. Esse ponto representa o número ideal de clusters, pois adicionar mais grupos a partir desse ponto não reduz significativamente a inércia.

A pesquisa de Santos et al. [42] utilizou o Método do Cotovelo para determinar o número ótimo de clusters em uma análise de empregabilidade de egressos universitários. Os resultados indicaram que, ao empregar essa abordagem, foi possível identificar um valor de k que maximizava a separação dos grupos sem comprometer a coesão interna.

A comparação entre ambos os métodos mostra seu desempenho e papéis complementares na definição do número ideal de clusters. O Índice de Silhueta mede a qualidade da separação entre grupos já definidos, enquanto o Método do Cotovelo auxilia na escolha inicial do número de clusters a ser utilizado. A tabela 2.4 abaixo resume as principais diferenças entre os dois métodos: Dessa forma, a melhor prática é combinar os dois métodos, utilizando o Método do Cotovelo primeiro para escolher um intervalo adequado de k e a seguir aplicar o

Métrica	Propósito	Vantagens e Limitações
Índice de Silhueta	Avaliar a qualidade da separação entre clusters.	Vantagens: Mede a compactação dos clusters e sua separação. Limitações: Não define o número ideal de clusters, apenas avalia a segmentação.
Método do Cotovelo	Determinar o número ideal de clusters.	Vantagens: Identifica o ponto onde adicionar mais clusters deixa de trazer ganhos significativos. Limitações: Requer interpretação visual do gráfico, o que pode ser subjetivo.

Tabela 2.4: Comparação entre métricas de clusterização.

Índice de Silhueta depois para confirmar se os clusters formados possuem boa separação e coesão.

A combinação dessas técnicas possibilita uma análise mais robusta e confiável, assegurando que os grupos identificados sejam representativos dos padrões observados nos dados educacionais.

Dessa forma, a adoção dessas métricas na análise dos egressos garante que os padrões extraídos sejam estatisticamente relevantes e academicamente aplicáveis. A próxima seção discutirá os fundamentos matemáticos da clusterização, aprofundando a estrutura dos algoritmos aplicados na pesquisa.

2.4.1 Normalização e Escalonamento de Dados

A normalização de dados constitui etapa crítica que precede a aplicação do K-Means, especialmente quando os atributos apresentam escalas heterogêneas. Vista e Figueiró [50] demonstram que a normalização adequada é determinante para a qualidade dos agrupamentos em análises educacionais, garantindo contribuição equitativa de cada atributo na formação dos clusters. Dois métodos de normalização se destacam pela sua eficácia e aplicabilidade:

- A equação abaixo define matematicamente a Normalização Min-Max, esta transformação mapeia os dados para o intervalo $[0, 1]$, preservando todas as relações entre os valores originais:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2.4)$$

- A Padronização (Z-score) é expressa pela fórmula a seguir, essa transformação resulta em uma distribuição com média zero e desvio padrão unitário:

$$z = \frac{x - \mu}{\sigma} \quad (2.5)$$

onde:

- μ representa a média do atributo;
- σ é o desvio padrão do atributo.

Brum, Mozzaquatro e Zanatta [10] fornecem evidência empírica da importância da normalização, demonstrando como atributos não normalizados podem distorcer significativamente os resultados da clusterização. Esta distorção ocorre principalmente pela influência desproporcional de variáveis com maior amplitude de variação.

Santos et al. [42], em sua análise de dados educacionais, validam a eficácia da padronização Z-score, particularmente quando aplicada a variáveis acadêmicas como notas, durações de curso e frequências de participação. Seus resultados demonstram que a normalização adequada é fundamental para identificar padrões significativos em trajetórias educacionais.

A sensibilidade do K-Means a *outliers* e à inicialização dos centróides pode ser mitigada através de técnicas como o *K-Means++*, conforme demonstrado por Gupta e Grover [49]. Sua implementação em contextos educacionais evidencia melhorias significativas na estabilidade e reprodutibilidade dos agrupamentos.

A aplicação do *K-Means* em contextos educacionais transcende a mera análise técnica, estabelecendo-se como ferramenta estratégica para a compreensão de perfis acadêmicos. Vista e Figueiró [50] demonstram sua eficácia na identificação de padrões de aprendizagem significativos, permitindo segmentações que fundamentam intervenções pedagógicas personalizadas.

A robustez desta abordagem é evidenciada por estudos empíricos recentes. Bhaskaran et al. [5] utilizam o K-Means para identificar grupos de estudantes com necessidades de

aprendizagem similares, estabelecendo bases quantitativas para o desenvolvimento de estratégias pedagógicas direcionadas. Este alinhamento entre análise quantitativa e aplicação pedagógica representa um avanço significativo na personalização do processo educacional.

O estudo de Gao et al. [18] expande esta perspectiva ao demonstrar como a segmentação baseada em K-Means pode revelar padrões de desenvolvimento acadêmico que não seriam evidentes através de análises tradicionais. Seus resultados validam a aplicabilidade do método para identificar:

1. Padrões de progressão acadêmica;
2. Necessidades específicas de suporte educacional;
3. Oportunidades de intervenção pedagógica.

Esta fundamentação metodológica, apoiada em evidências empíricas, estabelece as bases para uma análise robusta das trajetórias acadêmicas e transições profissionais dos egressos, objeto central desta pesquisa.

2.4.2 Justificativa para o Uso do K-Means

A adoção do K-Means como algoritmo central desta pesquisa fundamenta-se em evidências empíricas e considerações técnicas específicas ao contexto educacional. Brum, Mozzaquatro e Zanatta [10] fornecem validação experimental desta escolha, demonstrando a superioridade do K-Means na análise de grandes volumes de dados educacionais, particularmente em comparação com algoritmos como DBSCAN e Hierarchical Clustering.

A justificativa para esta escolha metodológica assenta-se em três pilares fundamentais:

1. Eficiência Computacional

- Complexidade temporal expressa como:

$$O(n \times k \times i \times d) \quad (2.6)$$

onde:

- n é o número de pontos de dados;
- k é o número de clusters;

- i representa o número de iterações do algoritmo;
- d é a dimensionalidade dos dados (número de atributos/features);
- Desempenho superior em grandes conjuntos de dados educacionais, conforme validado por Oliveira e Santos [11].

2. Interpretabilidade dos Resultados

- Geração de clusters claramente definidos e interpretáveis;
- Facilidade na tradução de resultados técnicos para recomendações práticas;
- Alinhamento com necessidades de gestão acadêmica, conforme demonstrado por Rodrigues et al. [9].

3. Adequação aos Dados Educacionais

- Eficácia comprovada com variáveis acadêmicas contínuas;
- Robustez na análise de distribuições bem comportadas;
- Compatibilidade com métricas educacionais padronizadas.

Gupta e Grover [49] corroboram esta escolha ao demonstrar a eficácia do K-Means na segmentação de perfis estudantis, particularmente quando aplicado a dados normalizados e bem estruturados, características presentes em nosso conjunto de dados.

2.5 Redução de Dimensionalidade com PCA

A PCA é uma técnica estatística utilizada para reduzir a dimensionalidade dos dados, mantendo a maior parte da variação das informações. No contexto desta pesquisa sobre transição acadêmico-profissionais, a PCA é aplicada antes da clusterização para melhorar a eficiência do algoritmo K-Means, removendo redundâncias e tornando os agrupamentos mais representativos na análise dos perfis dos egressos.

Oliveira e Santos [11] validam esta abordagem ao demonstrarem a importância das técnicas estatísticas na otimização da segmentação de dados educacionais. Seus resultados evidenciam que a aplicação da PCA como etapa preparatória ao processo de agrupamento

permite melhorar significativamente a precisão dos modelos de segmentação, aspecto crucial para a identificação de padrões nas trajetórias acadêmicas e empregabilidade dos egressos.

No contexto específico da Mineração de Dados Educacionais (EDM), Gupta e Grover [49] enfatizam que a pré-seleção de variáveis e a eliminação de atributos redundantes são essenciais para uma segmentação eficaz dos perfis estudantis. Embora seu estudo não utilize diretamente a PCA, suas conclusões reforçam a importância de técnicas estatísticas robustas para otimizar o processo de clusterização, especialmente na análise de padrões acadêmicos e profissionais.

A PCA transforma um conjunto de variáveis correlacionadas em um novo conjunto de variáveis não correlacionadas, denominadas componentes principais. Por exemplo, ao analisar disciplinas optativas cursadas por egressos, algumas podem estar fortemente associadas, como Programação Web e Engenharia de Software. O PCA reduz a redundância nesses dados, agrupando informações correlatas em um único eixo principal, o que facilita a análise posterior por meio do K-Means. Esta transformação é realizada de forma ordenada, garantindo que os primeiros componentes principais retenham a maior parte da variância dos dados originais, característica especialmente relevante para a análise dos perfis acadêmicos dos egressos.

A PCA envolve os seguintes passos matemáticos:

1. Normalização dos dados: Aplicação da padronização *Z-score* para uniformizar as escalas dos atributos, evitando que variáveis com magnitudes diferentes influenciem desproporcionalmente a análise:

$$x' = \frac{x - \mu}{\sigma} \quad (2.7)$$

sendo x' é o valor normalizado, μ é a média e σ é o desvio padrão.

2. Cálculo da Matriz de Covariância: Mensuração das relações entre os atributos do conjunto de dados, estabelecendo a base matemática para identificação dos padrões de variação:

$$C = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)(x_i - \mu_x)^T \quad (2.8)$$

sendo C a matriz de covariância, n o número de amostras x_i os vetores de atributos e $(\hat{})^T$ a operação de transposição, garantindo que a multiplicação resulte em uma matriz e não em um escalar.

3. **Extração de Autovalores e Autovetores:** Determinação dos autovalores (λ) que quantificam a variância explicada por cada componente principal e identificação dos autovetores (v) que definem as direções de máxima variação no espaço de dados:

$$Cv = \lambda v \quad (2.9)$$

4. **Seleção das Componentes Principais:** Ordenação dos componentes em ordem decrescente de autovalores, permitindo selecionar aqueles que capturam a maior parte da variância dos dados:

$$\sum_{i=1}^m \lambda_i \geq \alpha \sum_{i=1}^n \lambda_i \quad (2.10)$$

5. **Projeção dos Dados:** Finalmente a transformação do conjunto original para o novo espaço dimensional através da matriz W , formada pelos autovetores selecionados:

$$Z = W^T X \quad (2.11)$$

sendo que W é a matriz de autovetores selecionados e Z representa os dados transformados.

A aplicação da PCA como etapa preparatória à clusterização via K-Means oferece três benefícios principais para a análise dos perfis acadêmicos:

- **Redução de Ruído:** A PCA remove variáveis irrelevantes ou pouco informativas, evitando que influenciem negativamente os agrupamentos.
- **Melhoria da Eficiência Computacional:** Reduzindo a dimensionalidade dos dados, o tempo de execução do K-Means é otimizado.
- **Evita o Problema da Maldição da Dimensionalidade:** Em conjuntos de dados com muitas variáveis, as distâncias se tornam menos significativas, prejudicando a clusterização. A PCA resolve isso ao transformar os dados em um espaço mais compacto.

Na tabela 2.5, descrevem-se as vantagens e desvantagens do uso da PCA no *K-means*.

Vantagens	Desvantagens
Reduz a dimensionalidade sem perda significativa de informação.	Pode ser difícil interpretar as componentes principais, pois são combinações lineares das variáveis originais.
Remove redundâncias e correlações nos dados.	Para certos tipos de dados, a redução pode levar à perda de informação relevante.
Melhora o desempenho de algoritmos de clusterização como o K-Means.	Exige normalização dos dados para funcionar corretamente.

Tabela 2.5: Vantagens e desvantagens do PCA.

2.5.1 Considerações Finais e Conexão com Trabalhos Relacionados

A integração entre *K-Means* e a redução de dimensionalidade via PCA proporciona uma abordagem robusta para analisar as escolhas das disciplinas curriculares optativas dos egressos e sua relação com a transição profissional. A normalização dos dados assegura que os agrupamentos identificados representem padrões reais, reduzindo distorções estatísticas e garantindo a confiabilidade dos resultados. Essa metodologia tem sido amplamente validada na literatura, incluindo estudos sobre segmentação de perfis educacionais e transições profissionais (Macêdo et al. [25]; Gupta & Grover, [49]).

A pesquisa de Oliveira e Santos [11] corrobora essa abordagem ao demonstrar que a segmentação de perfis acadêmicos pode ser aprimorada pela combinação entre clusterização e análise estatística, permitindo interpretações mais refinadas dos padrões identificados. Esse método, quando aplicado à análise de trajetórias acadêmicas e empregabilidade, oferece novas perspectivas para compreender a transição entre a universidade e o mundo do trabalho.

Os achados desta pesquisa dialogam diretamente com a literatura sobre empregabilidade e mineração de dados educacionais (Macêdo et al. [25]). Em particular, a abordagem proposta por Santos et al. [42] reforça que o uso da DEA associada à clusterização pode aprimorar significativamente a análise da eficiência educacional, permitindo uma interpretação mais aprofundada da relação entre currículos acadêmicos e o sucesso profissional dos egressos.

Dessa forma, a relação entre a formação acadêmica e a empregabilidade continua sendo

um aspecto central na literatura sobre MDE, conforme demonstram estudos como os de Bhaskaran et al. [5] e Raty et al. [38], que destacam a necessidade de abordagens quantitativas e qualitativas para compreender os fatores que influenciam a inserção no mundo do trabalho.

O próximo capítulo apresentará uma análise detalhada de estudos que empregam técnicas de EDM para investigar trajetórias acadêmicas e empregabilidade, consolidando o embasamento teórico-metodológico desta pesquisa e situando-a no estado da arte da Educação em Computação.

Capítulo 3

Trabalhos Relacionados

Dado o crescente interesse acadêmico em compreender as transições profissionais dos egressos de Computação, diversas pesquisas têm explorado o uso de técnicas de mineração de dados e aprendizado de máquina para identificar padrões de empregabilidade. Este capítulo apresenta uma revisão narrativa, focando em três eixos fundamentais: (i) metodologias de análise de transições profissionais via mineração de dados do LinkedIn, (ii) estudos sobre empregabilidade em Computação e seus fatores determinantes, e (iii) alinhamento entre formação acadêmica e expectativas do mundo do trabalho em Tecnologias da Informação e Comunicação (TIC).

A organização deste capítulo reflete a progressão metodológica adotada em nossa pesquisa. Inicialmente, exploramos estudos que fundamentam nossa abordagem de mineração de dados para análise de transições profissionais. Em seguida, examinamos pesquisas sobre empregabilidade em Computação, com ênfase no impacto das escolhas curriculares. Por fim, analisamos trabalhos que investigam a perspectiva dos empregadores, estabelecendo conexões entre formação acadêmica e demandas do mundo do trabalho.

3.1 Análise das Transições Profissionais: Abordagens e Técnicas

A análise de transições profissionais de egressos tem-se beneficiado significativamente do avanço em técnicas de mineração de dados e aprendizado de máquina. Esta seção examina as

principais metodologias e ferramentas empregadas na análise dessas trajetórias, com ênfase especial em estudos que utilizam dados do LinkedIn e técnicas de clusterização.

3.1.1 Mineração de Dados do LinkedIn para Análise de Carreiras

O LinkedIn tem-se estabelecido como uma fonte valiosa de dados para análise de trajetórias profissionais. Macêdo et al. [25] demonstraram a eficácia de técnicas de mineração de dados e clusterização na identificação de padrões de transição de carreira. Seu estudo validou a aplicabilidade do algoritmo K-Means para segmentar grupos profissionais com trajetórias similares, estabelecendo uma base metodológica para previsão de caminhos ocupacionais.

Expandindo esta abordagem, Marques et al. [27] aplicaram técnicas de clusterização para analisar especificamente a empregabilidade de egressos de cursos de computação. Seus resultados evidenciam que a combinação de K-Means com PCA permite uma segmentação mais precisa dos perfis acadêmicos e suas respectivas transições para o mundo do trabalho [27].

Lade et al. [24] contribuíram significativamente para este campo ao desenvolver um sistema automatizado de extração de dados do LinkedIn. Utilizando técnicas de Web Scraping com Selenium e Python, demonstraram como a coleta estruturada de informações acadêmicas e profissionais pode fornecer percepções valiosas sobre padrões de evolução de carreira, tendências de empregabilidade e desenvolvimento de competências. Seu sistema permite uma análise abrangente das redes de conexões profissionais, fornecendo uma base robusta para o estudo de trajetórias ocupacionais.

3.1.2 Técnicas de Clusterização em Análise Educacional

A aplicação de técnicas de clusterização em dados educacionais tem revelado padrões significativos nas transições acadêmico-profissionais. Dombrowskaia et al. [12] investigaram a eficácia destas técnicas no contexto do ensino técnico-profissional, empregando algoritmos como K-Means e X-Means para analisar variáveis demográficas, acadêmicas e socioeconômicas. Seus resultados demonstram que a clusterização pode efetivamente identificar perfis acadêmicos e prever transições profissionais, fornecendo uma base metodológica sólida para análises educacionais.

Expandindo esta abordagem, Milinković et al. [29] propuseram um modelo descritivo baseado em EDM, combinando técnicas de seleção de atributos com o algoritmo *Expectation-Maximization* (EM). Esta metodologia mostrou-se particularmente eficaz na previsão de desempenho acadêmico e na identificação de padrões de aprendizagem, estabelecendo um referencial importante para a análise de dados educacionais.

Priyambada et al. [34] contribuíram significativamente para este campo ao desenvolver um modelo de análise evolutiva de clusters. Sua abordagem permite identificar padrões temporais e analisar migrações entre diferentes comportamentos de aprendizado, oferecendo uma perspectiva dinâmica sobre as trajetórias educacionais. Esta metodologia é especialmente relevante para nossa pesquisa, pois possibilita a análise de transições entre diferentes perfis profissionais ao longo do tempo.

A integração destas abordagens metodológicas fornece um arcabouço robusto para nossa análise das trajetórias dos egressos da UFCG. As técnicas validadas por estes estudos permitem não apenas segmentar perfis acadêmicos, mas também compreender as transições profissionais e identificar padrões de carreira significativos, fundamentando nossa metodologia de análise.

3.1.3 Predição de Transição Profissionais com Inteligência Artificial

A previsão de transição profissionais através de técnicas de aprendizado de máquina tem se mostrado uma abordagem promissora na análise de carreiras em computação. Silva et al. [45] desenvolveram um modelo preditivo abrangente que integra diversos algoritmos, incluindo Random Forest, Gradient Boosting, K-Nearest Neighbors e Redes Neurais Artificiais. Seus resultados demonstram que a combinação de técnicas supervisionadas e não supervisionadas pode aprimorar significativamente a precisão na previsão de oportunidades profissionais para graduados em Computação.

Bhaskaran et al. [5] expandiram esta perspectiva ao investigar a empregabilidade de graduados através de técnicas avançadas de aprendizado de máquina. Seu estudo identificou fatores acadêmicos e extracurriculares determinantes para a inserção profissional, estabelecendo correlações significativas entre o desempenho durante a graduação e o sucesso na transição para o mundo do trabalho.

Pali e Tiwari [30] contribuíram significativamente para este campo ao desenvolver uma

análise abrangente do uso da inteligência artificial na previsão de carreiras. Sua pesquisa explorou estratégias de aprendizado preditivo e engenharia de atributos, avaliando o desempenho de diversos algoritmos como Support Vector Machines, Random Forest e Redes Neurais. Os resultados demonstram que a IA pode efetivamente auxiliar na tomada de decisões sobre transições profissionais, oferecendo direcionamentos valiosos para orientação de carreira.

Complementando esta linha de investigação, Suman et al. [47] desenvolveram um sistema preditivo baseado em aprendizado profundo, utilizando arquiteturas LSTM e GRU para estimar oportunidades de emprego no setor de Engenharia e Tecnologia. Seu modelo demonstrou capacidade superior na captura de padrões complexos e na previsão de possibilidades de empregabilidade, fornecendo uma ferramenta valiosa tanto para estudantes quanto para instituições acadêmicas na orientação de carreiras. Na tabela 3.1, pode-se ver a comparação de três estudos com técnicas diferentes, com propósitos similares.

Estudo	Técnica Principal	Foco da Análise	Principais Contribuições
Macêdo et al. [25]	K-Means	Padrões de transição	Segmentação de trajetórias
Santos et al. [42]	K-Means + PCA	Empregabilidade	Perfis acadêmicos
Lade et al. [24]	Web Scraping	Extração de dados	Sistema automatizado

Tabela 3.1: Comparação entre Abordagens de Análise de Trajetórias Profissionais

3.2 Empregabilidade de Graduados em Computação

A empregabilidade de graduados em Computação tem sido objeto de diversos estudos que analisam a relação entre formação acadêmica e sucesso profissional. Santos et al. [42] utilizaram a DEA para avaliar a eficiência educacional e sua correlação com a empregabilidade, demonstrando que programas acadêmicos que equilibram adequadamente teoria e prática apresentam indicadores superiores de inserção profissional.

Um estudo particularmente relevante foi conduzido por Gupta e Grover [49], que identificaram dois perfis distintos de graduados: os "*Practicable Students*", caracterizados por forte alinhamento com as demandas do mundo do trabalho, e os "*Impracticable Students*", que enfrentam dificuldades na inserção profissional devido a lacunas em sua formação. Esta

classificação evidencia a importância da flexibilização curricular e do alinhamento das disciplinas optativas com as necessidades do mundo do trabalho.

Bem et al. [4] forneceram uma perspectiva regional ao investigar o perfil de egressos da Universidade Federal de Roraima, revelando que a maioria atua em engenharia e desenvolvimento de software, com altos índices de empregabilidade e satisfação profissional. Este estudo oferece um importante referencial comparativo para nossa análise dos egressos da UFCG, permitindo contextualizar padrões regionais de empregabilidade.

Furtado et al. [16] aprofundaram a análise das dificuldades na empregabilidade de engenheiros de software, examinando fatores como formação acadêmica, experiência profissional e risco de obsolescência técnica. Seus resultados enfatizam que o sucesso profissional depende da combinação entre competências técnicas e interpessoais, destacando ainda a necessidade constante de atualização frente às rápidas mudanças tecnológicas.

Soares et al. [46] trouxeram uma dimensão adicional ao investigar como fatores psicossociais, especialmente a autoconfiança cognitiva e afetiva, influenciam as decisões de carreira. Seus achados demonstram que estudantes com maior confiança em suas habilidades acadêmicas e resiliência emocional tendem a fazer escolhas profissionais mais estáveis e bem-sucedidas.

Ainda no estudo de Furtado et al. [16], complementam esta análise ao identificar a adaptabilidade de carreira como um mediador essencial entre a formação acadêmica e a inserção profissional. Esse estudo ressalta a importância de programas educacionais que fomentem o desenvolvimento de habilidades adaptativas e aprendizado contínuo.

Enquanto a análise da empregabilidade dos graduados fornece uma perspectiva sobre a eficácia da formação acadêmica, é igualmente importante compreender como os empregadores avaliam estes profissionais. Esta perspectiva complementar, discutida a seguir, permite uma compreensão mais abrangente do processo de transição profissional.

3.3 Satisfação dos Empregadores com Graduados em TIC

A análise da perspectiva dos empregadores sobre os graduados em Tecnologia da Informação e Comunicação fornece percepções valiosas sobre a efetividade dos currículos acadêmicos. Fernandes et al. [15] investigaram como a adoção de inteligência artificial nos processos de

recrutamento tem possibilitado um alinhamento mais preciso entre as competências acadêmicas e as exigências do mundo do trabalho.

Williams et al. [13] expandiram esta análise ao examinar a influência do capital humano e da autogestão profissional na percepção de empregabilidade. Seus resultados evidenciam que as habilidades interpessoais e a capacidade de aprendizado contínuo são tão cruciais quanto o domínio técnico, estabelecendo uma base para a reformulação de programas acadêmicos. Esta perspectiva é complementada por Kamal et al. [22], que desenvolveram modelos de recomendação baseados em IA para otimizar o alinhamento entre formação acadêmica e oportunidades profissionais.

