



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS QUIXADÁ**  
**BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

**IAGO VICTOR GOMES HOLANDA**

**ANÁLISE E DESCOBERTA DO CONHECIMENTO DO PERFIL PROFISSIONAL  
DOS EGRESSOS DA UFC-QUIXADÁ**

**QUIXADÁ**  
**2018**

IAGO VICTOR GOMES HOLANDA

ANÁLISE E DESCOBERTA DO CONHECIMENTO DO PERFIL PROFISSIONAL DOS  
EGRESSOS DA UFC-QUIXADÁ

Monografia apresentada no curso de Sistemas de Informação da Universidade Federal do Ceará, como requisito parcial à obtenção do título de bacharel em Sistemas de Informação. Área de concentração: Computação.

Orientador: Prof. Dr. Antônio Joel Ramiro de Castro

Co-Orientador: Leonardo Torres Marques

QUIXADÁ

2018

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- H669a Holanda, Iago Victor Gomes.  
Análise e descoberta do conhecimento do perfil profissional dos egressos da UFC- Quixadá / Iago Victor Gomes Holanda. – 2018.  
49 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Sistemas de Informação, Quixadá, 2018.  
Orientação: Prof. Dr. Antônio Joel Ramiro de Castro.  
Coorientação: Prof. Leonardo Torres Marques.
1. Mercado de trabalho - Ceará. 2. Perfil do egresso.. 3. Mineração de dados (Computação). I. Título.  
CDD 005
-

IAGO VICTOR GOMES HOLANDA

ANÁLISE E DESCOBERTA DO CONHECIMENTO DO PERFIL PROFISSIONAL DOS  
EGRESSOS DA UFC-QUIXADÁ

Monografia apresentada no curso de Sistemas de Informação da Universidade Federal do Ceará, como requisito parcial à obtenção do título de bacharel em Sistemas de Informação. Área de concentração: Computação.

Aprovada em: \_\_\_\_ / \_\_\_\_ / \_\_\_\_\_

BANCA EXAMINADORA

---

Prof. Dr. Antônio Joel Ramiro de Castro (Orientador)  
Campus Quixadá  
Universidade Federal do Ceará – UFC

---

Leonardo Torres Marques (Co-Orientador)  
Campus Mossoró  
Universidade Federal do Semi-Árido - UFERSA

---

Prof. Dr. Francisco Holanda Soares Junior  
Campus Limoeiro do Norte  
Instituto Federal do Ceará - IFCE

---

Prof. Dr. Jefferson de Carvalho Silva  
Campus Quixadá  
Universidade Federal do Ceará - UFC

A Deus e a minha família por todo o apoio durante a graduação, em especial meus pais Charles Holanda Cavalcante e Maria Erinalva Gomes.

## **AGRADECIMENTOS**

À Deus, pela honra e glória da vida e por me dar a oportunidade de viver a realização do sonho de me formar no curso que sempre sonhei.

Ao professor Antônio Joel Ramiro de Castro, pela orientação. Ao meu amigo e futuro professor Leonardo Torres Marques por toda a ajuda durante minha jornada acadêmica.

Aos professores participantes da banca Prof. Dr. Francisco Holanda Soares Junior e Prof. Dr. Jefferson de Carvalho Silva, pelas orientações, sugestões e colaborações fundamentais para pesquisa.

Aos meus pais, que sempre lutaram para me dar condições de poder alcançar este grande sonho.

Aos meus amigos de graduação, por toda a parceria e apoio durante a graduação.

“A persistência é o caminho do êxito.”

(Charles Chaplin)

## RESUMO

No cenário atual do mercado de trabalho, a tecnologia da informação (TI) tem exercido um papel importante no contexto empresarial. As empresas vêm buscando, um perfil de profissionais de TI qualificados, com versatilidade e capacidade para suprir as necessidades das empresas. Nesse contexto, o objetivo deste trabalho foi conhecer o perfil profissional dos egressos da Universidade Federal do Ceará (UFC) de Quixadá-CE, perante o mercado de trabalho. O estudo foi realizado com a utilização da técnica de descoberta de conhecimento (KDD), com análises de clusterização e descritiva, com o auxílio da ferramenta *Weka*. Os dados para o trabalho, foram levantados por meio de um questionário aplicado em uma população de 249 egressos e obtendo-se 118 respostas, aproximadamente 47,4% das respostas possíveis. A partir das análises, observou-se uma forte tendência de que os formados em Sistemas de Informação tenham melhores salários, mais facilidade para ingressar no mercado de trabalho e melhores planos de carreiras, em detrimento dos formados no curso de Engenharia de Software. Contudo, não foi possível mais análises por conta da pouca quantidade de dados obtidos na pesquisa e a pouca exploração de algoritmos e ferramentas de mineração de dados (MD) para tratar do assunto.

**Palavras chaves:** Mercado de Trabalho - Ceará. Perfil do egresso. Mineração de dados (Computação).

## ABSTRACT

In the current context of the job market, information technology (IT) has played an important role in the business context. Companies are looking for a profile of qualified IT professionals with versatility and capacity to meet the needs of the companies. In this context, the goal of this work was to know the professional profile of the graduates of the Federal University of Ceará (UFC) of Quixadá before the job market. The study was carried out using the Knowledge Discovery Technique (KDD), with clustering and descriptive analyzes, with the help of the Weka tool. The data for the study was collected using a questionnaire applied to a population of 249 graduates and obtaining 118 responses, approximately 47.4% of the possible answers. From the analyzes, a strong tendency was observed that those graduated in Information Systems tend to have better salaries, had more facility to enter the job market, had better career plans, in relation to those graduated in the Software Engineering course. However, no further analysis was possible because of the low amount of data obtained in the research and the little exploitation of algorithms and data mining (DM) tools to deal with the subject.