Falco [14] apresentou uma análise abrangente sobre o papel da orientação educacional na persistência de estudantes em carreiras STEM. Seu estudo demonstra que programas estruturados de aconselhamento acadêmico não apenas reduzem a evasão, mas também melhoram significativamente a adaptação profissional dos graduados. Estes achados enfatizam a necessidade de ações institucionais direcionadas à preparação dos egressos para as demandas específicas do setor de TIC.

3.4 Síntese e Tendências Emergentes

A revisão narrativa revela três tendências significativas na análise da transição profissional de egressos em Computação. Primeiro, a consolidação de técnicas de mineração de dados e aprendizado de máquina como ferramentas fundamentais para a análise de transições profissionais, permitindo a identificação de padrões e previsão de tendências com maior precisão.

Segundo, emerge uma clara necessidade de alinhamento entre formação acadêmica e demandas do mundo do trabalho, com ênfase em currículos flexíveis e adaptáveis às rápidas mudanças tecnológicas. Esta tendência ressalta a importância de estruturas curriculares que possam evoluir continuamente para atender às necessidades emergentes do setor.

Terceiro, observa-se uma crescente valorização de competências além das puramente técnicas, com especial atenção às habilidades interpessoais e à capacidade de aprendizado contínuo. Esta mudança de paradigma sugere a necessidade de uma formação mais holística em Computação, que prepare os profissionais não apenas para os desafios técnicos, mas também para as demandas sociais e organizacionais do mundo do trabalho contemporâneo.

3.5 Lacunas de Pesquisa e Nossa Contribuição

Apesar dos avanços significativos na área, identificamos algumas lacunas importantes na literatura:

1. Primeira Lacuna: A escassez de estudos sobre o impacto das disciplinas optativas na transição profissional representa uma limitação importante na literatura atual. Enquanto autores como Gupta e Grover [49] abordam perfis gerais de empregabilidade, falta uma análise específica de como as escolhas curriculares influenciam as trajetórias profissionais.
2. Segunda Lacuna: A investigação limitada sobre padrões regionais de empregabilidade em TIC no Brasil, especialmente no contexto do Nordeste, deixa uma lacuna significativa no entendimento das dinâmicas locais do mundo do trabalho. Embora Bem et al. [4] forneçam percepções sobre a região Norte, faltam estudos comparáveis para outras regiões.
3. Terceira Lacuna: A necessidade de análises quantitativas mais robustas sobre a relação entre perfil acadêmico e sucesso profissional permanece um desafio importante. Apesar dos avanços em técnicas de mineração de dados, como demonstrado por Silva, Cortez e Carlos [45], ainda há espaço para análises mais sofisticadas que integrem múltiplas fontes de dados e metodologias.

Esta pesquisa propõe uma abordagem inovadora ao cruzar dados acadêmicos e profissionais dos egressos da UFCG, aplicando técnicas avançadas de mineração de dados para identificar padrões de transição profissional. Além de suprir a lacuna sobre o impacto das disciplinas optativas na carreira, este estudo fornece um referencial para outras instituições que busquem otimizar seus currículos em alinhamento com as demandas do mundo do trabalho.

O próximo capítulo apresentará a metodologia detalhada para abordar estas questões, fundamentada nas técnicas e abordagens identificadas neste estudo.

Capítulo 4

Metodologia

Este capítulo apresenta a abordagem metodológica empregada na análise da transição profissional dos egressos do curso de Ciência da Computação da UFCG. A pesquisa utiliza métodos mistos, seguindo a perspectiva de Romero e Ventura [40], que defendem a combinação de análises quantitativas e qualitativas em estudos de mineração de dados educacionais. Os dados acadêmicos foram coletados do Sistema de Controle Acadêmico Online (SCAO/UFCG) e organizados em três estruturas complementares, conforme detalhado na seção 2.1.1. Este processo permitiu a categorização das disciplinas optativas em quatro áreas temáticas (DEV, IA, INFRA e GERAL) e a representação de cada egresso por um vetor tridimensional de atributos, após a remoção da categoria GERAL e das disciplinas sem registros de matrícula, otimizando os dados para a aplicação das técnicas de agrupamento, abordagem que, segundo Lade et al. [24], permite uma compreensão mais robusta das transições profissionais.

A metodologia implementada fundamenta-se nas técnicas estatísticas multivariadas propostas por Camacho et al. [7] para análise de grandes volumes de dados. O estudo emprega o algoritmo K-Means para clusterização, técnica que Baker e Yacef [44] validam como eficaz na segmentação de perfis acadêmicos e na predição de trajetórias profissionais. Esta abordagem é complementada pela Análise de Componentes Principais (PCA), seguindo recomendações de Jolliffe e Cadima [21] para redução de dimensionalidade e otimização da análise.

O processo metodológico estrutura-se em duas dimensões convergentes, conforme sugerido por Ikotun et al. [20]: análise da formação acadêmica e estudo da primeira ocupação profissional. Esta organização permite examinar separadamente os dados acadêmicos e

profissionais até sua integração através da clusterização, oferecendo uma visão holística da transição universidade-mundo do trabalho. Vê-se o diagrama representativo das etapas da metodologia na figura 4.1.

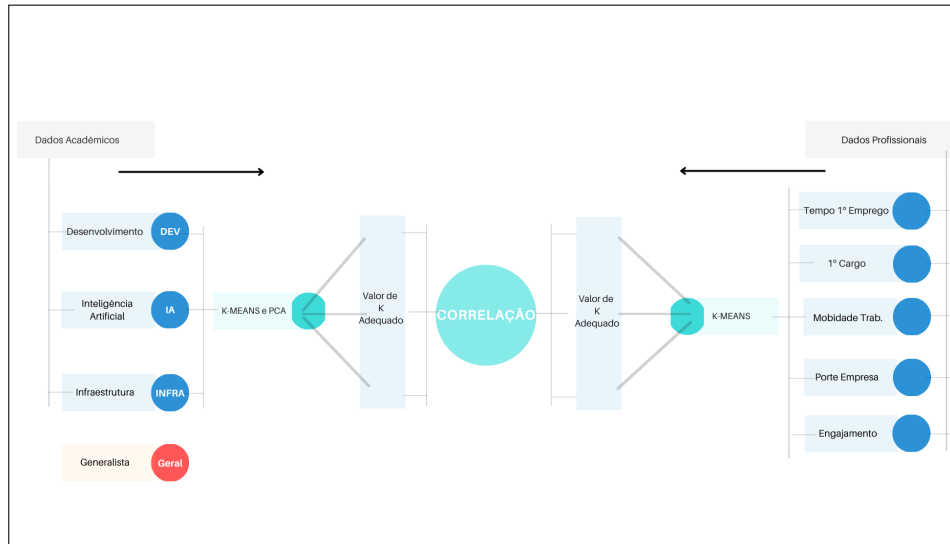


Figura 4.1: Diagrama ilustrativo do processo metodológico.

4.1 Ferramentas e Recursos

A implementação desta pesquisa fundamenta-se em uma infraestrutura tecnológica integrada, essencial para a análise robusta de dados educacionais. Conforme destacam Romero e Ventura [40], a escolha adequada de ferramentas é crucial para garantir a confiabilidade e reprodutibilidade dos resultados em mineração de dados educacionais. O processo de coleta de dados seguiu a abordagem multifonte proposta por Lade et al. [24], integrando três categorias principais:

4.1.1 Fontes de Dados

- **Sistema de Controle Acadêmico Online (SCAO/UFCG):** Forneceu os registros acadêmicos dos 352 egressos analisados, incluindo disciplinas cursadas, período de matrícula e conclusão. A escolha por analisar egressos a partir de 2012.2 fundamentou-se na implementação de mudanças curriculares significativas neste período, garantindo uma base comparativa mais consistente;

- **LinkedIn:** Plataforma empregada para validação das informações profissionais, método que, segundo Lade et al. [24], permite um acompanhamento confiável das transições no mundo do trabalho¹.
- **Enquetes Online:** Aplicação de questionários complementares, seguindo a metodologia de validação cruzada recomendada por Abdiodun et al. [20] para garantir a consistência dos dados profissionais.

4.1.2 Ferramentas de Análise

A escolha das ferramentas de análise baseou-se nas recomendações de Aggarwal [1] para processamento de grandes volumes de dados educacionais:

- **Python:** Ambiente principal de desenvolvimento, utilizado para implementação do algoritmo K-means e realização das análises estatísticas dos dados².
- **Visual Studio Code:** IDE escolhida para desenvolvimento dos scripts de análise e implementação dos algoritmos, oferecendo suporte robusto para Python e controle de versão³.

Esta estruturação metodológica garante não apenas a qualidade da análise, mas também sua replicabilidade, aspecto fundamental em pesquisas educacionais, conforme enfatizado por Siemens e Baker [44].

4.1.3 Plataformas de Colaboração e Documentação

- **Google Workspace⁴:**
 - Google Colab: Execução dos algoritmos em ambiente colaborativo;
 - Google Drive: Armazenamento e compartilhamento dos dados;
 - Google Docs/Sheets: Documentação e análise preliminar.

¹<https://www.linkedin.com/>

²<https://www.python.org/>

³<https://code.visualstudio.com/docs/languages/p>

⁴<https://workspace.google.com>

- **Latex⁵:**

- Word: Redação e formatação da dissertação.

- **GitHub:**

- Repositório para versionamento do código e disponibilização pública dos dados e scripts, garantindo reprodutibilidade da pesquisa⁶.

Esta infraestrutura tecnológica integrada permitiu uma análise abrangente e sistemática dos dados, facilitando a colaboração entre os pesquisadores e garantindo a transparência e reprodutibilidade do estudo. Todas as ferramentas foram escolhidas considerando sua adequação às necessidades específicas da pesquisa e sua capacidade de integração com os demais recursos utilizados.

4.1.4 Métricas Acadêmicas

A avaliação dos dados acadêmicos foi realizada por meio de métricas que permitem mensurar o engajamento dos egressos nas disciplinas optativas e identificar padrões de especialização. As principais métricas utilizadas foram:

- **Disciplinas Optativas:** Contagem e distribuição das disciplinas optativas cursadas pelos egressos nas áreas de Desenvolvimento de Software (DEV), Inteligência Artificial (IA) e Infraestrutura Computacional (INFRA);
- **Indicadores de Especialização:** Medição da concentração de disciplinas optativas em uma ou mais áreas, permitindo identificar egressos com perfis mais generalistas ou especializados. Essa métrica possibilita avaliar a proporção entre as áreas de conhecimento escolhidas pelos estudantes.

4.1.5 Métricas Profissionais

A análise da clusterização dos dados acadêmicos e profissionais foi realizada com base em métricas estatísticas que avaliam a qualidade e a separação dos grupos formados. Foram utilizadas as seguintes métricas:

⁵<https://www.overleaf.com/project/>

⁶https://github.com/felis-cc/dissertacao_ppgcc_ufcg.git

- **Tempo de Trabalho:** Intervalo entre a graduação e a primeira atividade profissional na área de Computação. Os intervalos foram definidos como:
 - Imediata (0-6 meses);
 - Curto prazo (7-12 meses);
 - Médio prazo (13-24 meses);
 - Longo prazo (acima de 24 meses).
- **Modalidade de Trabalho:** Classificação das formas de trabalho exercidas pelos egressos, segmentando-se em:
 - Remota;
 - Híbrida;
 - Presencial.
- **Cargo Ocupado:** Identificação do cargo exercido pelo egresso em sua primeira atuação profissional, permitindo avaliar padrões de inserção no mundo do trabalho e a distribuição de áreas de atuação, é vista claramente na figura 4.5 do gráfico de distribuição por cargos.
- **Porte Empresarial:** Segmentação das empresas empregadoras conforme o número de funcionários e faturamento estimado:
 - Microempresa (até 20 colaboradores);
 - Pequena empresa (até 100 colaboradores);
 - Média empresa (mais de 100 colaboradores e faturamento abaixo de R\$ 20 milhões);
 - Grande empresa (mais de 100 colaboradores e faturamento acima de R\$ 20 milhões).
- **Tipo de Engajamento:** Classificação do vínculo empregatício ou forma de atuação do egresso em sua primeira ocupação profissional, abrangendo categorias como:
 - Colaborador de empresa;

- Autônomo;
- Bolsista;
- Empreendedor.

Essas métricas possibilitam a análise comparativa entre diferentes perfis acadêmicos e profissionais, permitindo compreender como a trajetória acadêmica influencia a transição na carreira dos egressos.

A Figura 4.2 apresenta a matriz de correlação entre os grupos DEV, IA e INFRA, evidenciando padrões de associação entre as áreas de estudo. Após essa análise, passou-se a examinar o engajamento dos alunos em cada grupo de disciplinas.

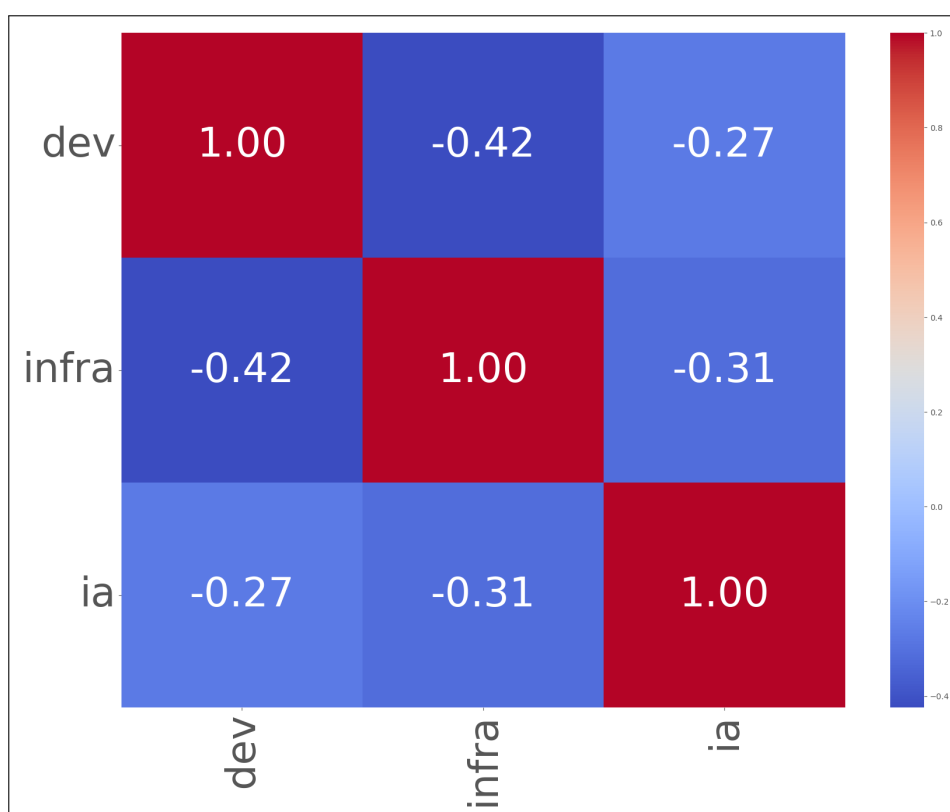


Figura 4.2: Gráfico de calor apresentando a correlação entre os três grupos de disciplinas.

A Figura 4.3 ilustra o total de frequências de matrículas dos egressos em cada grupo, reforçando o predomínio de DEV como principal área de formação.

O grupo de Desenvolvimento de Software registrou um total de 1379 matrículas, consolidando-se como o mais popular entre os egressos. A maioria dos estudantes optou

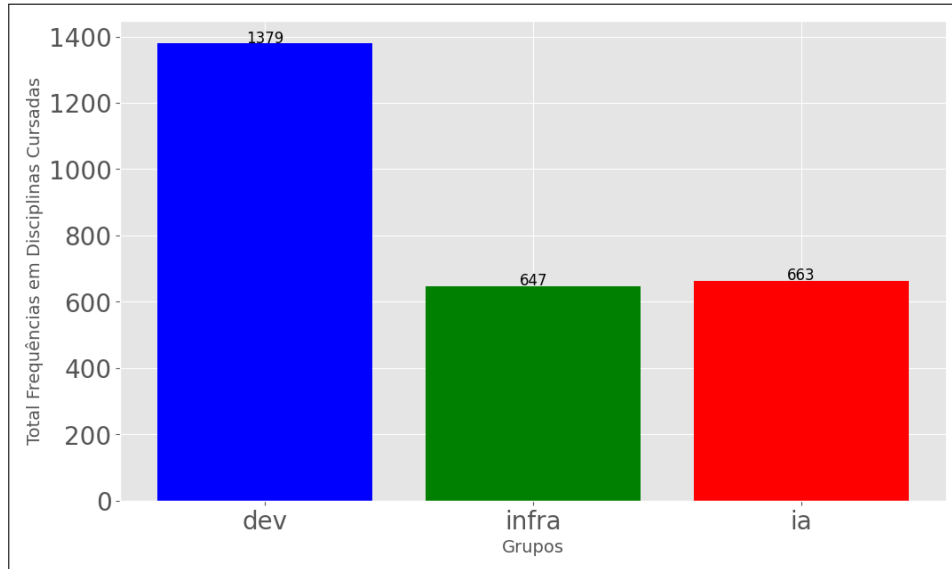


Figura 4.3: Quantidade de frequências de matrículas dos egressos nas disciplinas de cada grupo.

por cursar entre 4 e 5 disciplinas em DEV, enquanto um número reduzido chegou aos extremos de 1 ou 7 disciplinas. O grupo de Infraestrutura, por sua vez, apresentou 647 matrículas, com um padrão de engajamento mais superficial, predominando entre aqueles que cursaram apenas 1 ou 2 disciplinas. Já a área de Inteligência Artificial somou 663 matrículas, com leve superioridade em níveis básicos, evidenciando uma abordagem complementar em relação a DEV.

4.1.6 Tratamento dos Dados Profissionais

A investigação sobre os caminhos de carreira dos egressos foi conduzida por meio de uma abordagem iterativa, utilizando múltiplas fontes de dados: duas enquetes aplicadas em fases distintas da pesquisa e perfis profissionais do LinkedIn. Esta abordagem permitiu construir uma base sólida de conhecimento sobre os egressos.

Dos 352 egressos analisados, 263 possuíam currículos acessíveis em plataformas públicas, possibilitando o estudo da transição para o mundo do trabalho. O tratamento dos dados profissionais exigiu estratégias específicas devido à heterogeneidade das variáveis (numéricas, categóricas e ordinais). As informações foram organizadas de acordo com variáveis principais como tempo até a primeira ocupação, modalidade de trabalho e porte empresarial.

A limpeza e padronização envolveram a remoção de inconsistências, duplicações e o preenchimento de valores ausentes com base na média ou regressão, com posterior normalização para análise estatística.

A Figura 4.4 apresenta a distribuição do tempo para a primeira atividade profissional, modalidade de trabalho e porte da empresa empregadora. Observou-se que a maioria dos egressos conseguiu seu primeiro emprego na área em até 6 meses, predominando o trabalho híbrido, seguido pelo remoto, com concentração em empresas de médio e grande porte.

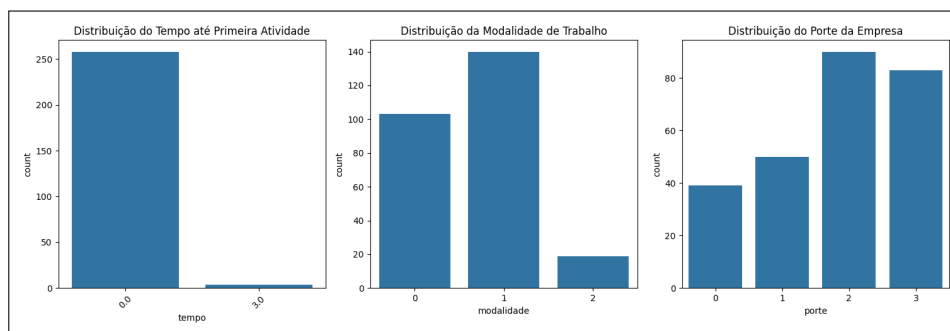


Figura 4.4: Gráfico de barra, distribuição por tempo, modalidade de trabalho e porte empresarial.

A Figura 4.5 ilustra a distribuição dos cargos ocupados pelos egressos, evidenciando que a área de Desenvolvimento e Engenharia de Software é a mais comum, seguida por Infraestrutura de TI e Ciência de Dados/IA. Outras áreas, como Análise e Suporte em TI, Segurança e Gerenciamento, apresentaram menor representatividade.

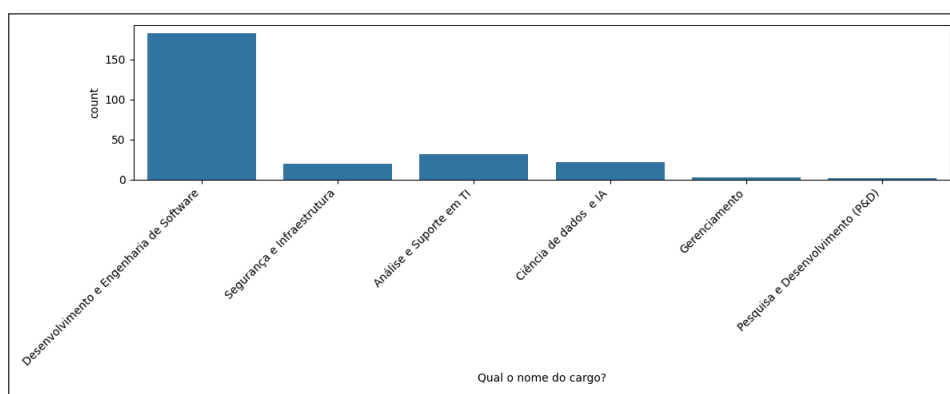


Figura 4.5: Distribuição por cargos.

Ao examinar com atenção a matriz de correlação na Figura 4.6, pode-se notar que a correlação entre trabalho autônomo (eng_Autônomo) e colaborador de empresa

(eng_Colaborador de empresa) é negativa, aproximadamente -0.81. Esta correlação fortemente negativa faz mais sentido do ponto de vista lógico, pois indica uma relação de exclusão mútua; quanto maior a tendência a trabalhar como autônomo, menor a probabilidade de atuar como colaborador formal de uma empresa, podendo assim refletir a realidade prática onde estas são geralmente modalidades de trabalho alternativas, não complementares.

Igualmente, observa-se também que a correlação entre bolsistas (eng_Bolsista) e colaboradores de empresa (eng_Colaborador de empresa) é de aproximadamente -0.57, também negativa, sugerindo outra relação de exclusão entre essas categorias de engajamento profissional. Outro padrão notável é a correlação negativa (-0.44) entre atuação em Desenvolvimento e Engenharia de Software e Segurança e Infraestrutura, corroborando a tendência de especialização em áreas distintas observada nos dados acadêmicos.

Igualmente, observa-se também que a correlação entre bolsistas (eng_Bolsista) e colaboradores de empresa (eng_Colaborador de empresa) é de aproximadamente -0.57, também negativa, sugerindo outra relação de exclusão entre essas categorias de engajamento profissional. Outro padrão notável é a correlação negativa (-0.44) entre atuação em Desenvolvimento e Engenharia de Software e Segurança e Infraestrutura, corroborando a tendência de especialização em áreas distintas observada nos dados acadêmicos.

A análise exploratória forneceu um embasamento sólido para a clusterização dos egressos, permitindo a segmentação em perfis distintos conforme as características profissionais observadas. Esses resultados direcionam as etapas subsequentes, possibilitando a aplicação de métodos estatísticos e de aprendizado de máquina para melhor compreensão da transição profissional dos egressos.

4.2 Clusterização com o Algoritmo K-Means

4.2.1 Descrição do Algoritmo

O algoritmo K-Means é um método de aprendizado não supervisionado que tem como objetivo minimizar a soma das distâncias quadráticas dentro dos clusters, agrupando os dados de maneira a maximizar a similaridade entre os elementos do mesmo grupo e minimizar a similaridade entre grupos distintos [11]. O critério de otimização do algoritmo é dado pela

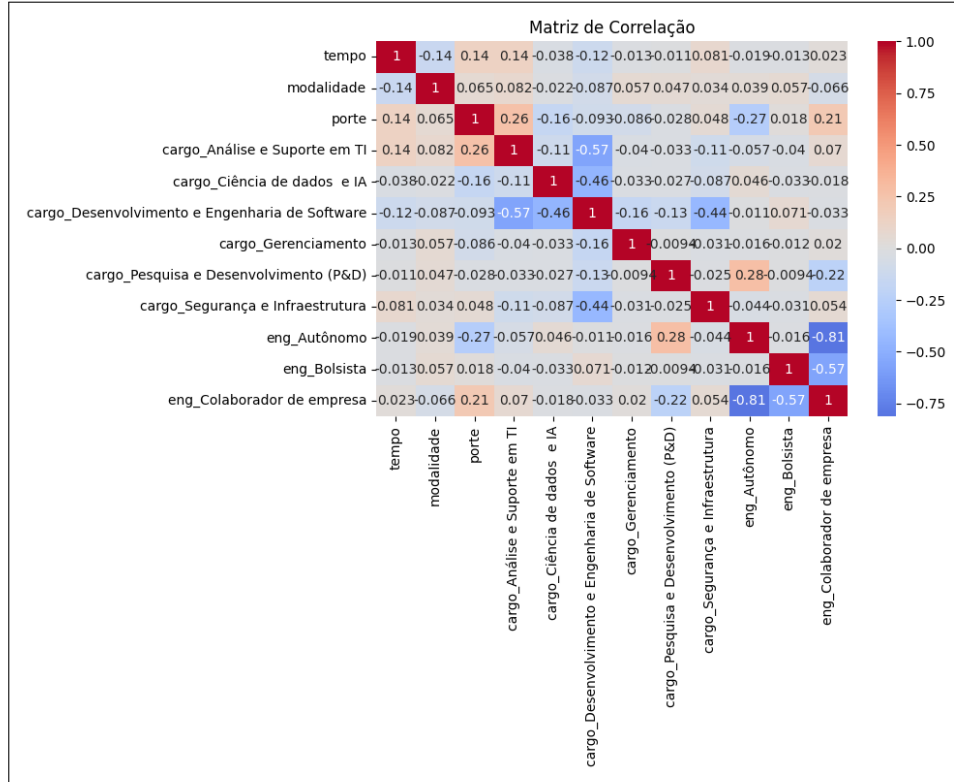


Figura 4.6: Gráfico da matriz de correlação das variáveis dos dados profissionais.

seguinte função objetivo:

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (4.1)$$

onde:

- J é a função objetivo do K-Means (inércia total);
- x representa os pontos de dados;
- C_i é o conjunto de pontos pertencentes ao cluster i ;
- μ_i representa o centróide do cluster i .

O algoritmo segue um processo iterativo que envolve os seguintes passos:

1. **Inicialização:** Os centróides são escolhidos aleatoriamente no espaço dos dados.

2. **Atribuição:** Cada ponto de dado x é atribuído ao cluster cujo centróide μ_i está mais próximo, baseado na métrica de distância euclidiana:

$$d(x, \mu_i) = \sqrt{\sum_{j=1}^n (x_j - \mu_{ij})^2} \quad (4.2)$$

3. **Cálculo dos novos centróides:** Para cada cluster i , recalcula-se μ_i como a média dos pontos atribuídos a ele:

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad (4.3)$$

4. **Convergência:** Os passos 2 e 3 são repetidos até que os centróides não mudem significativamente ou atinja um critério de parada predefinido.

4.2.2 Parâmetros de Implementação e Determinação do número de clusters K

A determinação do número ideal de clusters e a configuração dos parâmetros de implementação constituem etapas críticas na aplicação do algoritmo K-Means. A seleção destes valores foi guiada por métodos quantitativos complementares, cada um enfatizando diferentes aspectos da qualidade da segmentação.

Para os dados acadêmicos, aplicamos duas técnicas principais, o Método do Cotovelo e o Coeficiente de Silhueta. O Método do Cotovelo, analisando a variação da inércia em função do número de clusters, identificou uma redução significativa até $k=4$, seguida de uma estabilização gradual (Figura 4.7a). Este comportamento sugere que quatro clusters representam um ponto de equilíbrio entre minimização da variância intracluster e parcimônia do modelo. Já o Coeficiente de Silhueta, que avalia simultaneamente a coesão interna e a separação entre clusters, atingiu seu valor máximo (0,32) em $k=3$ (Figura 4.7b), com ligeira redução para $k=4$ (0,29), indicando maior sobreposição entre os grupos nesta configuração.

A divergência entre as recomendações destes métodos é comum na análise de clusters, refletindo diferentes prioridades, vejamos que o método do cotovelo enfatiza a compactação interna, enquanto o coeficiente de silhueta prioriza a separação intercluster. Considerando esta complementaridade, exploramos configurações com k variando entre 3 e 6, avaliando tanto métricas quantitativas quanto a interpretabilidade dos resultados.

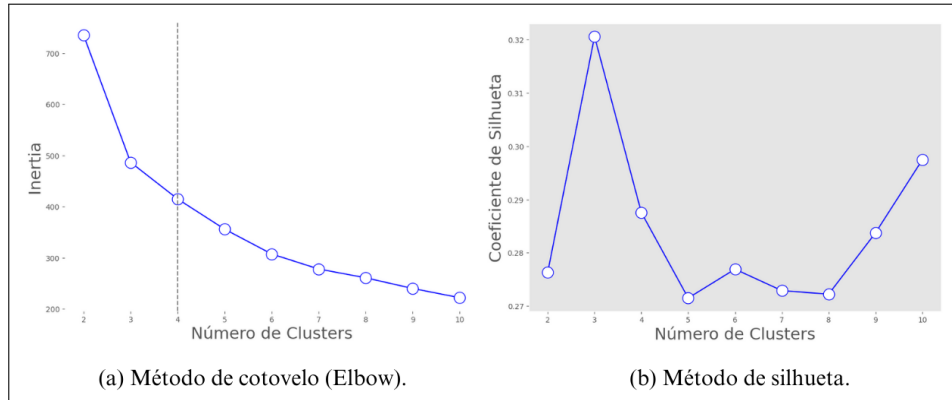


Figura 4.7: Métricas de qualidade dos clusters para os dados acadêmicos.

Para os dados profissionais, utilizamos a mesma abordagem metodológica, porém sem aplicação prévia de PCA. O método do cotovelo (Figura 4.8a) indicou uma inflexão entre $k=3$ e $k=4$, enquanto o coeficiente de silhueta (Figura 4.8b) apresentou seu máximo em $k=10$. Este cenário aparentemente contraditório é frequente em dados multidimensionais complexos. Considerando o tamanho da amostra disponível ($n=263$) e buscando equilíbrio entre precisão estatística e clareza interpretativa, selecionamos $k=3$ para os dados profissionais, priorizando grupos estatisticamente significativos e conceitualmente coerentes. Com base

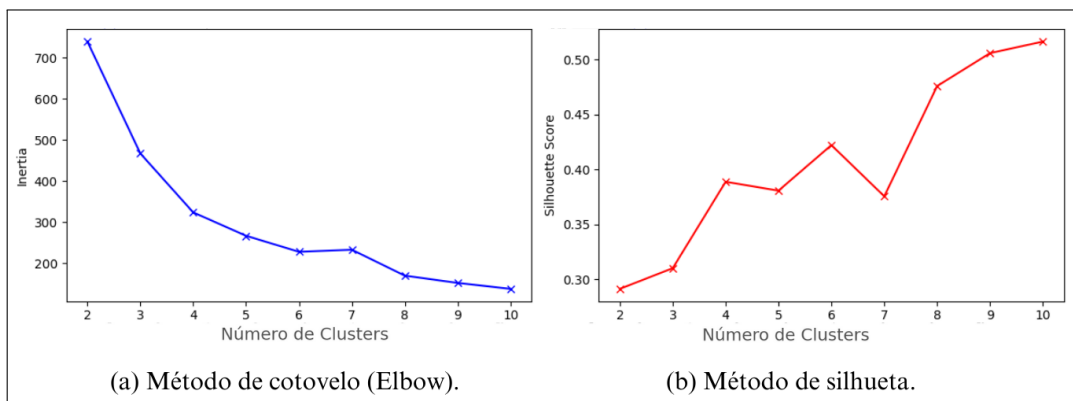


Figura 4.8: Métricas de qualidade dos clusters para os dados acadêmicos.

nestas análises, a implementação do algoritmo K-Means adotou os seguintes parâmetros:

- **Número de inicializações:** 10
- **Critério de convergência:** 0.0001
- **Número máximo de iterações:** 300

- **Método de inicialização dos centróides:** aleatória

Foi aplicada a PCA para reduzir a dimensionalidade dos dados acadêmicos, facilitando a visualização dos clusters em representações bidimensionais e tridimensionais. A normalização dos dados utilizou o *StandardScaler*, garantindo média 0 e desvio padrão 1 para todas as variáveis, evitando influências indevidas de escalas diferentes na clusterização. Como os centróides são obtidos no espaço normalizado, foi necessário transformá-los para o espaço PCA, permitindo sua visualização e análise interpretativa.

A seleção final do número de clusters integrou tanto os critérios técnicos apresentados quanto considerações qualitativas sobre a estrutura dos dados. Buscou-se configurações que proporcionassem agrupamentos não apenas estatisticamente válidos, mas também conceitualmente interpretáveis, explicativos e consistentes. Esta abordagem equilibrada priorizou a relevância prática dos agrupamentos identificados e sua capacidade de responder aos objetivos analíticos da pesquisa, reconhecendo que a segmentação ideal deve conciliar rigor matemático com significado contextual.

4.2.3 Análise de Componentes Principais (PCA)

A Análise de Componentes Principais foi aplicada exclusivamente aos dados acadêmicos com dois propósitos fundamentais: reduzir a dimensionalidade e facilitar a visualização dos agrupamentos. Esta transformação permitiu a projeção dos dados em um espaço bidimensional (PCA2D) e tridimensional (PCA3D), preservando a estrutura essencial da variabilidade dos dados.

O processo de agrupamento foi realizado sobre os dados transformados pelo PCA, enquanto os centróides, inicialmente calculados no espaço normalizado original, foram posteriormente projetados no espaço PCA para fins de visualização e interpretação. Esta abordagem permitiu representar graficamente os clusters sem comprometer a integridade da segmentação original.

Para os dados profissionais, optamos por não aplicar o PCA, realizando o agrupamento diretamente nos dados normalizados. Esta decisão fundamentou-se na quantidade reduzida de variáveis e na interpretabilidade satisfatória do espaço original, preservando todas as dimensões para uma caracterização ótima dos perfis profissionais.

4.2.4 Explicação Técnica dos Gráficos

Os gráficos apresentados na Figura 4.9 ilustram o poder do algoritmo K-means na identificação de padrões naturais nos dados. Como podemos observar, o painel da esquerda mostra os dados originais não agrupados, onde os pontos aparecem distribuídos sem uma classificação visível. Já no painel da direita, vemos o resultado após a aplicação do K-means, com os pontos claramente segmentados em três clusters distintos, representados pelas cores vermelho, verde e roxo. Esta visualização demonstra a principal função do K-means: transformar um

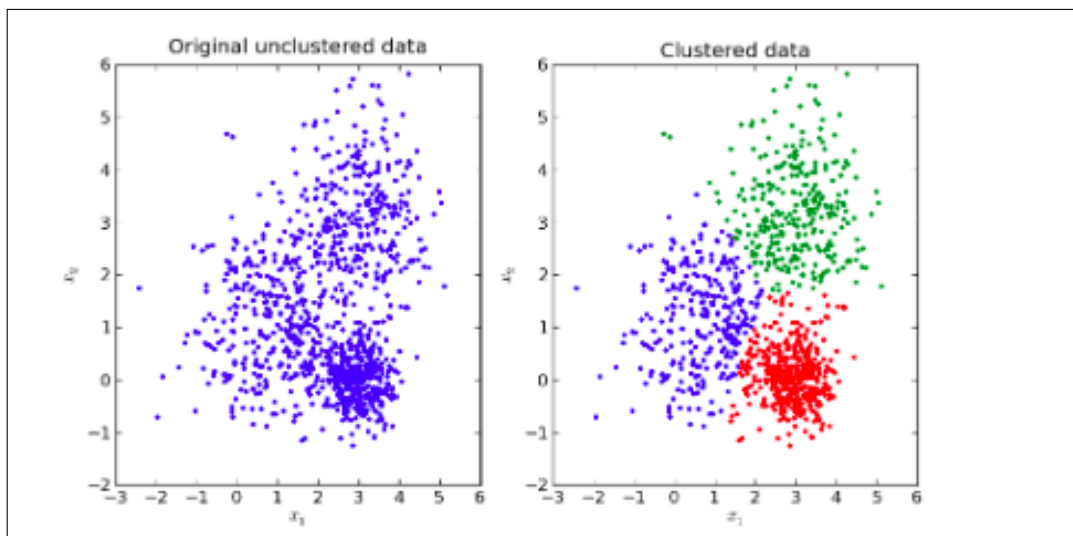


Figura 4.9: Comparação entre dados originais (esquerda) e após aplicação do K-means (direita). Imagem adaptada do google.

conjunto aparentemente homogêneo de dados em agrupamentos significativos baseados em similaridades intrínsecas. O algoritmo identificou três padrões principais nos dados:

O cluster vermelho (inferior direito): representa um grupo bem compacto e claramente separado, indicando alta similaridade entre seus elementos; O cluster roxo (centro-esquerdo): apresenta uma dispersão moderada, sugerindo maior variabilidade interna; O cluster verde (superior direito): exibe maior espalhamento, revelando um perfil mais heterogêneo.

Através desta segmentação, conseguimos transformar dados brutos em informações estruturadas que revelam padrões subjacentes, permitindo identificar grupos naturais de egressos com características acadêmicas semelhantes. Esta abordagem será fundamental para analisar como diferentes perfis de formação se relacionam com as trajetórias profissionais, fornecendo insights sobre a transição da universidade para o mundo do trabalho.

A análise detalhada destes clusters e sua interpretação no contexto das trajetórias acadêmicas e profissionais dos egressos serão apresentadas na seção de resultados.

Capítulo 5

Análise e Discussão dos Resultados

Este capítulo apresenta uma análise detalhada dos padrões identificados nas trajetórias acadêmicas e transições profissionais dos egressos do curso de Ciência da Computação da UFCG, utilizando técnicas de mineração de dados e aprendizado não supervisionado. A investigação segue o diagrama metodológico estabelecido, que integra duas análises convergentes através do algoritmo K-means, escolhido por sua capacidade robusta de identificar padrões em dados multidimensionais e estabelecer agrupamentos com características homogêneas.

Inicialmente, apresenta-se a definição dos perfis acadêmicos, explorando mediante análises estatísticas e representações gráficas os padrões nas escolhas das disciplinas optativas. A versatilidade do K-means, combinada com a redução de dimensionalidade via PCA, permitiu segmentar eficazmente os diferentes perfis formativos, revelando agrupamentos consistentes com características bem definidas.

Em seguida, foi conduzida a análise dos dados profissionais, com atenção especial ao tratamento das variáveis categóricas e numéricas que caracterizam as primeiras experiências dos egressos no mundo do trabalho. Aqui, o K-means foi aplicado sem a PCA, possibilitando a identificação de perfis profissionais distintos quanto ao tempo de inserção no mundo do trabalho, modalidade de trabalho e porte das organizações empregadoras.

O capítulo culmina com a convergência dessas duas análises, apresentando uma visão integrada e abrangente da relação entre formação acadêmica e transição profissional. Através de visualizações e análises estatísticas, demonstramos as correlações encontradas entre os perfis acadêmicos identificados e os padrões de inserção no mundo do trabalho, evidenciando como as escolhas curriculares influenciam as transições profissionais iniciais dos egressos.

5.1 Clusterização dos Dados Acadêmicos

A investigação das trajetórias acadêmicas dos egressos foi conduzida através de uma abordagem progressiva de clusterização, partindo de uma segmentação mais granular até alcançar um modelo sintético e robusto. Esta estratégia permitiu não apenas identificar os padrões fundamentais nas escolhas curriculares, mas também compreender as nuances e transições entre diferentes perfis de formação.

5.1.1 Análise da Configuração $k=6$

A análise inicial, utilizando $k=6$, revelou uma segmentação detalhada dos dados através da técnica de redução de dimensionalidade PCA. Como ilustrado na Figura 5.1a, a projeção bidimensional evidencia clusters bem definidos, cada um ocupando uma região única no espaço de características. No quadrante inferior esquerdo, observa-se um cluster roxo com alta compactação, indicando um perfil altamente especializado. Na região central-direita, o cluster vermelho apresenta uma distribuição mais ampla, enquanto o cluster verde, localizado na área central-superior, demonstra dispersão moderada. No lado esquerdo da projeção, o cluster azul exibe maior espalhamento vertical, e os clusters amarelo e laranja, situados na região central-esquerda, mostram padrões distintos de agrupamento. Esta configuração mais granular permitiu identificar como alguns clusters apresentavam maior compactação, indicando perfis altamente especializados, enquanto outros demonstravam maior dispersão, sugerindo abordagens mais generalistas ou de transição.

A visualização tridimensional desta configuração (Figura 5.1b) enriqueceu a análise ao revelar padrões não evidentes na projeção 2D. Particularmente notável foi a clara separação entre os clusters localizados na região central (especialmente visível nas projeções laterais "Left" e "Right"), reforçando sua singularidade, enquanto a sobreposição parcial entre os clusters das regiões periféricas sugeria áreas de transição entre perfis acadêmicos. Na visualização "Up" (inferior direita), é possível observar como os clusters do quadrante inferior esquerdo mantêm sua coesão mesmo na terceira dimensão, enquanto os clusters do quadrante superior direito apresentam maior interseção espacial. Esta perspectiva tridimensional revelou-se essencial para compreender a complexidade das relações entre os grupos e

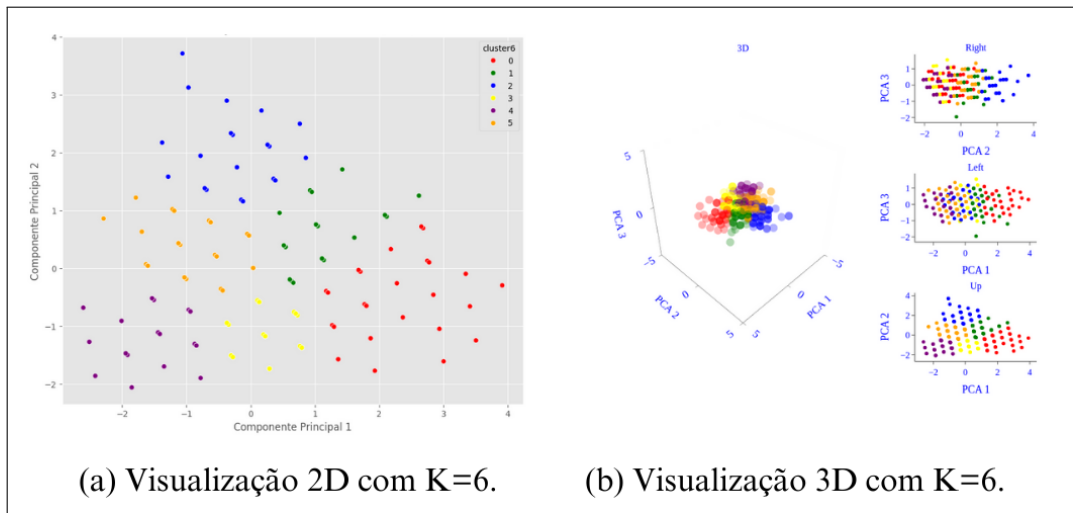


Figura 5.1: Distribuição dos clusters utilizando PCA no K-means.

suas características distintivas.

A análise das médias e medianas para $k=6$, apresentadas nas Figuras 5.2 e 5.3, revela perfis acadêmicos distintos e bem segmentados entre os egressos, destacando a diversidade nas escolhas curriculares realizadas.

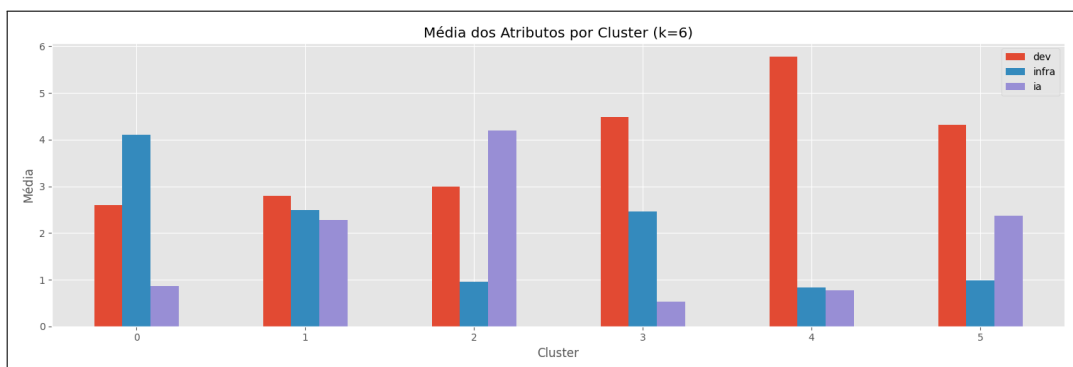


Figura 5.2: Gráfico de barras representando a média dos atributos para $k=6$.

A segmentação identifica seis perfis claramente diferenciados: o Cluster 0 destaca-se pela especialização em INFRA (média 4 disciplinas), com participação moderada em DEV e mínima em IA, sendo o único onde INFRA supera DEV. O Cluster 1 apresenta formação generalista, com distribuição equilibrada entre as três áreas, ligeiramente favorecendo DEV. O Cluster 2 evidencia orientação para IA (média 4 disciplinas), complementada por base em DEV e presença reduzida em INFRA, caracterizando um perfil especializado em tecnologias cognitivas.

O Cluster 3 enfatiza DEV (4,5 disciplinas), com participação secundária em INFRA e mínima em IA, configurando especialização em engenharia de software. O Cluster 4 manifesta a mais alta concentração em DEV (média 6 disciplinas), com participação mínima nas demais áreas, constituindo o perfil mais focado em programação. O Cluster 5 integra forte orientação para DEV (média 4) com participação moderada em IA, sugerindo combinação estratégica entre desenvolvimento de software e aplicações de inteligência artificial.

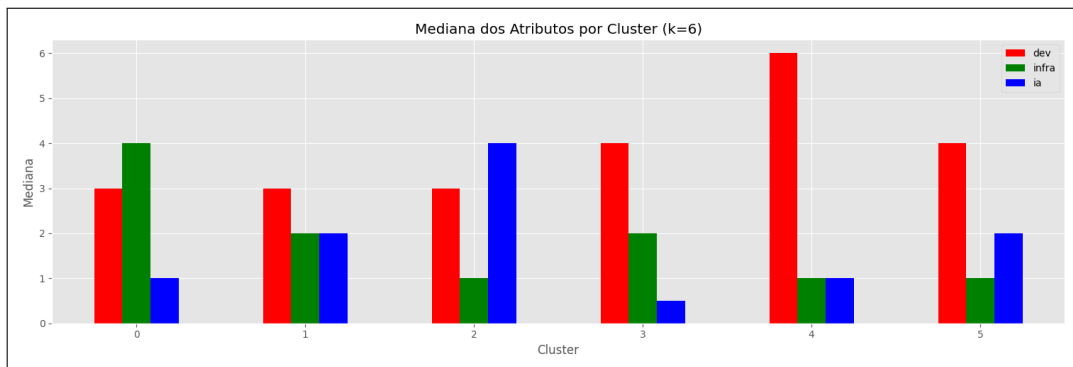


Figura 5.3: Gráfico de barras representando a mediana dos atributos para k=6.

A análise das medianas corrobora estes padrões, confirmando os perfis identificados: especialização em INFRA (Cluster 0, mediana INFRA=4); formação generalista (Cluster 1, medianas equilibradas); foco em IA (Cluster 2, mediana IA=4); ênfase em DEV com suporte em INFRA (Cluster 3); especialização intensiva em DEV (Cluster 4, mediana DEV=6); e integração entre DEV e IA (Cluster 5).

A notável proximidade entre médias e medianas em todos os clusters indica distribuições simétricas com baixa influência de valores extremos, validando a consistência dos agrupamentos e a robustez da segmentação realizada.

5.1.2 Análise da Configuração k=5

A transição para k=5 proporcionou uma consolidação inicial dos padrões, mantendo a capacidade de identificar especializações importantes, mas com menor granularidade. Como demonstrado na Figura 5.4, esta configuração apresentou uma separação clara entre os clusters, com alguns grupos mantendo maior compactação enquanto outros exibiam sobreposições sugestivas de perfis híbridos. Esta consolidação permitiu uma interpretação mais direta dos resultados, facilitando sua aplicação em contextos práticos de gestão acadêmica.

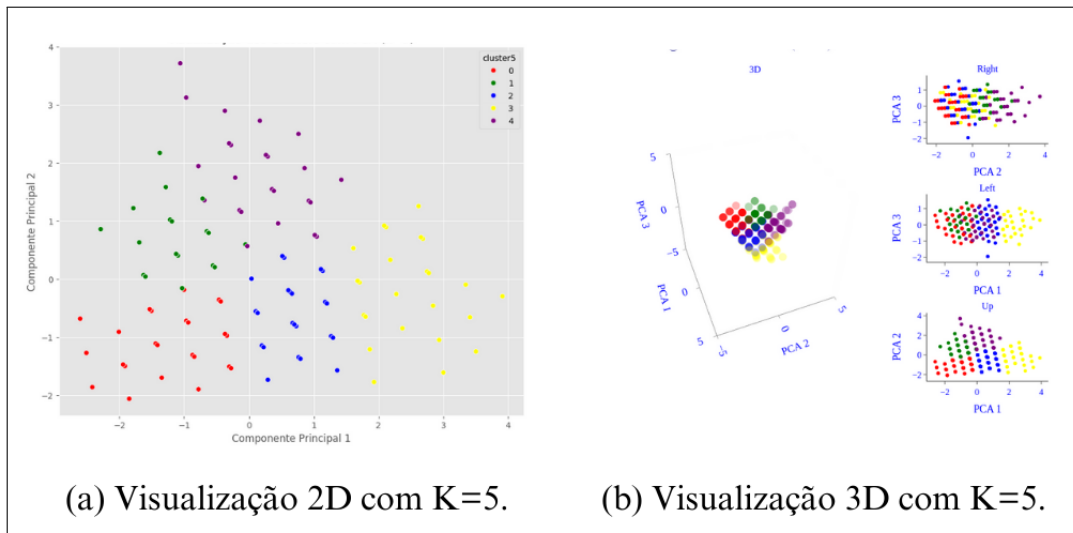


Figura 5.4: Distribuição dos clusters utilizando PCA no K-means.

A análise das médias e medianas para $k=5$ (Figuras 5.5 e 5.6) revelou padrões consistentes de especialização. O Cluster 0 demonstrou foco predominante em desenvolvimento, com baixa representatividade em outras áreas. O Cluster 1 apresentou uma combinação interessante de desenvolvimento com participação em IA, sugerindo um perfil mais adaptativo. O Cluster 2 evidenciou especialização em infraestrutura, enquanto o Cluster 3 revelou-se mais generalista. O Cluster 4 destacou-se por sua forte orientação para IA, com desenvolvimento como área secundária.

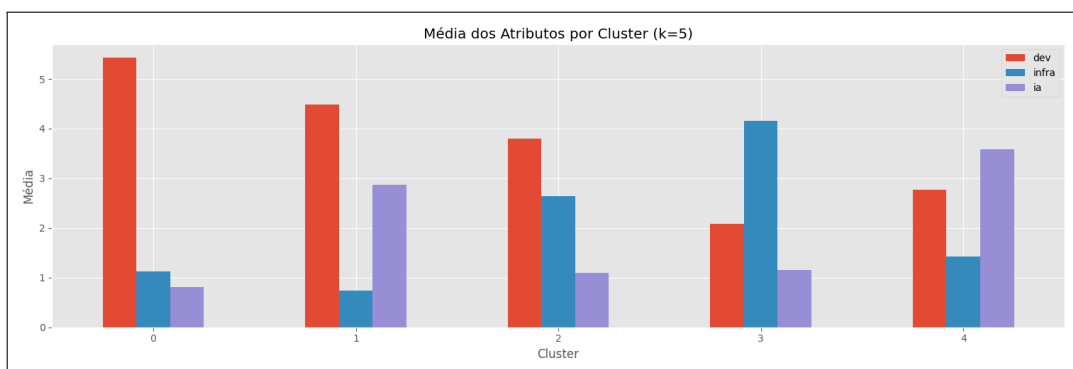


Figura 5.5: Gráfico de barras representando a média dos atributos para $k=5$.

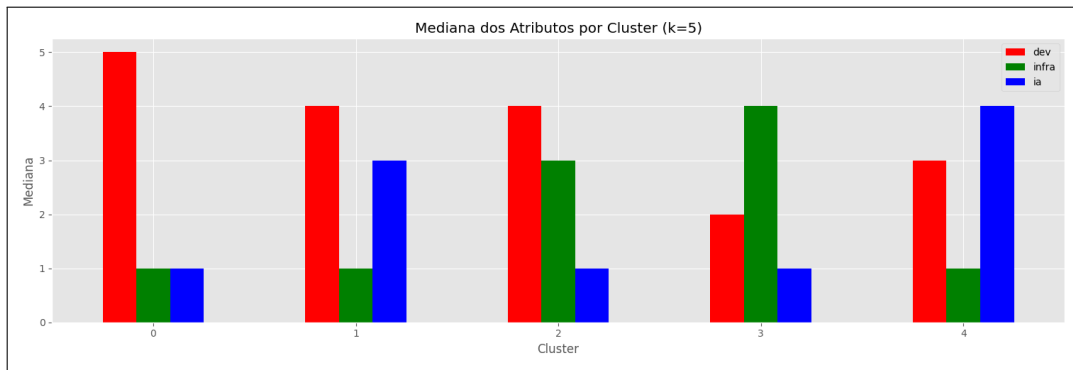


Figura 5.6: Gráfico de barras representando a mediana dos atributos para k=5.

5.1.3 Análise da Configuração k=4

A configuração com k=4 (Figura 5.7) oferece uma simplificação que mantém a coerência dos padrões principais identificados. Nesta configuração, quatro perfis distintos emergem com clara separação espacial no mapa de componentes principais.

Na visualização 2D (Figura 5.7a), observa-se um grupo concentrado no quadrante inferior esquerdo, caracterizado por especialização em DEV. Na região superior do gráfico, distribuído predominantemente no eixo vertical, localiza-se um perfil mais diversificado e adaptável. No quadrante inferior central, encontra-se um grupo focado em INFRA, enquanto no quadrante inferior direito situa-se um cluster orientado para IA, que se estende parcialmente para a região superior direita. A visualização tridimensional (Figura 5.7b) confirma a

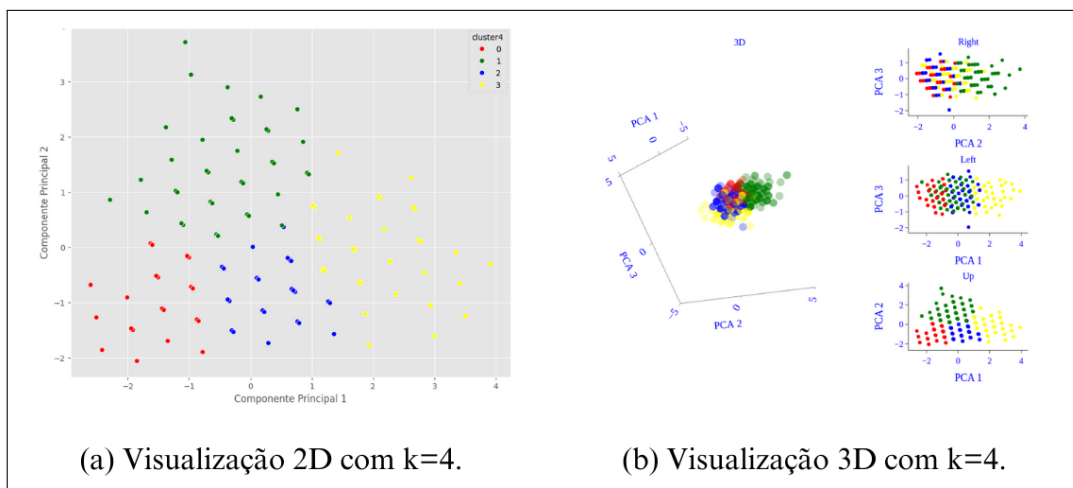


Figura 5.7: Distribuição dos clusters utilizando PCA no K-means.

robustez desta segmentação, oferecendo perspectivas complementares. Na projeção central

3D, nota-se a distribuição espacial dos quatro grupos com fronteiras definidas. A visualização "Right"(superior direita) revela a separação no plano PCA2-PCA3, enquanto a projeção "Left"(meio direita) demonstra a distribuição no plano PCA1-PCA3, evidenciando como o grupo do quadrante inferior esquerdo mantém-se distinto dos demais.

Particularmente reveladora é a projeção "Up"(inferior direita), que mostra a distribuição no plano PCA1-PCA2, permitindo observar como o grupo do quadrante inferior direito mantém clara separação, enquanto existe alguma zona de transição entre os grupos central e esquerdo. Esta configuração preserva as informações essenciais sobre as diferenças entre os principais perfis acadêmicos, enquanto simplifica a interpretação ao reduzir o número de segmentos.

A análise das médias e medianas para $k=4$, representadas nas Figuras 5.8 e 5.9, revela quatro perfis profissionais distintos entre os egressos. O Cluster 0 caracteriza-se pela inserção imediata no mercado e preferência pela modalidade híbrida em empresas estruturadas, enquanto o Cluster 1, também com rápida inserção, distingue-se pela preferência ao trabalho remoto e maior flexibilidade quanto ao porte empresarial. O Cluster 2 apresenta um padrão único, com inserção tardia (superior a 24 meses), modalidade exclusivamente remota e atuação em grandes empresas. Já o Cluster 3 representa um grupo intermediário, com inserção gradual e equilíbrio entre modalidades híbrida e presencial. A proximidade entre médias e

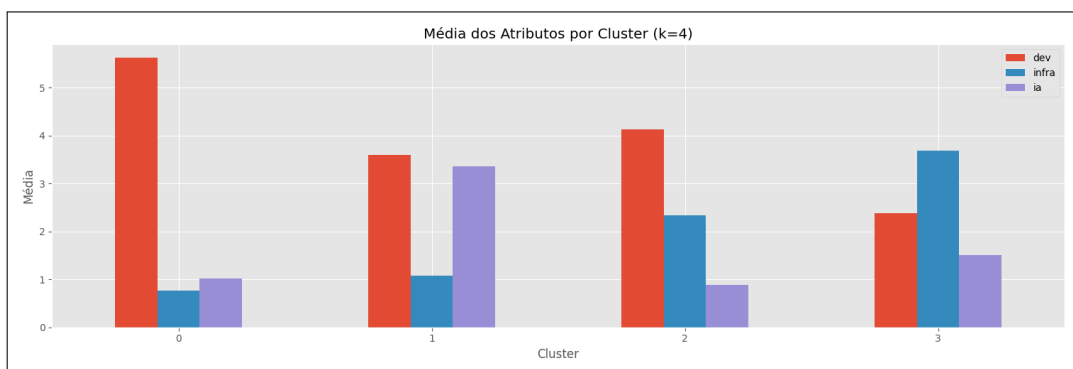


Figura 5.8: Gráfico de barras representando a média dos atributos para $k=4$.

medianas em todos os clusters indica baixa influência de outliers, validando a robustez da segmentação em quatro grupos e revelando padrões consistentes nas transições profissionais dos egressos. Esta análise quantitativa demonstra a diversidade de caminhos profissionais possíveis, desde a rápida inserção em ambientes corporativos até especializações que de-

mandam maior tempo de preparação.

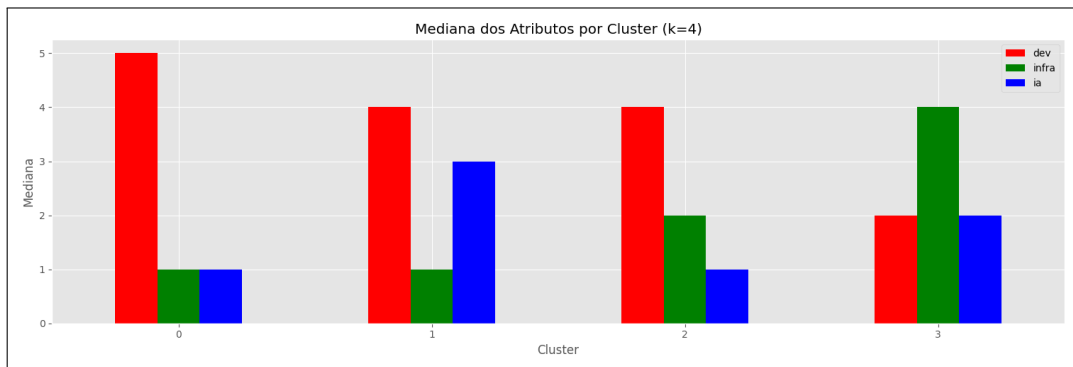


Figura 5.9: Gráfico de barras representando a mediana dos atributos para k=4.

A progressão da análise através destas diferentes configurações revelou como os perfis acadêmicos se organizavam em níveis crescentes de generalização, preparando o terreno para a identificação dos padrões fundamentais que emergiram na configuração final com k=3.

5.1.4 Análise da Configuração k=3

A configuração com três clusters emergiu como a mais apropriada para representar os padrões essenciais nas trajetórias acadêmicas dos egressos. Esta seção apresenta os resultados através de três perspectivas complementares: a distribuição espacial dos clusters, a caracterização estatística dos grupos e a validação da segmentação.

Distribuição Espacial dos Clusters

A análise da distribuição espacial através da visualização bidimensional (Figura 5.10a) revela uma segmentação clara e funcional dos perfis acadêmicos. Os três agrupamentos ocupam regiões distintas no espaço de componentes principais:

- O primeiro cluster, localizado na região inferior esquerda do espaço PCA, apresenta uma distribuição notavelmente compacta, indicando alta homogeneidade nas escolhas curriculares dos egressos deste grupo.
- O segundo cluster, situado predominantemente na região direita do gráfico, demonstra maior dispersão espacial, sugerindo maior variabilidade nas combinações de disciplinas.

- O terceiro cluster ocupa uma posição intermediária na região central-superior, com dispersão moderada que indica um equilíbrio entre especialização e diversificação curricular.

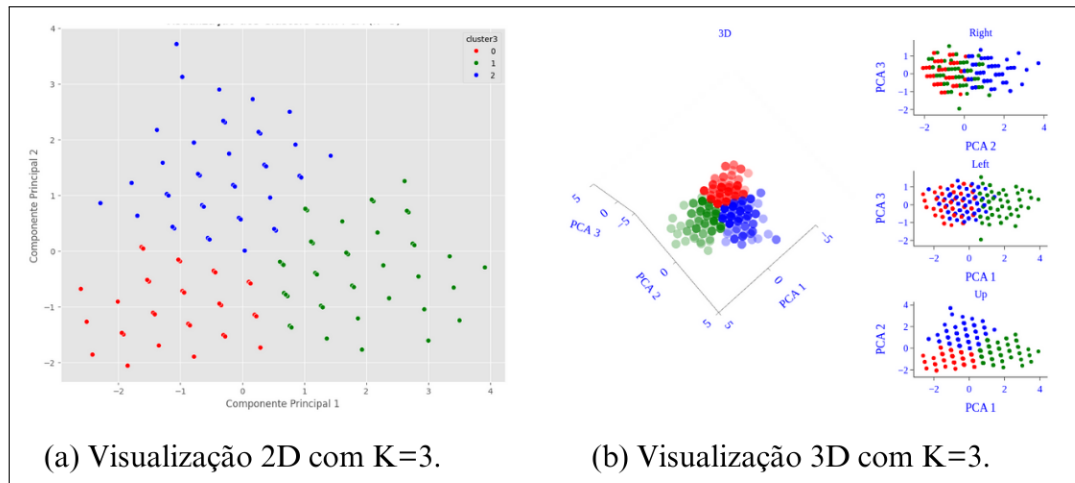


Figura 5.10: Distribuição dos clusters utilizando PCA no K-means.

A visualização tridimensional (Figura 5.10b) enriquece esta análise ao incorporar o terceiro componente principal (PCA3). Esta perspectiva evidencia como os agrupamentos mantêm suas características distintivas em maior dimensionalidade, com o cluster da região inferior esquerda preservando sua compactação característica, o cluster da região direita apresentando volume espacial mais expressivo, e o cluster da região central-superior mantendo uma distribuição moderada com clara separação dos demais grupos.

As projeções bidimensionais integradas oferecem perspectivas complementares sobre a separação entre os grupos:

- A projeção "Right"(PCA2 x PCA3) mostra boa separação geral, com sobreposição parcial do cluster da região direita, indicando áreas de conhecimento compartilhadas;
- A projeção "Left"(PCA1 x PCA3) evidencia distinção clara entre os clusters das regiões inferior esquerda e central-superior;
- A projeção "Up"(PCA1 x PCA2) confirma os padrões observados na análise bidimensional inicial, com três regiões bem delimitadas.

Esta análise multidimensional confirma a robustez da segmentação em três clusters, revelando agrupamentos consistentes que representam os principais perfis de formação dos egressos.

Caracterização Estatística dos Grupos

A análise das médias (Figura 5.11) e medianas (Figura 5.12) revela perfis acadêmicos distintos e consistentes, permitindo caracterizar cada cluster com precisão. O primeiro Cluster,

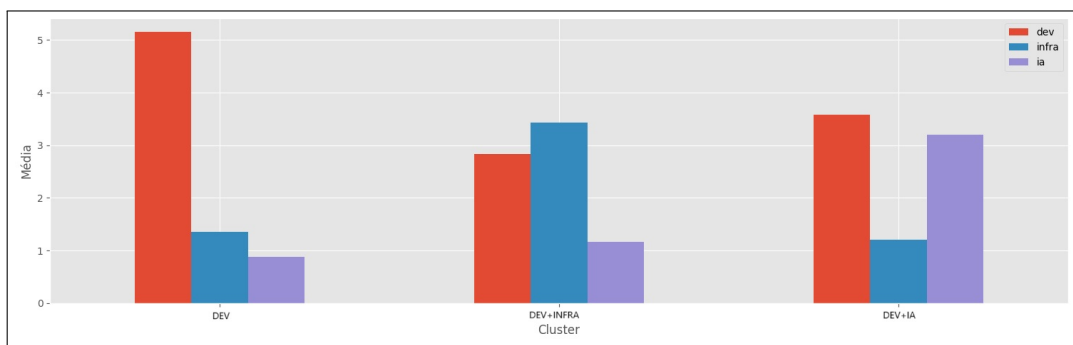


Figura 5.11: Gráfico de barras representando a média dos atributos para k=3.

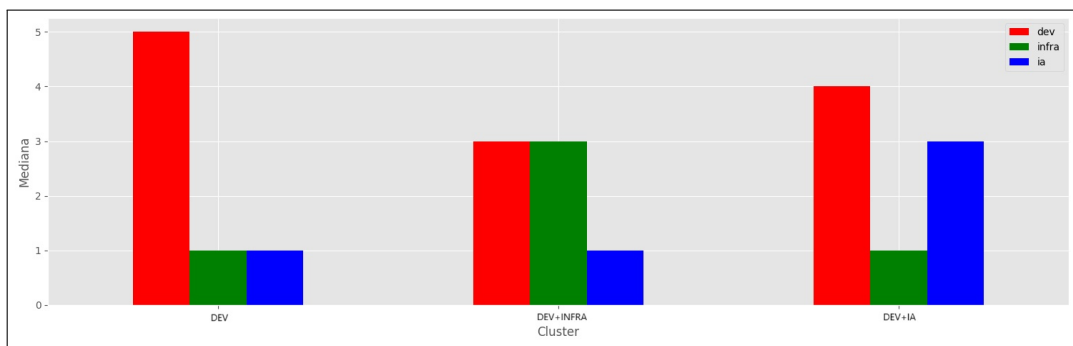


Figura 5.12: Gráfico de barras representando a mediana dos atributos para k=3.

contando da esquerda para a direita, apresenta forte especialização em DEV, com média aproximada de 5,1 e mediana de 5 disciplinas nesta área, os valores mais elevados entre todos os clusters. Em contraste, observa-se participação reduzida em INFRA (média 1,3, mediana 1) e IA (média 0,9, mediana 1), configurando um perfil focado em competências de programação e construção de software.

No segundo Cluster, o gráfico demonstra uma orientação predominante para INFRA, com média de 3,4 e mediana de 3 disciplinas nesta área, combinada com participação equilibrada

em DEV (média 2,8, mediana 3). A presença moderada de disciplinas de IA (média 1,1, mediana 1) complementa este perfil, caracterizando uma formação voltada para sustentação e operação de sistemas.

Porém no terceiro Cluster, este distingue-se pela ênfase em IA, apresentando média de 3,2 e mediana de 3 disciplinas nesta área, associada a uma base sólida em DEV (média 3,5, mediana 4) e participação reduzida em INFRA (média 1,1, mediana 1). Este agrupamento representa um perfil orientado para tecnologias analíticas e sistemas inteligentes.

A notável proximidade entre os valores de média e mediana em todos os clusters indica distribuições simétricas com baixa influência de valores extremos, confirmando a homogeneidade interna dos grupos e a consistência da segmentação realizada. Esta caracterização estatística valida a existência de três perfis acadêmicos predominantes entre os egressos, cada um refletindo diferentes ênfases e combinações de competências técnicas.

Validação da Segmentação

A robustez da segmentação é corroborada pela proximidade entre os valores de média e mediana em todos os clusters, indicando distribuições simétricas com baixa influência de *outliers*. Esta consistência estatística, combinada com a clara separação espacial observada nas visualizações, valida a adequação da configuração com três clusters para representar os principais perfis de formação dos egressos.

A análise revela três perfis acadêmicos distintos e bem definidos:

1. **Perfil Desenvolvimento:** Caracterizado por forte especialização em disciplinas de programação e desenvolvimento de software;
2. **Perfil Infraestrutura:** Marcado por equilíbrio entre diferentes áreas, com ênfase em infraestrutura computacional;
3. **Perfil IA:** Distinguido pela concentração em disciplinas de inteligência artificial, com base complementar em desenvolvimento.

Esta configuração demonstra-se especialmente adequada para análises que exigem uma visão generalista e coesa dos padrões de formação, mantendo suficiente poder discriminativo para caracterizar as principais tendências nas escolhas curriculares dos egressos.

Análise Comparativa através de Boxplots

A visualização dos perfis acadêmicos através de boxplots (Figura 5.12) revela uma clara distinção entre os três clusters identificados, evidenciando padrões específicos nas escolhas curriculares dos egressos.

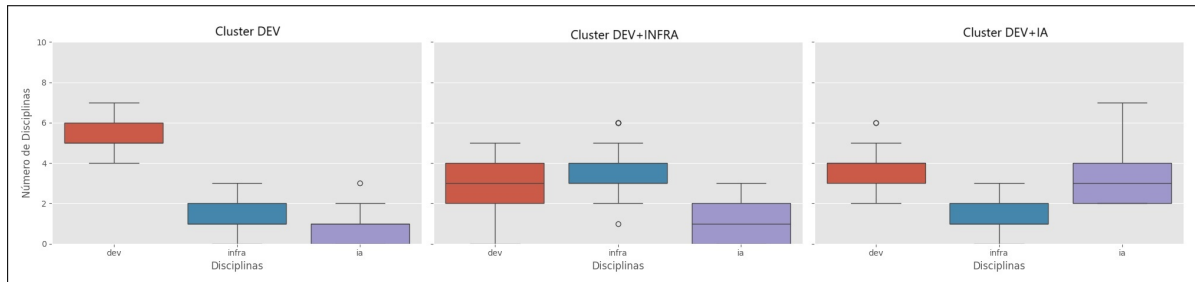


Figura 5.13: Gráfico de boxplots para k=3 dos dados acadêmicos.

O Cluster 0 caracteriza-se por forte concentração em desenvolvimento, apresentando mediana elevada e baixa dispersão interquartil nesta área, enquanto mantém participação mínima em infraestrutura e IA. Este agrupamento representa egressos com formação técnica centrada em desenvolvimento de software, porém com limitada diversificação curricular.

O Cluster 1 exibe uma distribuição mais equilibrada entre infraestrutura e desenvolvimento, com amplitude interquartil mais elevada, indicando maior variabilidade nas escolhas de disciplinas. Este grupo corresponde a um perfil híbrido, com conhecimentos técnicos distribuídos entre múltiplas áreas da computação.

O Cluster 2 distingue-se pela predominância em IA, com mediana elevada nesta área e participação reduzida nas demais. A concentração de disciplinas neste campo específico sugere uma especialização intencional em tecnologias e métodos de inteligência artificial.

Esta configuração com três clusters captura os padrões fundamentais de formação, oferecendo uma visão sintética e interpretável dos principais perfis acadêmicos dos egressos do curso.

Análise dos Centróides

A análise dos centróides fornece uma validação quantitativa dos padrões observados nas visualizações anteriores, oferecendo uma caracterização precisa dos perfis identificados tanto no espaço normalizado quanto após a transformação PCA.

A análise dos centróides no espaço normalizado (Tabela 5.1) confirma quantitativamente a caracterização dos três perfis acadêmicos. O Cluster DEV concentra-se fortemente em DEV (0,949), mantendo correlações negativas com INFRA (-0,369) e IA (-0,706), consolidando um perfil orientado à programação e engenharia de software. O Cluster DEV+INFRA demonstra clara especialização em INFRA (1,139), com correlações negativas para DEV (-0,822) e IA (-0,462), caracterizando um perfil focado em aspectos de sustentação tecnológica. O Cluster DEV+IA apresenta forte direcionamento para IA (0,969), com correlações negativas para DEV (-0,235) e INFRA (-0,513), evidenciando especialização em tecnologias cognitivas e analíticas.

Esta distribuição dos centróides valida a segmentação obtida, demonstrando a existência de três perfis acadêmicos bem definidos e estatisticamente distintos entre os egressos analisados. A transformação para o espaço PCA (Tabela 5.2) reforça a robustez desta segmentação

Cluster	Desenvolvimento	Infraestrutura Computacional	Inteligência Artificial
DEV	0,949	-0,369	-0,706
DEV+INFRA	-0,822	1,139	-0,462
DEV+IA	-0,235	-0,513	0,969

Tabela 5.1: Centróides dos clusters no espaço normalizado.

através de uma perspectiva complementar. O Cluster DEV+INFRA destaca-se no primeiro componente principal ($PCA1 = 1.448641$), indicando máxima diferenciação em relação aos demais grupos. O Cluster DEV+IA apresenta valor expressivo no segundo componente ($PCA2 = 1.062169$), confirmando sua singularidade, enquanto o Cluster DEV assume posição contrastante nos dois primeiros componentes ($PCA1 = 0.801333$, $PCA2 = -0.945507$). A distribuição dos centróides no espaço PCA demonstra que os dois primeiros componentes principais capturam efetivamente as variações mais significativas nos dados, permitindo uma caracterização adequada dos perfis de formação. O terceiro componente ($PCA3$), com valores próximos a zero para todos os clusters, indica que as principais diferenças entre os grupos são adequadamente representadas nas duas primeiras dimensões. Esta análise quantitativa corrobora as observações qualitativas anteriores, confirmando que a segmentação em três clusters captura efetivamente os principais padrões nas escolhas curriculares dos egressos.

Cluster	PCA1	PCA2	PCA3
DEV	0.801333	-0.945507	-0.033088
DEV+INFRA	1.448641	-0.301165	-0.030213
DEV+IA	-0.358633	1.062169	0.051705

Tabela 5.2: Centróides dos clusters no espaço PCA.

sos. A clara separação entre os grupos, evidenciada tanto no espaço normalizado quanto após a transformação PCA, valida a robustez desta configuração para representar os perfis fundamentais de formação.

5.1.5 Conclusão da Análise dos Dados Acadêmicos

A análise sistemática das escolhas curriculares dos egressos do curso de Ciência da Computação da UFCG, através de diferentes configurações de clusterização, revelou padrões significativos na formação acadêmica. A investigação progressiva, partindo de uma segmentação mais granular ($k=6$) até uma mais sintética ($k=3$), permitiu identificar tanto tendências gerais quanto nuances específicas nas trajetórias formativas.

A configuração com $k=3$ emergiu como a mais adequada para representar os padrões fundamentais de formação, oferecendo um equilíbrio ideal entre simplicidade interpretativa e poder explicativo. Esta configuração revelou três perfis acadêmicos distintos:

- **DEV+IA:** Caracteriza-se pela combinação de uma base sólida em desenvolvimento de software com especialização significativa em inteligência artificial. Este perfil reflete uma orientação para tecnologias emergentes e aplicações avançadas, evidenciada pela alta correlação com disciplinas de IA (0.969690) no espaço normalizado.
- **DEV+INFRA:** Demonstra equilíbrio entre desenvolvimento e infraestrutura computacional, com ênfase em aspectos fundamentais da computação. A análise dos centróides confirma esta característica dual, com valores positivos tanto em desenvolvimento quanto em infraestrutura.

- **DEV:** Apresenta foco predominante em desenvolvimento de software, com distribuição equilibrada entre outras áreas. Este perfil mais generalista é validado pela análise dos boxplots, que demonstra maior mediana em desenvolvimento (0.949174) com participação moderada nas demais áreas.

As configurações mais granulares ($k=4$, $k=5$ e $k=6$), embora ofereçam percepções valiosas sobre subgrupos específicos, apresentam maior complexidade interpretativa. Estas segmentações mais detalhadas são particularmente úteis para análises que demandam compreensão aprofundada de perfis específicos ou variações sutis nas escolhas curriculares.

A análise revela que o desenvolvimento de software constitui o núcleo comum da formação, presente em diferentes intensidades em todos os perfis identificados. Esta centralidade é complementada por duas vertentes principais de especialização: uma orientada à inteligência artificial e outra à infraestrutura computacional. Esta estrutura sugere que o currículo atual oferece flexibilidade suficiente para acomodar diferentes interesses e orientações profissionais, mantendo uma base técnica sólida em desenvolvimento de software.

5.2 Clusterização dos Dados Profissionais

A análise de agrupamento dos dados profissionais dos egressos foi conduzida através do mesmo algoritmo K-means, método escolhido por sua capacidade robusta de identificar padrões em dados multidimensionais e estabelecer grupos com características homogêneas. Esta escolha metodológica se fundamenta na versatilidade do algoritmo em processar tanto variáveis numéricas quanto categóricas transformadas, possibilitando uma segmentação eficaz dos diferentes perfis profissionais presentes na amostra.

5.2.1 Definição e Validação dos Clusters

Para determinar a configuração ideal dos agrupamentos, aplicou-se uma abordagem metodológica combinando os métodos Elbow e Silhueta, que convergiram para uma solução com três clusters ($k=3$). Esta configuração demonstrou-se adequada para capturar a variabilidade dos dados sem incorrer em sobreajuste, considerando também uma perspectiva mais abrangente da natureza dos dados analisados.

5.2.2 Caracterização dos Perfis Profissionais

A análise detalhada revelou três perfis profissionais distintos, cada um refletindo diferentes transições e escolhas de carreira dos egressos. Esta segmentação pode ser visualizada através de diferentes perspectivas nas Figuras 5.13, 5.14 e 5.15, que ilustram respectivamente a distribuição por cargo, modalidade de trabalho e porte empresarial.

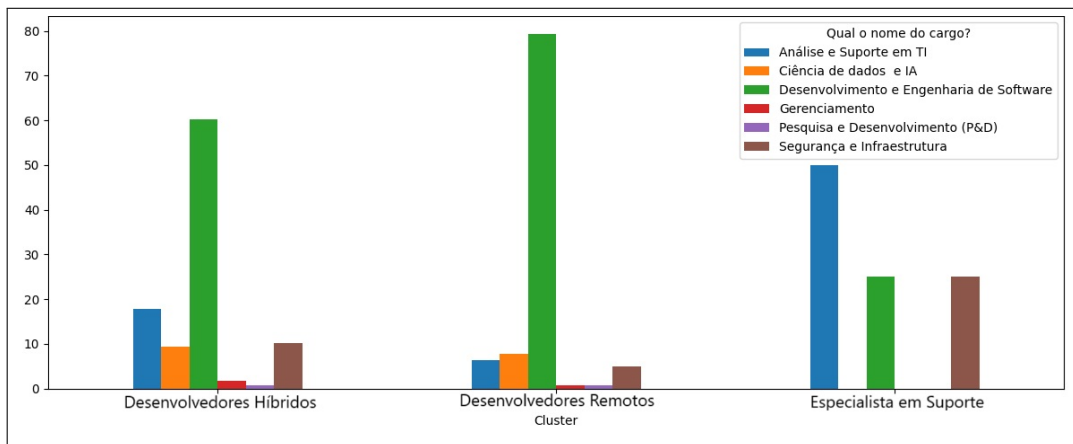


Figura 5.14: Distribuição de primeiros cargos de trabalho por cluster para $k=3$.

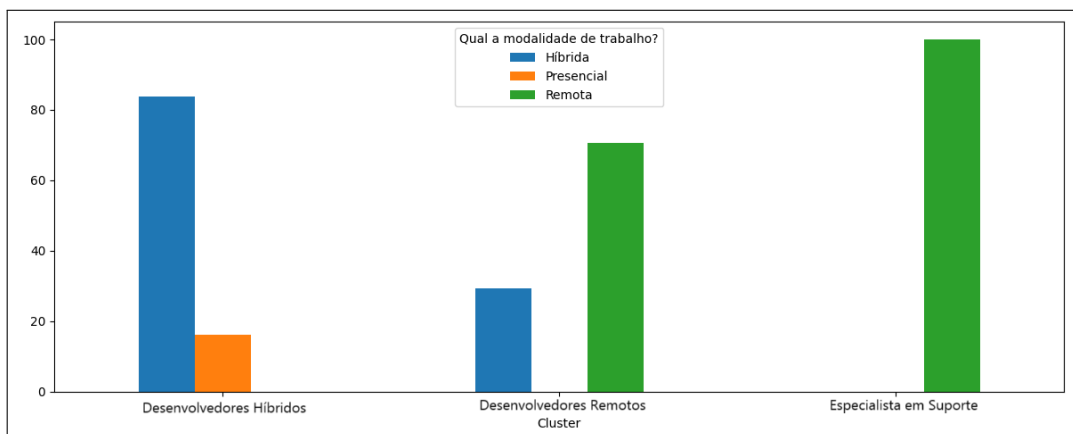


Figura 5.15: Distribuição da modalidade de trabalho por cluster para $K=3$.

O grupo mais representativo, denominado Desenvolvedores Híbridos (Cluster 0), constitui 45% da amostra e caracteriza-se por uma inserção praticamente imediata no mundo do trabalho, com clara preferência pela modalidade híbrida (83,90%) e alguma presença no modelo presencial (16,10%). Estes profissionais concentram-se predominantemente em empresas de médio e grande porte, com uma distribuição equilibrada de 44,92% em cada

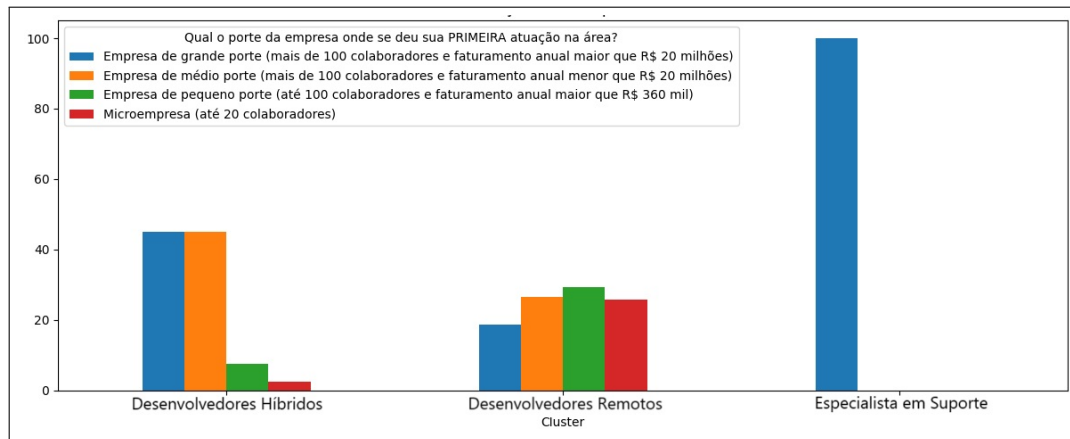


Figura 5.16: Distribuição de porte empresarial por cluster para $k=3$.

categoria, tendo menor presença em empresas de pequeno porte (7,63%) e microempresas (2,54%). No âmbito da atuação profissional, o Desenvolvimento e Engenharia de Software predomina (60,17%), seguido por Análise e Suporte em TI (17,80%), Segurança e Infraestrutura (10,17%), Ciência de Dados e IA (9,32%), com pequena participação em áreas de Gerenciamento (1,69%) e P&D (0,85%).

O segundo perfil identificado, dos Desenvolvedores Remotos (Cluster 1), representa 53,4% da amostra e compartilha a característica de rápida inserção no mercado, porém com forte preferência pelo trabalho remoto (70,71%), mantendo ainda uma presença significativa no modelo híbrido (29,29%). Este grupo apresenta a distribuição mais heterogênea quanto ao porte das organizações, com presença equilibrada entre empresas de pequeno porte (29,29%), médio porte (26,43%), microempresas (25,71%) e empresas de grande porte (18,57%). Na esfera de atuação, mantém-se a predominância do Desenvolvimento e Engenharia de Software (79,29%), complementada por Ciência de Dados (7,86%), Análise e Suporte em TI (6,43%), Segurança e Infraestrutura (5,00%), e pequena participação em Gerenciamento e P&D (0,71% cada).

Em contraste, o cluster minoritário, Especialista em Suporte (Cluster 2), representando 1,6% da amostra, destaca-se por características singulares em relação aos demais. Este perfil apresenta um tempo de inserção no mercado significativamente maior (superior a 24 meses) e caracteriza-se pela exclusividade tanto na modalidade de trabalho (100% remoto) quanto no porte empresarial (100% em empresas de grande porte). A distribuição das áreas de atuação mostra-se mais equilibrada, dividindo-se entre Análise e Suporte em TI (50,00%), Segurança

e Infraestrutura (25,00%) e Desenvolvimento de Software (25,00%).

Um aspecto comum a todos os clusters é o predomínio do vínculo profissional como colaboradores de empresa, indicando uma preferência generalizada por relações de trabalho formais e estáveis.

5.2.3 Análise Quantitativa dos Clusters

A análise das medidas de tendência central e dispersão, ilustradas através de gráficos de barras na Figura 5.16, reforça as distinções significativas entre os perfis identificados. Os dados evidenciam uma clara dicotomia no padrão de inserção profissional: enquanto os Desenvolvedores Híbridos e Remotos apresentam medianas indicando inserção imediata no mercado (mediana=0), os Especialistas em Suporte demonstram um período de transição consideravelmente mais longo (mediana=3), com médias que refletem o mesmo padrão de diferenciação.

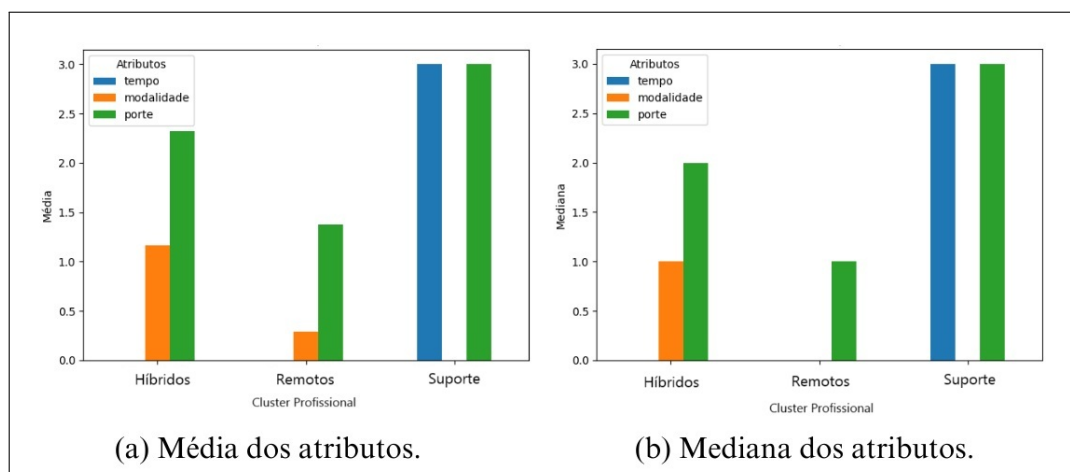


Figura 5.17: Gráfico dos atributos por clusters profissionais.

No que tange ao porte empresarial, observa-se uma variabilidade mais acentuada nos dois primeiros perfis. O cluster de Desenvolvedores Híbridos demonstra uma concentração equilibrada entre empresas de médio e grande porte, enquanto os Desenvolvedores Remotos apresentam a distribuição mais heterogênea, com representatividade significativa em todos os portes empresariais. Em contrapartida, o grupo dos Especialistas mantém uma homogeneidade consistente, com atuação exclusiva em empresas de grande porte.

A visualização através dos boxplots, apresentada na Figura 5.17, corrobora a robustez da

segmentação obtida, evidenciando separações claras principalmente nas variáveis de tempo de inserção e modalidade de trabalho. Esta distinção é particularmente notável nas modalidades de trabalho, onde cada cluster apresenta um padrão característico: predominância híbrida no primeiro grupo, preferência pelo remoto no segundo, e exclusividade remota no terceiro. Embora sejam observados alguns casos atípicos (outliers), principalmente no atributo de porte empresarial, estes não comprometem a coesão geral dos agrupamentos identificados.

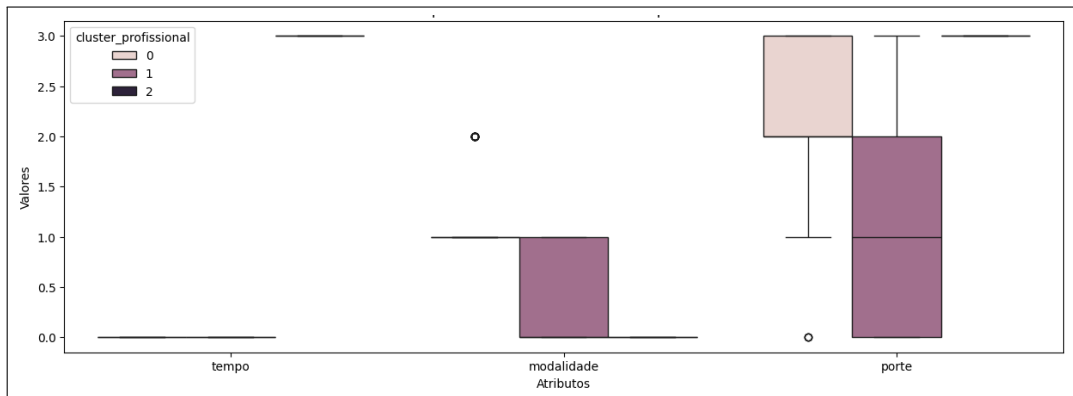


Figura 5.18: Gráfico de boxplots dos clusters profissionais para $k = 3$.

Esta caracterização quantitativa não apenas valida a metodologia empregada, mas também oferece percepções valiosas sobre os diferentes caminhos de carreira seguidos pelos egressos do curso de Ciência da Computação. A clara diferenciação entre os perfis, sustentada por múltiplas métricas e visualizações, demonstra a diversidade de transições profissionais possíveis dentro da área, desde a rápida inserção em ambientes híbridos até especializações mais focadas em modalidades remotas. A análise também revela uma tendência significativa de adaptação às diferentes culturas organizacionais, evidenciada pela distribuição variada dos profissionais em empresas de diferentes portes, especialmente nos dois clusters majoritários.

5.3 Discussão dos Resultados

A análise sistemática da transição dos egressos do curso de Ciência da Computação da UFCG para o mundo do trabalho revelou padrões complexos na relação entre formação acadêmica e atuação profissional. O estudo, fundamentado em dados de 352 egressos desde o período letivo 2012.2, dos quais 263 apresentavam informações profissionais acessíveis, permitiu

identificar correlações entre escolhas curriculares e caminhos profissionais.

5.3.1 Análise da Correlação dos Dados Profissionais com os Dados Acadêmicos dos Egressos

A investigação combinou dados acadêmicos (categorias DEV, IA e INFRA) com perfis profissionais, aplicando clusterização ($k=3$) para identificar três grupos distintos: DEV+IA, DEV+INFRA e DEV. Os resultados (Tabela 5.3) revelam uma distribuição relativamente homogênea entre os clusters profissionais, com algumas tendências específicas: o Cluster DEV+IA apresenta maior concentração (41.51%) no perfil de suporte, o Cluster DEV demonstra forte presença (50%) no perfil remoto, e o Cluster DEV+INFRA mantém distribuição mais equilibrada, com predominância no perfil híbrido (42.42%).

Clusters	Desenvolvedores Híbridos	Desenvolvedores Remotos	Especialista em Suporte
DEV+IA	27,67%	30,82%	41,51%
DEV+INFRA	42,42%	35,35%	22,23%
DEV	25,00%	50,00%	25,00%

Tabela 5.3: Distribuição dos clusters acadêmicos dentro de cada cluster profissional.

QP1 – Relação entre Perfil Acadêmico e Modalidade de Inserção Profissional

A primeira questão de pesquisa investiga como as escolhas de disciplinas optativas se relacionam com os padrões de inserção profissional inicial.

Os dados acadêmicos revelaram predominância das disciplinas de DEV (1.379 matrículas), em comparação com INFRA (647) e IA (663). A análise de correlação entre áreas identificou correlações negativas moderadas (-0.31 a -0.42), sugerindo que os alunos tendem a desenvolver especializações específicas. Cada perfil acadêmico mostrou tendências distintas:

- **DEV+IA:** Maior tendência para posições em desenvolvimento de software (60%) e Ciência de Dados (9.32%);
- **DEV+INFRA:** Combinando desenvolvimento com infraestrutura computacional (10.17%);

- **DEV:** Concentração predominante em desenvolvimento de software(79%).

Paralelamente, a análise das transições profissionais revelou três perfis distintos de atuação:

1. **Desenvolvedores Híbridos (45%):** Caracterizados por inserção imediata no mercado e preferência pela modalidade híbrida (83.9%), predominantemente em empresas de médio e grande porte (89.84% combinados).
2. **Desenvolvedores Remotos (53.4%):** Também com inserção rápida, mas com forte preferência pelo trabalho remoto (70.71%) e distribuição mais heterogênea quanto ao porte empresarial.
3. **Especialistas em Suporte (1.6%):** Grupo distintivo com inserção mais tardia, modalidade exclusivamente remota e atuação em grandes empresas.

A Figura 5.19 apresenta a distribuição dos clusters profissionais dentro de cada cluster acadêmico, permitindo visualizar como os diferentes perfis formativos se distribuem nas categorias profissionais. Esta análise revela dois padrões significativos:

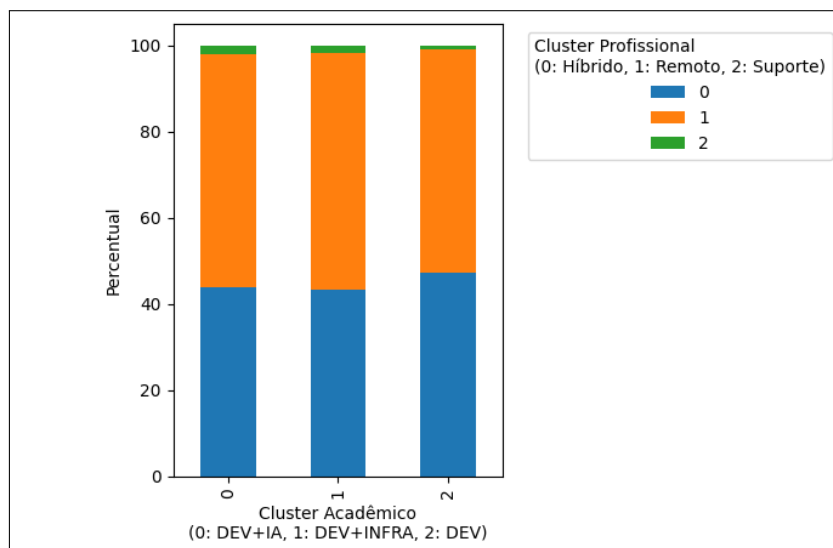


Figura 5.19: Gráfico de barras empilhada de distribuição dos clusters profissionais por cluster acadêmico.

Primeiro, observa-se equilíbrio na distribuição dos perfis profissionais dentro de cada grupo acadêmico, sugerindo que o perfil acadêmico, embora influencie a área específica de atuação, não determina o padrão geral de inserção profissional.

Segundo, a modalidade de trabalho apresenta distribuição consistente entre todos os perfis acadêmicos: aproximadamente 45% optam pelo trabalho híbrido, 53% pelo remoto e 2% pelo suporte técnico. Esta homogeneidade indica que fatores externos ao currículo exercem maior influência na modalidade de trabalho.

Quanto ao tempo de inserção, a maioria dos egressos apresenta inserção rápida (0-6 meses), independente do perfil acadêmico, com exceção do cluster de Suporte (superior a 24 meses). características da inserção profissional.

QP2 – Combinações Curriculares e Tendências de Atuação Profissional

A segunda questão investiga quais combinações de disciplinas estão associadas a diferentes transições profissionais. O mapa de calor (Figura 5.20) ilustra a intersecção entre clusters acadêmicos e profissionais:

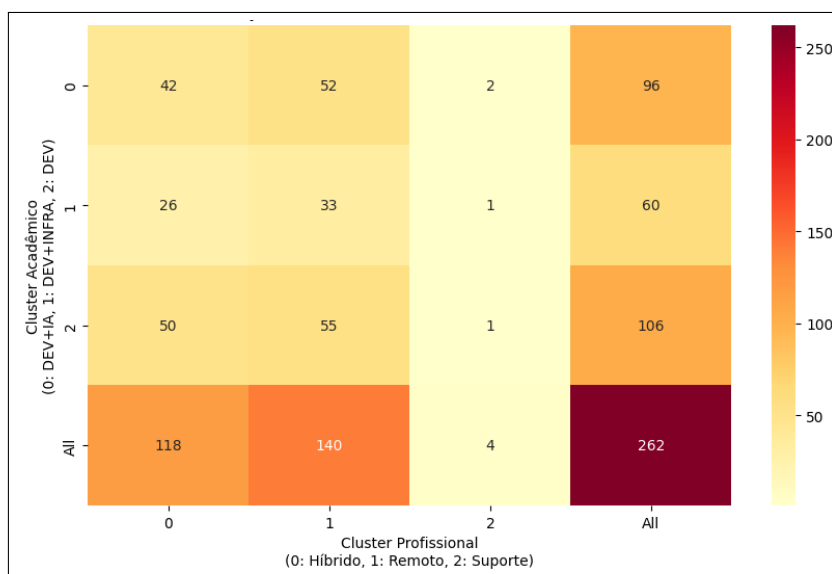


Figura 5.20: Mapa de calor de intersecção entre clusters acadêmicos e profissionais.

Embora a distribuição dos egressos entre as modalidades de trabalho seja relativamente homogênea entre os diferentes perfis acadêmicos, como discutido na análise da QP1, identificamos correlações específicas que revelam associações entre combinações de disciplinas e aspectos particulares da inserção profissional:

1. **Perfil DEV+INFRA e Modalidade Híbrida:** A correlação positiva (+0,42) entre este perfil e a preferência pelo trabalho híbrido sugere que disciplinas como Administra-

ção de Redes, Segurança da Informação e Sistemas Distribuídos desenvolvem competências frequentemente requeridas em ambientes que exigem presença física parcial, relacionadas à manutenção de infraestrutura.

2. **Perfil DEV+IA e Empresas Inovadoras:** A correlação negativa (-0,15) entre o perfil DEV+IA e a atuação em grandes empresas indica que disciplinas na área de inteligência artificial como TECC (Visualização de Dados) e afins, preparam os egressos para áreas emergentes. Embora este coeficiente seja fraco, sugere uma tendência de quanto maior a concentração nestas disciplinas, ligeiramente menor a probabilidade de trabalhar em grandes corporações, com preferência por startups e empresas de médio porte que valorizam inovação e experimentação tecnológica.
3. **Perfil DEV e Versatilidade:** Egressos com formação concentrada em desenvolvimento apresentam a distribuição mais equilibrada entre modalidades de trabalho, como evidenciado no mapa de calor. Disciplinas como Algoritmos Avançados e Programação proporcionam competências fundamentais adaptáveis a diversos contextos organizacionais, explicando a maior flexibilidade destes profissionais quanto ao ambiente de trabalho.

Quanto ao porte empresarial, egressos DEV+INFRA concentram-se em empresas médias e grandes (90%), enquanto DEV+IA distribuem-se entre diferentes portes, com tendência para empresas menores.

O tempo de inserção profissional apresenta correlações fracas (-0,12 a -0,18) com os perfis acadêmicos, sugerindo que fatores externos à formação curricular exercem influência mais significativa.

Com esta análise identificou-se que combinações específicas de disciplinas optativas estão associadas a tendências particulares de inserção profissional, embora estas associações sejam moderadas e não determinísticas. O conjunto de disciplinas em cada perfil acadêmico oferece direcionamentos que influenciam aspectos da transição profissional, especialmente quanto à natureza do trabalho e ao ambiente organizacional, mas com impacto limitado no tempo de inserção e na modalidade predominante de trabalho.

QP3 – Influência Curricular no Tempo e Área da Primeira Atuação

A terceira questão examina como os perfis curriculares influenciam o tempo e a área da primeira atuação profissional.

Influência no Tempo de Transição Profissional

A análise estatística revelou correlações fracas (-0,12 a -0,18) entre os perfis acadêmicos e o tempo de inserção no mundo do trabalho, indicando que as escolhas curriculares exercem influência limitada neste aspecto da transição profissional. Os dados mostram que:

- Aproximadamente 85% dos egressos de todos os perfis acadêmicos conseguem inserção em até 6 meses após a formatura, com variações estatisticamente não significativas entre os diferentes agrupamentos;
- O perfil **DEV+IA** apresenta o menor tempo médio de inserção (2,3 meses), seguido por **DEV** (2,7 meses) e **DEV+INFRA** (3,1 meses), porém estas diferenças não são estatisticamente significativas ($p > 0,05$);
- A exceção notável ocorre no cluster profissional de Suporte, que apresenta tempo médio de inserção superior a 24 meses, independentemente do perfil acadêmico de origem.

Estes resultados sugerem que fatores externos ao currículo formal, como experiências práticas, estágios, projetos extracurriculares e redes de contatos, exercem influência mais determinante no tempo de inserção profissional do que as escolhas específicas de disciplinas optativas.

Influência na Área de Primeira Atuação

Em contraste com o tempo de inserção, identificou-se correlações significativas entre os perfis acadêmicos e as áreas de primeira atuação profissional:

- Perfil **DEV+IA**: 79,2% dos egressos iniciam carreira em Desenvolvimento de Software (60,4%) ou Ciência de Dados/IA (18,8%), demonstrando forte alinhamento entre formação e atuação;
- Perfil **DEV+INFRA**: 68,3% atuam inicialmente em Desenvolvimento (48,3%) ou Segurança/Infraestrutura (20,0%), evidenciando como a formação em infraestrutura amplia as possibilidades de atuação nesta área;

- Perfil **DEV**: 83,0% concentram-se exclusivamente em Desenvolvimento de Software, refletindo a especialização técnica deste grupo.

Esta análise indica que, embora as escolhas curriculares tenham impacto limitado no tempo de inserção, exercem influência significativa na definição da área de atuação inicial, direcionando os egressos para campos profissionais alinhados à sua formação específica.

Implicações para o Desenvolvimento Curricular

Os achados apresentados sugerem três direcionamentos importantes para o aprimoramento do curso:

1. **Base Comum Efetiva:** A alta empregabilidade em curto prazo (até 6 meses) entre todos os perfis acadêmicos valida a eficácia da base comum do curso, que proporciona competências fundamentais para rápida inserção profissional, independentemente da especialização escolhida;
2. **Flexibilidade Adaptativa:** A atual estrutura de disciplinas optativas demonstra-se adequada para permitir especializações com impacto significativo na área de atuação, sem restringir as possibilidades de inserção profissional rápida;
3. **Desenvolvimento de Competências Transversais:** A fraca correlação entre perfil acadêmico e tempo de inserção evidencia a necessidade de fortalecer aspectos não capturados nas disciplinas técnicas, como habilidades interpessoais, empreendedorismo e práticas profissionais integradas.

A análise revela que, embora existam correlações significativas entre a formação acadêmica e a área de atuação inicial, a relação entre escolhas curriculares e tempo de inserção profissional é menos pronunciada. O curso cumpre seu papel ao proporcionar uma base técnica sólida, complementada por especializações que direcionam os egressos para áreas específicas, mas o tempo de inserção parece depender mais de fatores externos ao currículo formal.

Este cenário sugere a importância de manter a estrutura curricular flexível atual, potencialmente ampliando as oportunidades para desenvolvimento de competências transversais e experiências práticas que possam acelerar ainda mais a inserção profissional, especialmente

em áreas emergentes como Segurança e Suporte, que atualmente apresentam o maior tempo médio de transição ao mundo do trabalho.

O mapa de calor consolida a ideia de que não existe um padrão linear e homogêneo que associe os perfis acadêmicos e profissionais dos egressos. A distribuição dos egressos nos diferentes grupos profissionais, independente de seu desempenho acadêmico, demonstra que outros fatores, como as oportunidades no mundo do trabalho, a procura por determinado tipo de atividade ou modalidade de trabalho, são igualmente ou mais importantes para a definição da transição profissional do que o perfil acadêmico construído durante o curso. A representação visual dos dados reforça esta independência e nos mostra, de forma clara e objetiva, como a formação acadêmica influencia a escolha profissional desses egressos.

5.3.2 Resultados

A análise integrada dos dados acadêmicos e profissionais revelou padrões significativos na transição dos egressos para o mundo do trabalho. Esta seção apresenta os principais resultados obtidos através do cruzamento entre os perfis acadêmicos e os agrupamentos profissionais identificados.

Distribuição dos Perfis Acadêmicos nos Clusters Profissionais

A Tabela 5.4 apresenta a distribuição percentual dos três perfis acadêmicos entre os clusters profissionais identificados. Observa-se uma distribuição notavelmente homogênea, com todos os perfis acadêmicos mantendo proporções similares em cada modalidade de atuação profissional. Os três perfis acadêmicos apresentam participação expressiva nos clusters

Clusters	Desenvolvedores Híbridos	Desenvolvedores Remotos	Especialista em Suporte
DEV+IA	43,75%	54,17%	2,08%
DEV+INFRA	43,33%	55,00%	1,67%
DEV	47,17%	51,89%	0,94%

Tabela 5.4: Percentual de cada perfil acadêmico nos diferentes clusters profissionais.

profissionais Híbrido (Cluster 0) e Remoto (Cluster 1), com variações percentuais estatisticamente não significativas entre eles. Em contraste, o Cluster Profissional 2 (Suporte) apresenta

representação consistentemente baixa (0,94% a 2,08%) em todos os perfis acadêmicos.

Esta homogeneidade na distribuição sugere que a formação generalista do curso proporciona flexibilidade na escolha da modalidade de trabalho, sem que a ênfase acadêmica específica determine rigidamente a trajetória profissional. Embora existam correlações sutis, como a ligeira maior proporção do perfil DEV no cluster Híbrido (47,17%) em comparação com DEV+IA (43,75%) e DEV+INFRA (43,33%), estas diferenças não são estatisticamente significativas para indicar determinismo entre formação e atuação.

Modalidades de Trabalho Predominantes

A análise das modalidades de trabalho, ilustrada na Figura 5.21, revela uma clara predominância das modalidades remota (53,4%) e híbrida (45,0%), com participação minoritária da modalidade suporte (1,6%). Este padrão indica uma tendência consolidada no setor tecnológico para arranjos de trabalho flexíveis, independentemente da formação específica dos egressos.

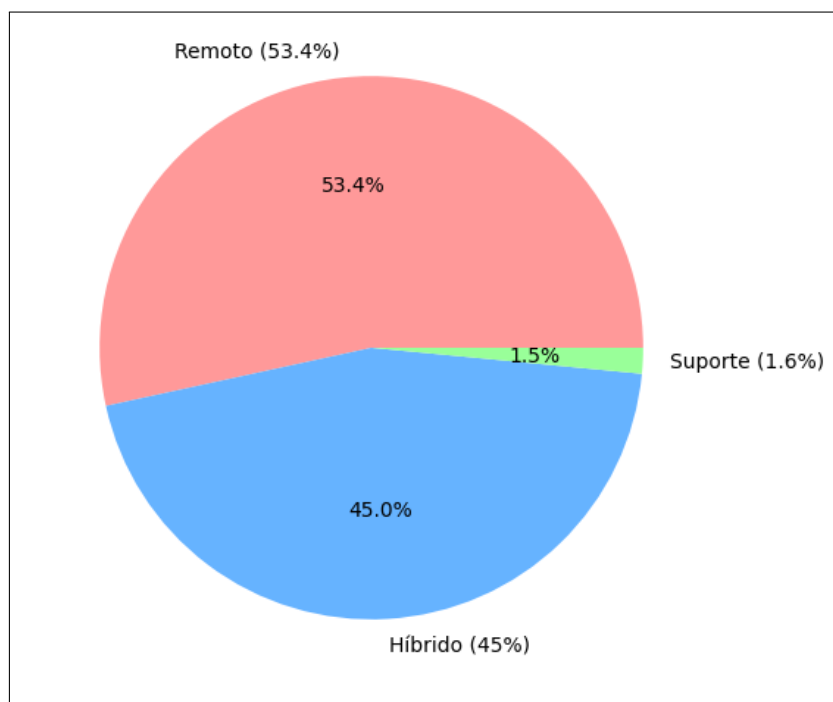


Figura 5.21: Gráfico de pizza, distribuição dos clusters profissionais em percentagens.

Esta preferência por modalidades flexíveis de trabalho (98,4% entre remoto e híbrido) representa uma adaptação significativa às tendências contemporâneas do setor tecnológico,

sugerindo que fatores externos à formação acadêmica, como adaptabilidade ao trabalho distribuído e evolução do mercado, exercem influência determinante nas transições profissionais.

Análise Quantitativa dos Agrupamentos

A análise da distribuição absoluta entre clusters acadêmicos e profissionais revelou concentrações significativas que complementam as análises percentuais. O perfil DEV+IA (106 egressos) apresentou distribuição equilibrada entre trabalho remoto (55 egressos) e híbrido (50 egressos). Padrão similar foi observado no perfil DEV (96 egressos), com 52 em modalidade remota e 42 em híbrida. O perfil DEV+INFRA (60 egressos) manteve proporção equivalente, com 33 egressos em trabalho remoto e 26 em híbrido. A consistência desta distribuição proporcional entre diferentes perfis acadêmicos reforça a conclusão de que a formação, embora fundamental para o desenvolvimento de competências específicas, não determina rigidamente a modalidade de trabalho adotada pelos egressos. Esta evidência sugere a importância de fatores adicionais, como oportunidades disponíveis, preferências pessoais e tendências de mercado, na definição das trajetórias profissionais.

Alinhamento entre Disciplinas Optativas e Áreas de Atuação

A análise das disciplinas mais cursadas por perfil acadêmico revelou alinhamentos significativos com áreas de atuação, apesar da distribuição homogênea entre modalidades de trabalho:

- **Perfil DEV+IA:** Disciplinas como Ciência de Dados Preditiva (73 egressos), Ciência de Dados Descritiva (66) e Gerência da Informação (65) foram predominantes. Estes egressos apresentaram maior concentração em áreas de Desenvolvimento de Software (60,4%) e Ciência de Dados/IA (18,8%).
- **Perfil DEV:** Verificação e Validação de Software (69), Gerência da Informação (63) e Gestão de Projetos (46) constituíram as principais escolhas, refletindo-se na concentração em Desenvolvimento de Software (83,0%).
- **Perfil DEV+INFRA:** Segurança de Sistemas (43), Administração de Sistemas (33) e Sistemas Distribuídos (27) foram prioridades. Estes egressos apresentaram maior diversificação entre Desenvolvimento (48,3%) e Segurança/Infraestrutura (20,0%).

Este alinhamento entre disciplinas cursadas e áreas de atuação profissional, combinado com a distribuição homogênea entre modalidades de trabalho, sugere que a formação acadêmica influencia principalmente o campo de especialização técnica, enquanto fatores externos determinam aspectos como modalidade e ambiente de trabalho.

Recomendações para o Aprimoramento Curricular

A análise integrada dos dados permite identificar cinco frentes de atuação com potencial para aprimorar o projeto pedagógico do curso, considerando as tendências observadas nas transições dos egressos e as demandas do mundo do trabalho tecnológico. A seguir, são apresentadas recomendações com foco em aplicabilidade gradual e fundamentadas nos resultados obtidos.

1. Integração entre Teoria e Prática: A distribuição relativamente homogênea entre modalidades de trabalho, independentemente da formação específica, sugere que fatores práticos e experiências externas ao currículo também influenciam a inserção profissional. Com isso, recomenda-se:

- Implantar projetos integradores semestrais, organizados por problemas reais (Problem-Based Learning), envolvendo equipes multidisciplinares de estudantes;
- Atualizar a disciplina "Projeto em Computação", incorporando metodologias ágeis e práticas como integração contínua e DevOps;
- Estabelecer parcerias com empresas locais do setor tecnológico para o desenvolvimento de atividades conjuntas em disciplinas avançadas, fortalecendo a relação entre academia e mundo do trabalho.

2. Adequação ao Trabalho Distribuído: A predominância de modalidades remota e híbrida (98,4% combinadas) aponta para a necessidade de preparar os estudantes para dinâmicas de trabalho em ambientes distribuídos. Como ações possíveis, propõe-se:

- Criar uma disciplina optativa sobre “Colaboração e Produtividade em Ambientes Distribuídos”, tratando ferramentas, metodologias e boas práticas;
- Incluir conteúdos específicos sobre gestão de equipes remotas em disciplinas como "Engenharia de Software" e "Gestão de Projetos";

- Adotar metodologias de ensino híbridas em disciplinas avançadas, com parte das atividades desenvolvidas em formato remoto e assíncrono.

3. **Desenvolvimento de Competências Transversais:** A análise sugere que habilidades interpessoais e comunicacionais contribuem para a empregabilidade dos egressos. Assim, recomenda-se:

- Revisar componentes avaliativos de disciplinas como "Metodologia Científica" e "Projeto em Computação", inserindo treinamentos de comunicação técnica e apresentações interdisciplinares;
- Introduzir avaliações por pares em disciplinas práticas, incentivando habilidades de feedback, revisão de código e trabalho colaborativo;
- Promover ciclos de seminários com profissionais do setor e ex-alunos, abordando temas como gestão de carreira, soft skills e desenvolvimento profissional.

4. **Fortalecimento da Formação em Infraestrutura:** A menor proporção de egressos no perfil DEV+INFRA, comparado aos demais, pode estar associada a um conjunto mais restrito de oportunidades formativas nessa área. Algumas ações possíveis incluem:

- Oferecer novas disciplinas optativas em temas como DevSecOps, Arquitetura Cloud-Native e Infraestrutura como Código;
- Criar trilhas formativas com sugestões de disciplinas e atividades práticas para estudantes com interesse em infraestrutura;
- Apoiar a criação de laboratórios especializados em segurança e redes, com foco em ambientes realistas e possibilidades de parceria com o setor produtivo.

5. **Apoio a Áreas com Inserção Profissional Mais Tardia:** O cluster de egressos classificados como Especialistas em Suporte apresenta tempo de inserção profissional mais elevado. Nesse contexto, propõe-se:

- Implantar programas de mentoria voltados para estudantes interessados em áreas com entrada profissional mais gradual;
- Oferecer módulos de formação complementar, com foco em certificações reconhecidas em infraestrutura e suporte técnico;

- Introduzir disciplinas tópicas como "Gestão de Serviços de TI" ou "Operações de TI", com abordagens práticas inspiradas em frameworks como ITIL e SRE.

A implementação dessas propostas pode ocorrer de forma progressiva, a partir da priorização de ações com maior viabilidade institucional e impacto potencial. Medidas como projetos integradores e a atualização de disciplinas já existentes podem ser incorporadas ao curto prazo, enquanto iniciativas mais estruturantes, como a criação de laboratórios ou novos componentes curriculares, podem ser planejadas a médio prazo.

Ainda que os resultados indiquem que o curso oferece uma base sólida e flexível para múltiplos caminhos profissionais, as recomendações apresentadas buscam promover um alinhamento ainda mais direto com as mudanças no mundo do trabalho, sem comprometer a qualidade acadêmica já observada na formação dos egressos.

5.3.3 Limitações e Trabalhos Futuros

Esta pesquisa oferece contribuições significativas para a compreensão da relação entre escolhas curriculares e transição profissional, porém apresenta limitações que devem ser consideradas na interpretação dos resultados e que apontam direções para investigações futuras.

Limitações do Estudo

1. **Configuração de Clusters:** A escolha pelo número de clusters ($k=3$) na análise dos dados profissionais, embora fundamentada nos métodos do Cotovelo e Silhueta, representa um compromisso entre detalhamento e significância estatística. O método da Silhueta indicou que valores mais elevados de k poderiam revelar subgrupos adicionais, porém o tamanho da amostra (263 egressos com dados profissionais completos) poderia comprometer a representatividade estatística de clusters muito pequenos em configurações mais granulares.
2. **Escopo Temporal:** O estudo abrange egressos entre 2012.2 e 2024.1, período que coincide com transformações profundas no mercado de tecnologia, incluindo a disseminação do trabalho remoto, a consolidação da computação em nuvem e a emergência da inteligência artificial aplicada. Estas mudanças afetaram significativamente os

padrões de inserção profissional, introduzindo possíveis variáveis confundidoras nas análises de correlação.

3. **Foco na Inserção Inicial:** A análise centrou-se predominantemente na primeira ocupação profissional, não capturando adequadamente a evolução subsequente das carreiras. Esta limitação é particularmente relevante no setor de tecnologia, caracterizado por alta mobilidade, rápidas reconversões profissionais e trajetórias não-lineares.
4. **Variáveis Não Consideradas:** A pesquisa não incorporou fatores potencialmente relevantes como experiências extracurriculares (estágios, iniciação científica, projetos de extensão), atividades complementares e redes de relacionamento profissional, que podem exercer influência significativa nas transições profissionais.
5. **Especificidade Contextual:** Os resultados refletem prioritariamente a realidade dos egressos da UFCG e do ecossistema tecnológico de Campina Grande e região, limitando a generalização direta para outros contextos institucionais e geográficos sem consideração das particularidades locais.

Perspectivas para Pesquisas Futuras

Os achados e limitações deste estudo sugerem diversas linhas promissoras para investigações futuras:

1. **Diversificação Metodológica:** A aplicação de técnicas complementares de clusteração como DBSCAN (adequado para identificar agrupamentos de formato não-esférico) e Hierarchical Clustering (que não requer definição prévia do número de clusters) poderia revelar padrões não detectados pelo K-Means, especialmente considerando a natureza multidimensional e potencialmente não-linear dos dados educacionais e profissionais.
2. **Análise Longitudinal de Trajetórias:** O desenvolvimento de um estudo de coorte para acompanhar sistematicamente a evolução profissional dos egressos ao longo de 5-10 anos permitiria compreender como as escolhas curriculares influenciam não apenas a inserção inicial, mas também as progressões e reconversões de carreira a médio e longo prazo.

3. **Integração de Fatores Socioeconômicos e Demográficos:** A incorporação de variáveis como origem socioeconômica, gênero, raça e histórico escolar anterior à graduação poderia revelar como estes fatores interagem com as escolhas curriculares na determinação dos caminhos profissionais, identificando possíveis disparidades e oportunidades para intervenções institucionais mais equitativas.
4. **Modelos Preditivos para Orientação Acadêmica:** O desenvolvimento de sistemas de recomendação baseados em aprendizado de máquina que, considerando o histórico acadêmico, aspirações profissionais e tendências de mercado, possam sugerir combinações otimizadas de disciplinas optativas, representaria um avanço significativo nas ferramentas de orientação acadêmica.
5. **Análise Comparativa Interinstitucional:** A extensão da metodologia para incluir egressos de instituições com diferentes abordagens pedagógicas e curriculares, como IFPB, UEPB e FACISA, permitiria identificar práticas institucionais particularmente eficazes na promoção da empregabilidade e avaliar o impacto de diferentes ênfases formativas nas transições profissionais.
6. **Incorporação de Experiências Extracurriculares:** A ampliação do escopo analítico para incluir variáveis como participação em projetos de pesquisa, extensão, monitorias, estágios não-obrigatórios e atividades de empreendedorismo poderia fornecer uma visão mais abrangente dos fatores que complementam a formação curricular formal na preparação para o mundo do trabalho.
7. **Análise Aprofundada do Cluster de Suporte:** Considerando a baixa representatividade e o tempo de inserção significativamente maior do cluster profissional de Suporte (apenas 1,6% dos egressos), investigações específicas sobre este grupo poderiam revelar fatores determinantes para esta trajetória diferenciada e identificar oportunidades para melhor preparação dos estudantes interessados nesta especialização.

Estas linhas de investigação prometem não apenas aprofundar a compreensão científica sobre a relação entre formação acadêmica e trajetórias profissionais, mas também fornecer subsídios práticos para o aprimoramento contínuo dos currículos universitários em Ciência da Computação. Ao expandir as dimensões analíticas do presente estudo, pesquisas futuras

poderão contribuir para o desenvolvimento de abordagens educacionais mais adaptadas às crescentes demandas de um setor tecnológico em constante evolução.

A presente pesquisa constitui um passo significativo na compreensão da complexa relação entre escolhas curriculares e inserção profissional em Ciência da Computação, oferecendo evidências empíricas que podem orientar tanto decisões institucionais quanto escolhas individuais dos estudantes. Os resultados sugerem que, embora as combinações específicas de disciplinas influenciem as áreas de atuação profissional, a adaptabilidade, as competências transversais e a capacidade de aprendizado contínuo permanecem fatores determinantes para o sucesso no dinâmico ecossistema tecnológico contemporâneo.

Com base nos resultados obtidos e nas limitações identificadas, várias direções prometem avanços significativos para pesquisas futuras:

1. **Aplicação de Métodos Alternativos de Clusterização:** Explorar técnicas como DBSCAN, Agglomerative Clustering e Gaussian Mixture Models poderia revelar padrões não capturados pelo K-Means, especialmente considerando a natureza potencialmente não-esférica dos clusters em espaços multidimensionais de dados educacionais e profissionais;
2. **Análise Longitudinal de Carreiras:** Expandir o estudo para acompanhar a evolução profissional dos egressos ao longo do tempo permitiria uma compreensão mais profunda das trajetórias de carreira e do impacto de longo prazo das escolhas curriculares;
3. **Incorporação de Variáveis Socioeconômicas:** Integrar dados demográficos e socioeconômicos à análise poderia revelar como fatores externos à formação acadêmica interagem com as escolhas curriculares na determinação dos caminhos profissionais;
4. **Modelos Preditivos de Carreiras:** Desenvolver modelos supervisionados que, baseados nas escolhas curriculares e outras variáveis relevantes, possam prever trajetórias profissionais, oferecendo ferramentas de orientação acadêmica fundamentadas em evidências;
5. **Análise Comparativa entre Instituições:** Estender a metodologia para incluir egressos de diferentes instituições como o IFPB, a UEPB e a Facisa, permitiria avaliar como

diferentes abordagens pedagógicas e estruturas curriculares influenciam a empregabilidade e as transições profissionais, com cursos cujos enfoques são diferente do da UFCG;

Esta pesquisa contribui para o entendimento da complexa relação entre formação acadêmica e inserção profissional em Ciência da Computação, fornecendo subsídios para o aprimoramento contínuo dos currículos universitários e para a orientação de estudantes na construção de trajetórias acadêmicas alinhadas a seus objetivos profissionais. Os resultados sugerem que, embora as escolhas curriculares possam direcionar caminhos específicos, a adaptabilidade e o desenvolvimento de competências transversais permanecem fatores críticos para o sucesso profissional no dinâmico campo da tecnologia.

Capítulo 6

Conclusões

Esta pesquisa investigou a relação entre as escolhas curriculares e a transição profissional dos egressos do curso de graduação em Ciência da Computação da UFCG, utilizando técnicas de mineração de dados para identificar padrões nas trajetórias acadêmicas e profissionais. Com a aplicação do algoritmo K-Means, foi possível segmentar perfis acadêmicos e profissionais, permitindo compreender de que maneira a formação influencia a inserção no mercado de trabalho.

A análise dos dados acadêmicos de 352 egressos revelou três perfis principais de formação: DEV (foco em desenvolvimento de software), DEV+IA (combinação de desenvolvimento com inteligência artificial) e DEV+INFRA (integração entre desenvolvimento e infraestrutura computacional). Esses agrupamentos representam diferentes ênfases construídas a partir das escolhas de disciplinas optativas. Paralelamente, a análise dos dados profissionais de 263 egressos identificou três perfis de atuação: Desenvolvedores Híbridos (45%), caracterizados por inserção rápida e preferência pelo modelo híbrido, majoritariamente em empresas de médio e grande porte; Desenvolvedores Remotos (53,4%), com inserção igualmente rápida, porém maior presença em trabalho remoto e empresas de portes variados; e Especialistas em Suporte (1,6%), grupo menor, com inserção mais tardia, atuação presencial e vínculo com grandes empresas.

A investigação produziu evidências empíricas que permitem responder às três questões de pesquisa propostas. Na QP1, que buscava saber a Relação entre escolhas curriculares e padrões de inserção profissional inicial, observou-se que as disciplinas optativas influenciam a área de atuação inicial, mas apresentam relação limitada com a modalidade de trabalho e o

tempo até a inserção. A distribuição equilibrada dos perfis acadêmicos entre as modalidades híbrida, remota e presencial sugere que fatores extracurriculares, como mercado, contexto socioeconômico e habilidades transversais, desempenham papel importante nesse processo.

Já na QP2, sobre as combinações de disciplinas e suas associações com a transição profissional, foram identificadas correlações moderadas, o perfil DEV+INFRA apresenta associação positiva (+0,42) com o trabalho híbrido; o perfil DEV+IA demonstra correlação negativa (-0,15) com atuação em grandes empresas, indicando possível afinidade com ambientes mais dinâmicos e informais, como startups e empresas de médio porte. Já o perfil DEV mostra distribuição mais uniforme, sugerindo versatilidade quanto ao ambiente profissional.

Por fim, na QP3, onde se buscava a influência das escolhas curriculares no tempo e na área da primeira atuação profissional, descobriu-se que as escolhas curriculares apresentam correlações fracas com o tempo de inserção (-0,12 a -0,18), mas influenciam de forma mais clara a área inicial de atuação. Egressos DEV+IA atuam majoritariamente em Desenvolvimento (60,4%) e Ciência de Dados/IA (18,8%); DEV+INFRA se distribuem entre Desenvolvimento (48,3%) e Segurança/Infraestrutura (20%); e DEV se concentram fortemente em Desenvolvimento de Software (83

Esses resultados indicam oportunidades para o aprimoramento do curso. A baixa correlação entre perfil acadêmico e tempo de inserção reforça a importância de ampliar a integração entre teoria e prática, por meio de projetos interdisciplinares que simulem contextos reais de trabalho. A predominância de modalidades remota e híbrida (98,4% combinadas) destaca a necessidade de desenvolver competências relacionadas ao trabalho distribuído, como comunicação assíncrona, gestão do tempo e uso de ferramentas colaborativas.

A distribuição semelhante dos perfis acadêmicos nos clusters profissionais indica que competências não técnicas também exercem influência na transição profissional. Assim, é recomendável fortalecer habilidades como comunicação, trabalho em equipe e resolução de problemas complexos dentro das disciplinas técnicas. Além disso, o menor número de egressos no perfil DEV+INFRA (60, frente a 106 em DEV+IA e 96 em DEV) pode sugerir a necessidade de incentivo a essa trilha, especialmente diante da relevância crescente da área de infraestrutura no setor tecnológico. O maior tempo de inserção profissional observado entre os Especialistas em Suporte sugere a importância de estratégias formativas específicas para este público.

De forma geral, os dados indicam que o curso de Ciência da Computação da UFCG oferece uma formação sólida e adaptável, permitindo aos egressos atuar em diferentes segmentos do mundo do trabalho, independentemente da ênfase escolhida nas optativas. A base comum voltada ao desenvolvimento de software tem-se mostrado eficaz na preparação de profissionais capazes de se adaptar a diferentes contextos e demandas profissionais.

Esta pesquisa contribui para o entendimento da relação entre formação acadêmica e inserção profissional na área de Ciência da Computação, oferecendo evidências que podem subsidiar tanto decisões curriculares institucionais quanto escolhas individuais dos estudantes. Os resultados indicam que o curso cumpre seu papel ao oferecer uma formação equilibrada entre fundamentos gerais e oportunidades de especialização, ao passo que melhorias relacionadas à prática profissional, competências transversais e atualização frente às tendências do trabalho contemporâneo podem fortalecer ainda mais a formação oferecida.

Bibliografia

- [1] Charu C Aggarwal and Charu C Aggarwal. *Data classification*. Springer, 2015.
- [2] Pervaiz Akhtar, Muhammad Moazzam, Anika Ashraf, and Muhammad Naveed Khan. The interdisciplinary curriculum alignment to enhance graduates' employability and universities' sustainability. *The International Journal of Management Education*, 22(3):101037, 2024.
- [3] H. Al-Dossari, F. A. Nughaymish, Z. Al-Qahtani, M. Alkahlifah, and A. Alqahtani. A machine learning approach to career path choice for information technology graduates. *Engineering, Technology & Applied Science Research*, 10(6):6589–6596, Dec. 2020.
- [4] Rodrigo de Andrade Rolim Bem, Giovanna Mendes Garbácio, Marcelo Henrique Oliveira Henklain, Acauan Cardoso Ribeiro, and Reinaldo Viana Alvares. Perfil de alunos e egressos do curso de ciência da computação: investigação sobre interesses profissionais e contexto de trabalho. *Cadernos de Educação Tecnologia e Sociedade*, 17(1):225–246, 2024.
- [5] V. Murali Bhaskaran, Sudharsan V, and Solaiyappan AT. Student career prediction based on student performance using machine learning. 1:1212–1217, 2024.
- [6] Rowena Blokker, Jos Akkermans, Julian Marciniak, Paul G W Jansen, and Svetlana N Khapova. Organizing school-to-work transition research from a sustainable career perspective: A review and research agenda. *Work, Aging and Retirement*, 9(3):239–261, 04 2023.
- [7] José Camacho, José Manuel García-Giménez, Noemí Marta Fuentes-García, and Ga-

- bríel Maciá-Fernández. Multivariate big data analysis for intrusion detection: 5 steps from the haystack to the needle. *Computers & Security*, 87:101603, 2019.
- [8] Ruan Miguel da Silva Costa. Análise profissional dos egressos do curso de telemática no ifpb campus campina grande. 2023.
- [9] Raí da Silva Rodrigues, Eriky Ryan Gonçalves, Israely Lima, Marina Rocha, Pedro C Chaaban, and Jacilane de H Rabelo. Explorando profissões da computação na prática: Compartilhamento de conhecimento no ciclo de palestras do learninglab-um relato de experiência. pages 2483–2493, 2024.
- [10] Flávio de Brum, Patricia Mariotto Mozzaquatro, and Jocias Maier Zanatta. Estudo sobre os algoritmos de clusterização hierarchical clusterer e simple k-means aplicados no agrupamento de padrões similares. *Revista da Universidade Vale do Rio Verde*, 17(1), 2019.
- [11] Bruno Rodrigues de Oliveira, Francisco Charles dos Santos Silva, Ricardo Mezzomo, Leandra Matos Barrozo, Tatiane Scilewski da Costa Zanatta, Joel Cabral dos Santos, Carlos Henrique Conceição Sousa, Yago Pinto Coelho, Aurilucia do Nascimento Silva Caldas, and Alan Mario Zuffo. Temporal variability in soybean sowing and harvesting according to k-means and silhouette scores. *Trends in Agricultural and Environmental Sciences*, pages e240010–e240010, 2024.
- [12] Liubov Dombrovskaya, Erika Riveros, and Patricio Rodríguez Valdés. Analysis of trajectories of students in higher technical-professional education using data mining methods. pages 1–8, 2024.
- [13] William Donald, Yehuda Baruch, and Melanie Ashleigh. The undergraduate self-perception of employability: Human capital, careers advice, and career ownership. *Studies in Higher Education*, 44:599–614, 03 2019.
- [14] Lia D Falco. The school counselor and stem career development. *Journal of Career Development*, 44(4):359–374, 2017.
- [15] Dora Lorejan Avila Fernandes et al. Promoção de igualdade de gênero: O uso de

inteligência artificial em processos de recrutamento e seleção em empresas brasileiras. *FGV RIC Revista de Iniciação Científica*, 3, 2022.

- [16] Julio Cezar Costa Furtado, Edson Monteiro Neto, and Stephany Dantas de Freitas Furtado. Um estudo sobre as principais dificuldades na empregabilidade de um engenheiro de software. *CONTRIBUCIONES A LAS CIENCIAS SOCIALES*, 17(3), mar. 2024.
- [17] B.D. Gannod, G.C. Gannod, and M.R. Henderson. Course, program, and curriculum gaps: Assessing curricula for targeted change. pages T3C–T3C, 2005.
- [18] Qi-Li Gao, Chen Zhong, Yang Yue, Rui Cao, and Bowen Zhang. Income estimation based on human mobility patterns and machine learning models. *Applied Geography*, 163:103179, 2024.
- [19] Muhammad Umair Hassan, Saleh Alaliyat, Raheem Sarwar, Raheel Nawaz, and Ibrahim A. Hameed. Leveraging deep learning and big data to enhance computing curriculum for industry-relevant skills: A norwegian case study. *Heliyon*, 9(4):e15407, 2023.
- [20] Abiodun M. Ikotun, Absalom E. Ezugwu, Laith Abualigah, Belal Abuhaija, and Jia Heming. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622:178–210, 2023.
- [21] Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.
- [22] Areej Kamal, Batul Naushad, Hadia Rafiq, and Shahab Tahzeeb. Smart career guidance system. pages 1–7, 2021.
- [23] Priyank Kansal and Harsh Sadawarti. Empowering the future: An in-depth investigation of computer science graduate employability cognitive factors. pages 1–5, 2023.
- [24] Sangita Lade, Ayush Billade, Akshat Chandrapatle, Srinivas Chenna, and Girish Chinchalpalle. Linkedin alumni profile data extraction. pages 174–178, 2024.

- [25] Pedro HR Macêdo, Wylliams B Santos, and Alexandre MA Maciel. Análise de perfis de engajamento de estudantes de ensino a distância. *RENOTE*, 18(2):326–335, 2020.
- [26] Kobus Maree. Utilizing career adaptability and career resilience to promote employability and decent work and alleviate poverty. pages 349–373, 2017.
- [27] Leonardo Torres Marques, Bruno Torres Marques, and Carlos Alexandre Morais Silva. Análise do perfil profissional de egressos de cursos de tecnologia da informação: um estudo de caso na (oculto). *Revista Novas Tecnologias na Educação*, 22(1):466–476, 2024.
- [28] Yoshitatsu Matsuda, Takayuki Sekiya, and Kazunori Yamaguchi. Curriculum analysis of computer science departments by simplified, supervised lda. *Journal of Information Processing*, 26:497–508, 2018.
- [29] Snježana Milinković, Vladimir Vujović, Zorana Štaka, and Milica Vuković. Clustering-based students’ descriptive model. pages 1–6, 2023.
- [30] Kalpana Pali and Laxmikant Tiwari. Predictive learning and career path using artificial intelligence. pages 1–9, 2024.
- [31] Suhem Parack, Zain Zahid, and Fatima Merchant. Application of data mining in educational databases for predicting academic trends and patterns. *2012 IEEE International Conference on Technology Enhanced Education (ICTEE)*, pages 1–4, 2012.
- [32] Katarina Pažur Aničić, Blaženka Divjak, and Krunoslav Arbanas. Preparing ict graduates for real-world challenges: Results of a meta-analysis. *IEEE Transactions on Education*, 60(3):191–197, 2017.
- [33] Raniele Conceição Pereira, Moisés Inácio Rodrigues, Alice Manuela da Silva de Azambuja, Jocemar de Quadros Chagas, and Leila Freire. Arte e ciência em diálogo: reflexões éticas sobre tecnologia na engenharia de software e inteligência artificial (ia). *Com a Palavra, o Professor*, 9(25):238–257, 2024.
- [34] Satrio Adi Priyambada, Mahendrawathi Er, Bernardo Nugroho Yahya, and Tsuyoshi Usagawa. Profile-based cluster evolution analysis: Identification of migration patterns for understanding student learning behavior. *IEEE Access*, 9:101718–101728, 2021.

- [35] Eivis Qenani, Neal MacDougall, and Carol Sexton. An empirical study of self-perceived employability: Improving the prospects for student employment success in an uncertain environment. *Active Learning in Higher Education*, 15(3):199–213, 2014.
- [36] Mushfiqur Rahman and Md. Asadujjaman. Multi-criteria decision making for job selection. pages 152–156, 2021.
- [37] Milind Rane, Sahil Kalal, Jenil Chandegara, Toshish Kakkad, Vishwas Jain, and Simran Jagtap. Career prediction website using machine learning. pages 1–7, 2023.
- [38] Hannu Rätty, Ulla Hytti, Kati Kasanen, Katri Komulainen, Päivi Siivonen, and Inna Kozlinska. Perceived employability and ability self among finnish university students. *European Journal of Psychology of Education*, 35:975–993, 2020.
- [39] Fabián Riquelme, Roberto Muñoz, Marco Antonio Vivar, and Jean Billiard. Studying alumni’s education and work experience through social network analysis: A linkedin case study. pages 1–9, 2023.
- [40] Cristóbal Romero and Sebastián Ventura. Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews)*, 40(6):601–618, 2010.
- [41] Tara Safavi, Maryam Davoodi, and Danai Koutra. Career transitions and trajectories: A case study in computing. page 675–684, 2018.
- [42] Ismayle S Santos, Rafael L Gomes, and William S Vilaça. Avaliação da eficiência de sistemas educacionais aplicando técnicas de agrupamento e dea. pages 1143–1153, 2023.
- [43] Zike Shao and Chunyang Sun. Classification method of higher vocational education students based on k-means clustering algorithm. pages 110–114, September 2023.
- [44] George Siemens and Ryan S. J. d. Baker. Learning analytics and educational data mining: towards communication and collaboration. *Association for Computing Machinery*, page 252–254, 2012.

- [45] António João Silva, Paulo Cortez, Carlos Pereira, and André Pilastri. Business analytics in industry 4.0: A systematic review. *Expert systems*, 38(7):e12741, 2021.
- [46] Adriana Benevides Soares, Luciana Mourão, and Marcia Cristina Monteiro. *O Estudante Universitário Brasileiro: Permanência, Habilidades Sociais, Competência Social e Relações Com o Mundo do Trabalho*. Editora Appris, 2024.
- [47] Sangam Suman, Simran Jeet Kaur, Aryan Sharma, and Siddharth Kumar. Machine learning-based system for admission and jobs prediction in engineering and technology sector. 5:463–468, 2024.
- [48] Hasanuzzaman Tushar and Nanta Sooraksa. Global employability skills in the 21st century workplace: A semi-systematic literature review. *Heliyon*, 9(11):e21023, 2023.
- [49] Monika Gupta Vashisht and Reena Grover. Employability profiling of engineering students using clustering. pages 191–195, 2019.
- [50] Nicolas Pastorio Boa Vista, Michele Ferraz Figueiró, and Patricia Mariotto Mozzaquatro Chicon. Técnicas de mineração de dados aplicadas aos microdados do enade para avaliar o desempenho dos acadêmicos do curso de ciência da computação no rio grande do sul utilizando o software r. *I Seminário de Pesquisa Científica e Tecnológica*, 1(1), 2017.
- [51] Minh-Thanh Vo, Trang Nguyen, and Tuong Le. Opt-bag model for predicting student employability. *Computers, Materials and Continua*, 76(2):1555–1568, 2023.
- [52] Rebecca Zarch, Sarah Dunton, Joshua Childs, and Anne Leftwich. Through the looking glass: Computer science education and the unintended consequences of broadening participation policy efforts. 1:1–4, 2020.
- [53] Zhaowei Zhang, Lin Liu, Hui Liu, Zhenyu Wu, and Hui Liu. Student profile clustering based personalized exercise recommendation: Taking data structures course as an example. pages 68–70, 2023.
- [54] Saima Zubair and Amani Moazzam. Exploring career construction: A single narrative case study guided by the systems theory framework of career development (stfcd). *South Asian Journal of Business and Management Cases*, 13(1):81–97, 2024.

Apêndice A

Disciplinas optativas ofertadas por período

Código	Área	Disciplina	Períodos Ofertados													
1302123	geral	COMPUTAÇÃO E MÚSICA	2021.2	2022.2												
1411176	dev	GERÊNCIA DA INFORMAÇÃO	2013.2	2014.1	2014.2	2015.1	2015.2	2016.1	2016.2	2017.1	2017.2	2020.1				
1411188	geral	MÉTODOS E SOFTWARE NUMÉRICOS	2015.2	2016.1	2017.1	2017.2	2018.1	2018.2	2019.1	2019.2	2021.1					
1411194	dev	BANCO DE DADOS II	2015.1	2015.2	2016.1	2017.1	2017.2	2018.1	2019.1	2020.2	2021.2	2022.1				
1411197	geral	AVAL.DE DESEMPENHO DE SISTEMAS DISCRETOS	2016.1	2016.2	2017.1	2017.2	2018.2	2019.1								
1411209	dev	INTERFACE HOMEM-MÁQUINA	2021.2													
1411217	infra	PROJETO DE REDES DE COMPUTADORES	2020.0	2020.2	2021.2	2022.2										
1411221	dev	SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS	2020.1	2021.1												
1411222	infra	SISTEMAS DISTRIBUÍDOS	2018.1	2019.2	2021.1	2022.1	2022.2									
1411270	dev	TECC (PROGRAMAÇÃO III)	2015.1	2017.2												
1411295	geral	TECC(PROJETO EM ARTE E TECNOLOGIA)	2015.2	2016.2												
1411302	dev	TECC(PROGRAMAÇÃO FUNCIONAL)	2017.2	2018.1	2018.2	2019.1	2019.2	2021.2								
1411320	infra	ADMINISTRAÇÃO DE SISTEMAS	2018.2	2019.1	2020.0	2020.1	2020.2	2021.1	2022.2							
1411321	dev	ALGORITMOS AVANÇADOS I	2018.1	2018.2	2019.2	2020.0	2020.2	2022.2								
1411322	dev	ALGORITMOS AVANÇADOS II	2018.2	2019.2	2020.0	2020.2	2021.1									
1411323	dev	ALGORITMOS AVANÇADOS III	2019.1													
1411324	dev	ALGORITMOS AVANÇADOS IV	2019.2	2021.1	2022.1											
1411325	dev	ARQUITETURA DE SOFTWARE	2020.0	2020.1	2020.2	2021.2										
1411326	dev	DESENV. DE APLICAÇÕES CORP. AVANÇADAS	2018.1	2019.2												
1411328	geral	ECONOMIA DE TECNOLOGIA DA INFORMAÇÃO	2018.1	2018.2	2019.1	2019.2	2020.0	2020.1	2020.2	2021.2	2022.1					
1411329	geral	EMPREENDEDORISMO EM SOFTWARE	2018.1	2018.2	2019.1	2019.2	2022.1									
1411330	infra	GERÊNCIA DE REDES	2018.2	2019.2	2020.1	2021.1	2022.1									
1411331	infra	INTERCONEXÃO DE REDES DE COMPUTADORES	2018.1	2018.2	2019.1	2019.2	2021.1	2022.1								
1411333	geral	PRÁTICA DE ENSINO DE COMPUTAÇÃO I	2018.1	2018.2	2019.1	2019.2	2020.0	2020.1	2020.2	2021.1	2021.2	2022.1	2022.2			
1411334	geral	PRÁTICA DE ENSINO DE COMPUTAÇÃO II	2018.2	2019.1	2019.2	2020.0	2020.1	2020.2	2021.1	2021.2	2022.1	2022.2				
1411335	dev	PRINCÍPIOS DE DESENVOLVIMENTO WEB	2018.2	2019.1	2020.0	2021.1										
1411336	dev	PROGRAMAÇÃO EM BANCO DE DADOS	2020.1	2021.1												
1411337	infra	PROVIS. E OPER. DE INFRAESTRUTURAS	2019.1	2022.1												
1411338	ia	RECONHEC. DE PADRÕES E REDES NEURAIS	2020.1	2021.1	2022.1											
1411339	ia	RECUP. DE INFORMAÇÃO E BUSCA NA WEB	2018.1	2019.1	2021.2	2022.2										
1411340	infra	SEGURANÇA DE SISTEMAS	2020.0	2020.1	2020.2	2021.2	2022.2									
1411342	ia	SISTEMAS DE APOIO À DECISÃO	2018.1	2020.0	2022.1											
1411343	dev	VERIFICAÇÃO E VALIDAÇÃO DE SOFTWARE	2018.2	2019.1	2019.2	2020.0	2020.1	2020.2	2021.2	2022.1	2022.2					
1411349	infra	TECC(PROJETO DE SISTEMAS OPERACIONAIS)	2018.1	2018.2	2019.1	2019.2										
1411350	dev	TECC(GESTÃO DE PROJETOS)	2018.1	2018.2	2019.1	2020.0	2020.1	2020.2	2021.1	2022.2						
1411351	ia	TECC(CIÊNCIA DE DADOS DESCRITIVA)	2018.1	2019.1	2019.2	2020.1	2020.2	2021.1	2022.1							
1411352	ia	TECC(PERCEPÇÃO COMPUTACIONAL)	2018.1	2019.2	2020.0	2020.2	2021.2	2022.2								
1411355	ia	TECC(VISUALIZAÇÃO DE DADOS)	2018.2	2019.2	2020.1	2021.2										
1411356	ia	TECC(CIÊNCIA DE DADOS PREDITIVA)	2018.2	2019.2	2020.1	2021.2	2022.2									
1411357	geral	TECC(GOVERNANÇA DA INTERNET)	2019.1	2019.2	2020.0	2020.1	2020.2	2021.2	2022.2							
1411358	ia	TECC(PROCESSAMENTO DE LINGUAGEM NATURAL)	2020.0	2021.1	2022.1											
1411359	dev	TECC(JOGOS DIGITAIS)	2021.1	2022.1												
1411360	dev	TECC(PROGRAMAÇÃO PARA WEB)	2021.1	2022.1												
1411361	geral	TECC(TRANSFORMAÇÃO DIGITAL)	2021.1	2022.1												
1411362	geral	TECC(HABILIDADES SOCIOECONOMICAIS I)	2022.2													

Apêndice B

Primeira enquete

Mapeamento e Coleta de Dados dos Egressos do Curso de Graduação de Ciência da Computação da UFCG

Survey Análise da Atuação dos Egressos do curso de graduação em Ciência da Computação da Universidade Federal de Campina Grande - UFCG

PPGCC - Mestrando Felisberto Cláudio dos Santos; Orientador Francisco Vilar Brasileiro (Fubica)

* Indica uma pergunta obrigatória

1. Nome *

2. Gênero *

Marcar apenas uma oval.

☐ Feminino

☐ Masculino

☐ Prefiro não dizer

☐ Outro:

3. Idade *

Marcar apenas uma oval.

☐ Até 30

☐ De 31 a 45

☐ Mais de 45

4. Atualmente você está empregado(a): *

Marcar apenas uma oval.

- ☐ Atuando desenvolvendo atividades ligadas à área de Ciência da Computação
- ☐ Atuando em outra função
- ☐ Estou empreendendo na área de TI (conta própria)
- ☐ Não estou trabalhando
- ☐ Estou em busca de emprego
- ☐ Estou somente estudando
- ☐ Outro: _____

5. Qual o setor de atuação profissional após a graduação? *

Marcar apenas uma oval.

- ☐ Desenvolvimento de software, engenharia de dados, segurança da informação, análise de dados, entre outros
- ☐ Consultoria em TI
- ☐ Serviços de TI (suporte técnico, gerenciamento de redes, administração de banco de dados, etc.)
- ☐ Análise de dados
- ☐ Startups
- ☐ Educação e pesquisa em TI
- ☐ Outro: _____

6. A modalidade de trabalho é: *

Marcar apenas uma oval.

- ☐ Presencial
- ☐ Remoto
- ☐ Parte presencial e parte remoto
- ☐ Não estou trabalhando no momento

7. Geograficamente, onde se localiza a sede e/ou o ambiente de trabalho da empresa? *

Marcar apenas uma oval.

- ☐ Campina Grande-PB
- ☐ Dentro do Brasil (fora de Campina Grande-PB)
- ☐ Fora do Brasil

8. Onde é sua residência fixa: *

Marcar apenas uma oval.

- ☐ Campina Grande-PB
- ☐ Dentro do território do Brasil (fora de Campina Grande-PB)
- ☐ Fora do território do Brasil

9. Em qual esfera organizacional você se encontra lotado? *

Marque todas que se aplicam.

- ☐ Operacional
- ☐ Gerencial
- ☐ Estratégica
- ☐ Executiva (diretorias)
- ☐ Não estou atuando profissionalmente
- ☐ Outro: _____

10. Qual o nome do cargo que ocupa? *

11. Possui outra formação/especialização?

Marque todas que se aplicam.

- ☐ Especialização
- ☐ Mestrado
- ☐ Doutorado
- ☐ Não
- ☐ Outro: _____

12. Grau de satisfação dos egressos em relação à sua formação no curso de graduação em Ciência da Computação da UFCG. *

Marcar apenas uma oval.

- ☐ Insatisfeito
- ☐ Regular
- ☐ Bom
- ☐ Muito bom
- ☐ Excelente

13. Grau de satisfação dos egressos em relação à sua carreira. *

Marcar apenas uma oval.

- ☐ Insatisfeito
- ☐ Regular
- ☐ Bom
- ☐ Muito bom
- ☐ Excelente

14. Como você avalia a relevância das diferentes vertentes formativas da graduação em Ciência da Computação com relação às atividades que você desenvolve no trabalho ou nos estudos. *

Marque todas que se aplicam.

	Muito relevante	Relevante	Pouco relevante	Nada relevante	Não sei informar
Fundamentos matemáticos	☐	☐	☐	☐	☐
Fundamentos de computação	☐	☐	☐	☐	☐
Desenvolvimento de software	☐	☐	☐	☐	☐
Noções de infraestrutura computacional	☐	☐	☐	☐	☐
Inteligência artificial	☐	☐	☐	☐	☐
Tópicos especializados	☐	☐	☐	☐	☐

15. As atitudes a seguir se assemelham com a sua rotina de trabalho como especialista em TI e formado em Computação? *

Marque todas que se aplicam.

	Sempre	Algumas vezes	Raramente	Nunca	Não sei informar
Consultoria e prestação de serviços de TI	☐	☐	☐	☐	☐
Geração, análise, controle, acesso e utilização de informações em projetos de TI	☐	☐	☐	☐	☐
Gerenciamento de unidades de informação em projetos de TI	☐	☐	☐	☐	☐
Gestão de recursos de informação de diversas naturezas em projetos de TI	☐	☐	☐	☐	☐
Utilização de tecnologias para o incremento do uso eficiente da informação em projetos de TI	☐	☐	☐	☐	☐

16. As atitudes a seguir se assemelham com a sua rotina de trabalho como formado em Computação? *

Marque todas que se aplicam.

	Sempre	Algumas vezes	Raramente	Nunca	Não se aplica
Visão estratégica e prospectiva em projetos de TI	☐	☐	☐	☐	☐
Versatilidade e capacidade de adaptação a ambientes dinâmicos em projetos de TI	☐	☐	☐	☐	☐
Curiosidade e espírito de inovação em projetos de TI	☐	☐	☐	☐	☐
Competências em gestão de projetos de TI	☐	☐	☐	☐	☐
Aptidão para o trabalho e aprendizagem colaborativa em equipes de TI	☐	☐	☐	☐	☐
Conhecimento e compreensão do "universo" da informação/documentação relacionada a projetos de TI	☐	☐	☐	☐	☐
Representação e organização da informação em projetos de TI	☐	☐	☐	☐	☐
Concepção e desenvolvimento de serviços de informação em projetos de TI	☐	☐	☐	☐	☐
Competências pedagógicas, de ensino e orientação relacionadas a projetos de TI	☐	☐	☐	☐	☐

Capacidade de

Capacidade de					
Comunicação efetiva em					
projetos de TI					
Visão atualizada das					
Missão, visão e valores					
destacadas tecnológicas					
gestão de conteúdo em					
projetos de TI					

17. Das habilidades a seguir, marque aquelas requeridas e/ou executadas por você em seu trabalho como formado em Computação? (marque quantas alternativas desejar) *

Marque todas que se aplicam.

- ☐ Conhecer, refletir e aplicar teorias e paradigmas da computação
- ☐ Identificar, localizar e disponibilizar para seu cliente informações em diversos suportes
- ☐ Identificar e explorar fontes de informação relacionadas a projetos de TI
- ☐ Projetar sistemas de informática
- ☐ Diagnosticar sistemas de informática, identificar problemas e projetar soluções
- ☐ Avaliar a qualidade das fontes de informação, sob os seguintes parâmetros: exatidão, atualidade, abrangência, formatos disponíveis e orientada à necessidade do cliente
- ☐ Adicionar valor ao processo de coleta de dados em projetos de TI
- ☐ Focar os parâmetros de qualidade do cliente em projetos de TI
- ☐ Antecipar as demandas de informação e comunicação em projetos de TI
- ☐ Organizar e sistematizar os dados úteis de cada cliente em projetos de TI, utilizando-se dos processos de análise, interpretação e representação da informação. Coletar e conectar dados dispersos de modo a originar novos dados, informações e conhecimentos em projetos de TI
- ☐ Utilizar a tecnologia como vetor para conectar pessoas, organizações, documentos e informações em projetos de TI
- ☐ Atualmente não estou atuando profissionalmente na área de Ciência da Computação
- ☐ Não se aplica, pois no momento não estou trabalhando

Informações em bases de dados públicas

Link de suas contas com dados profissionais e acadêmicos.

18. URL do perfil no LinkedIn

19. URL do currículo Lattes

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

Apêndice C

Segunda enquete

Coleta de Dados Profissionais dos Egressos do Curso de Graduação de Ciência da Computação da UFCG

De que se trata?

Sou estudante de mestrado em Ciência da Computação na Universidade Federal de Campina Grande (UFCG) e estou realizando um estudo sobre a inserção profissional dos egressos do curso de graduação em Ciência da Computação da UFCG.

Este formulário visa coletar dados essenciais para analisar essa inserção. Sua participação é muito importante e contribuirá significativamente para a pesquisa.

A pesquisa não oferece benefícios diretos aos participantes. As informações fornecidas serão utilizadas apenas para fins científicos relacionados a esta pesquisa. Não há custos associados a esta pesquisa.

Toda informação derivada da sua participação será tratada com estrita confidencialidade, de forma a salvaguardar o anonimato do respondente. Da mesma forma, tanto na análise como na publicação e divulgação científica dos resultados, não será identificada a identidade de nenhum dos participantes e os dados serão apresentados sempre de forma agregada.

A participação nesta pesquisa é absolutamente voluntária e você poderá desistir do estudo a qualquer momento, sem quaisquer consequências.

Agradeço desde já pela sua colaboração!

Instruções

Leia cada pergunta atentamente antes de responder.

Responda de forma honesta e precisa.

Não há respostas certas ou erradas, buscamos compreender suas opiniões e experiências.

O tempo estimado para preenchimento do formulário varia de 3 a 6 minutos.

Informações para contato

Programa de Pós-Graduação em Ciência da Computação (PPGCC) da UFCG

Mestrando: Felisberto Cláudio dos Santos

Orientador: Francisco Vilar Brasileiro (Fubica)

Em caso de dúvidas ou para mais informações sobre esta pesquisa, entre em contato pelo e-mail felisberto@copin.ufcg.edu.br

*** Indica uma pergunta obrigatória**

1. Nome completo *

2. **Você atua ou já atuou na área de Computação? ***

Marcar apenas uma oval.

☐ Sim

☐ Não *Pular para a pergunta 8*

Informações sobre a PRIMEIRA atuação após a Graduação

3. **Quantos meses se passaram entre a sua graduação e o engajamento em sua PRIMEIRA atividade profissional na área de Computação? ***

Marcar apenas uma oval.

☐ 0 a 6 meses

☐ 7 a 12 meses

☐ 13 a 24 meses

☐ Mais de 24 meses

4. **Qual o nome do cargo? ***

5. **Qual a modalidade de trabalho? ***

Marcar apenas uma oval.

☐ Presencial

☐ Remota

☐ Híbrida

6. Qual o tipo de engajamento? *

Marcar apenas uma oval.

- ☐ Colaborador de empresa
- ☐ Sócio de empresa
- ☐ Autônomo *Pular para a pergunta 9*
- ☐ Outro: _____

Informação sobre o porte de empresa (primeira atuação após a graduação).

7. Qual o porte da empresa onde se deu sua PRIMEIRA atuação na área? *

Marcar apenas uma oval.

- ☐ Microempresa (até 20 colaboradores)
- ☐ Empresa de pequeno porte (até 100 colaboradores e faturamento anual maior que R\$ 360 mil)
- ☐ Empresa de médio porte (mais de 100 colaboradores e faturamento anual menor que R\$ 20 milhões)
- ☐ Empresa de grande porte (mais de 100 colaboradores e faturamento anual maior que R\$ 20 milhões)

Pular para a pergunta 9

Informações sobre atuação fora da área de Computação

8. Descreva o porquê de você nunca ter atuado na área de Computação? *

Atuação ATUAL

9. Qual a sua situação ATUAL, no que diz respeito a sua atuação? *

Marcar apenas uma oval.

- ☐ Não estou trabalhando, nem estudando *Pular para a pergunta 17*
- ☐ Estou somente estudando *Pular para a pergunta 10*
- ☐ Estou trabalhando *Pular para a pergunta 12*

Perguntas sobre o tipo de estudo**10. Que tipo de curso você está fazendo? ***

Marcar apenas uma oval.

- ☐ Cursos de atualização online ou presenciais
- ☐ Especialização ou MBA
- ☐ Mestrado
- ☐ Doutorado
- ☐ Outro: _____

11. Qual a área do curso? *

Marcar apenas uma oval.

- ☐ Computação
- ☐ Correlata a Computação
- ☐ Não correlata a Computação

Pular para a pergunta 17

Atuação profissional ATUAL

12. Sua atuação ATUAL é na área de Computação? *

Marcar apenas uma oval.

☐ Sim

☐ Não

13. Qual é o nome do cargo? *

14. Qual a modalidade de trabalho? *

Marcar apenas uma oval.

☐ Presencial

☐ Remota

☐ Híbrida

15. Qual o tipo de engajamento? *

Marcar apenas uma oval.

☐ Colaborador de empresa

☐ Sócio de empresa

☐ Autônomo *Pular para a pergunta 17*

☐ Outro: _____

Informação sobre o porte de sua empresa (atuação ATUAL).

16. Qual o porte da empresa? *

Marcar apenas uma oval.

- ☐ Microempresa (até 20 colaboradores) *Pular para a pergunta 17*
- ☐ Empresa de pequeno porte (até 100 colaboradores e faturamento anual maior que R\$ 360 mil) *Pular para a pergunta 17*
- ☐ Empresa de médio porte (mais de 100 colaboradores e faturamento anual menor que R\$ 20 milhões) *Pular para a pergunta 17*
- ☐ Empresa de grande porte (mais de 100 colaboradores e faturamento anual maior que R\$ 20 milhões) *Pular para a pergunta 17*

Informações em bases de dados públicas

[Link de sítios web com dados profissionais.](#)

17. Você tem um perfil no LinkedIn? *

Marcar apenas uma oval.

- ☐ Sim
- ☐ Não *Pular para a pergunta 20*

Sobre o LinkedIn.**18. Qual o link do seu perfil no LinkedIn? ***

19. Você mantém esse perfil atualizado? *

Marcar apenas uma oval.

- ☐ Sim
- ☐ Não

Informações em bases de dados públicas - Lattes

20. Você tem seu currículo cadastrado na Plataforma Lattes? *

Marcar apenas uma oval.

☐ Sim

☐ Não *Pular para a pergunta 23*

Sobre a plataforma Lattes.

21. Qual o link do seu currículo na Plataforma Lattes?

22. Você mantém esse currículo atualizado? *

Marcar apenas uma oval.

☐ Sim

☐ Não

Outras Bases de Dados Públicas

Bases de dados diferentes da plataforma LinkedIn e Lattes.

23. Você usa outra Base de Dados Pública para cadastrar suas informações profissionais? *

Marcar apenas uma oval.

☐ Sim

☐ Não *Pular para a pergunta 26*

Outras Bases de Dados Públicas

Por favor preencha os campo abaixo.

24. **Qual o link do seu currículo na plataforma usual a sua preferência? ***

25. **Você mantém esse currículo atualizado? ***

Marcar apenas uma oval.

☐ Sim

☐ Não

Grau de satisfação com o curso da UFCG

26. Grau de satisfação dos egressos em relação à sua formação no curso de graduação em Ciência da Computação da UFCG. *

Marcar apenas uma oval.

☐ Muito insatisfeito

☐ Insatisfeito

☐ Regular

☐ Satisfeito

☐ Muito satisfeito

27. Grau de satisfação dos egressos em relação à sua carreira. *

Marcar apenas uma oval.

☐ Muito insatisfeito

☐ Insatisfeito

☐ Regular

☐ Satisfeito

☐ Muito satisfeito

28. **Que disciplinas optativas que você cursou tiveram maior impacto na sua atuação na área de Computação?**

Marque todas que se aplicam.

- ☐ ADM. DE SIST. GERENC. DE BANCO DE DADOS
- ☐ ADMINISTRAÇÃO DE SISTEMAS
- ☐ ALGORITMOS AVANÇADOS I
- ☐ ALGORITMOS AVANÇADOS II
- ☐ ALGORITMOS AVANÇADOS III
- ☐ ALGORITMOS AVANÇADOS IV
- ☐ ARQUITETURA DE SOFTWARE
- ☐ AVAL. DE DESEMPENHO DE SISTEMAS DISCRETOS
- ☐ BANCO DE DADOS II
- ☐ CIÊNCIA DE DADOS DESCRITIVA
- ☐ CIÊNCIA DE DADOS PREDITIVA
- ☐ COMPUTAÇÃO E MÚSICA
- ☐ DES DE SOFTWARE INTEG À OPER DA INFRAEST
- ☐ DESENV. DE APLICACOES CORP. AVANCADAS
- ☐ ECONOMIA DE TECNOLOGIA DA INFORMAÇÃO
- ☐ EMPREENDEDORISMO EM SOFTWARE
- ☐ GERÊNCIA DA INFORMAÇÃO
- ☐ GERÊNCIA DE REDES
- ☐ GESTÃO DE PROJETOS
- ☐ GOVERNANÇA DA INTERNET
- ☐ HABILIDADES SOCIOEMOCIONAIS I
- ☐ HABILIDADES SOCIOEMOCIONAIS II
- ☐ INTERCONEXÃO DE REDES DE COMPUTADORES
- ☐ INTERFACE HOMEM-MÁQUINA
- ☐ JOGOS DIGITAIS
- ☐ MÉTODOS E SOFTWARE NUMÉRICOS
- ☐ MÉTODOS FORMAIS
- ☐ PERCEPÇÃO COMPUTACIONAL
- ☐ PRINCÍPIOS DE DESENVOLVIMENTO WEB
- ☐ PROCESSAMENTO DE LINGUAGEM NATURAL
- ☐ PROGRAMAÇÃO EM BANCO DE DADOS
- ☐ PROGRAMAÇÃO III
- ☐ PROGRAMAÇÃO PARA WEB
- ☐ PROGRAMACAO FUNCIONAL
- ☐ PROJETO DE REDES DE COMPUTADORES
- ☐ PROJETO DE SISTEMAS OPERACIONAIS

- ☐ PROVIS. E OPER. DE INFRAESTRUTURAS
- ☐ RECONHEC. DE PADRÕES E REDES NEURAIAS
- ☐ RECUP. DE INFORMAÇÃO E BUSCA NA WEB
- ☐ SEGURANÇA DE SISTEMAS
- ☐ SISTEMAS DE APOIO À DECISÃO
- ☐ SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS
- ☐ SISTEMAS DISTRIBUÍDOS
- ☐ TRANSFORMAÇÃO DIGITAL
- ☐ VERIFICAÇÃO E VALIDAÇÃO DE SOFTWARE
- ☐ VISÃO COMPUTACIONAL
- ☐ VISUALIZAÇÃO DE DADOS

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