**Keywords:** Job Market - Ceará. Egress profile. Data Mining (Computing).

## LISTA DE FIGURAS

Figura 1 – Processo de KDD . . . . .	18
Figura 2 – Tela inicial da ferramenta <i>Weka</i> . . . . .	22
Figura 3 – Tela de exploração do <i>Weka</i> . . . . .	23
Figura 4 – Egressos - UFC Quixadá . . . . .	26
Figura 5 – Gênero dos egressos . . . . .	31
Figura 6 – Idade dos egressos . . . . .	32
Figura 7 – Naturalidade dos egressos . . . . .	32
Figura 8 – Cidade atual onde reside e trabalha . . . . .	33
Figura 9 – Formação acadêmica dos egressos . . . . .	34
Figura 10 – Ano de formação dos egressos . . . . .	34
Figura 11 – Atual nível de formação dos egressos . . . . .	35
Figura 12 – Idiomas fluentes . . . . .	36
Figura 13 – Atual situação empregatícia dos egressos . . . . .	36
Figura 14 – Dificuldades para inserção no mercado de trabalho . . . . .	37
Figura 15 – Tempo para inserção no mercado de trabalho . . . . .	37
Figura 16 – Setor de atuação dos egressos . . . . .	38
Figura 17 – Área de atuação de atuação dos egressos . . . . .	38
Figura 18 – Remuneração dos egressos . . . . .	39
Figura 19 – Nível de satisfação da atual atividade . . . . .	39
Figura 20 – Identificação na área que trabalha . . . . .	40
Figura 21 – Possibilidade de crescimento profissional . . . . .	41
Figura 22 – Resultado geral . . . . .	42

## LISTA DE QUADROS

Quadro 1 – Variáveis selecionadas selecionadas . . . . .	27
Quadro 2 – Variáveis e verificação de problemas . . . . .	28
Quadro 3 – Variáveis do questionário e as respectivas transformações . . . . .	29
Quadro 4 – Antes e depois das alternativas com <i>missing values</i> ou <i>outliers</i> . . . . .	29

## LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
CE	Ceará
CC	Ciência da Computação
ES	Engenharia de Software
GUI	<i>Graphical User Interface</i>
IFF	Instituto Federal Fluminense
IES	Instituição de Ensino Superior
KDD	<i>Knowledge discovery in databases</i>
MD	Mineração de Dados
RC	Redes de Computadores
RS	Rio Grande do Sul
SI	Sistemas de Informação
TI	Tecnologia da Informação
UFC	Universidade Federal do Ceará

## LISTA DE SÍMBOLOS

% Porcentagem

+ Soma

/ Divisão

\* Multiplicação

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
<b>1.1</b>	<b>Objetivos</b>	<b>14</b>
<i>1.1.1</i>	<i>Objetivo Geral</i>	<i>14</i>
<i>1.1.2</i>	<i>Objetivos específicos</i>	<i>14</i>
<b>2</b>	<b>TRABALHOS RELACIONADOS</b>	<b>15</b>
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>17</b>
<b>3.1</b>	<b>Mercado da Tecnologia da Informação</b>	<b>17</b>
<b>3.2</b>	<b>Análise Descritiva</b>	<b>17</b>
<b>3.3</b>	<b>KDD - Knowledge discovery in databases</b>	<b>18</b>
<i>3.3.1</i>	<i>Seleção</i>	<i>19</i>
<i>3.3.2</i>	<i>Pré-Processamento e Limpeza</i>	<i>19</i>
<i>3.3.3</i>	<i>Transformação de Dados</i>	<i>19</i>
<i>3.3.4</i>	<i>Mineração de Dados</i>	<i>20</i>
<i>3.3.5</i>	<i>Avaliação dos Dados</i>	<i>20</i>
<b>3.4</b>	<b>Clusterização</b>	<b>20</b>
<b>3.5</b>	<b>Ferramenta de MD</b>	<b>22</b>
<b>4</b>	<b>MATERIAIS E MÉTODOS</b>	<b>24</b>
<b>4.1</b>	<b>Desenvolvimento e pré-teste do questionário</b>	<b>24</b>
<b>4.2</b>	<b>Aplicação do Questionário</b>	<b>25</b>
<b>4.3</b>	<b>Aplicação do KDD</b>	<b>27</b>
<i>4.3.1</i>	<i>Seleção dos Dados</i>	<i>27</i>
<i>4.3.2</i>	<i>Pré-Processamento e Transformação</i>	<i>28</i>
<i>4.3.3</i>	<i>Mineração dos Dados</i>	<i>30</i>
<i>4.3.4</i>	<i>Avaliação dos Dados</i>	<i>30</i>
<b>5</b>	<b>RESULTADOS</b>	<b>31</b>
<b>5.1</b>	<b>Resultados da Análise Descritiva</b>	<b>31</b>
<b>5.2</b>	<b>Análise de Clusterização</b>	<b>41</b>
<b>6</b>	<b>DISCUSSÃO</b>	<b>43</b>
<b>7</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>44</b>
	<b>REFERÊNCIAS</b>	<b>45</b>
	<b>APÊNDICE A – Perfil dos Egressos da UFC-Quixadá no Mercado de TI</b>	<b>47</b>

## 1 INTRODUÇÃO

Diante das transformações no mercado de trabalho, a TI tem evoluído continuamente, tornando-se essencial no cotidiano das pessoas, empresas e atingindo setores como saúde, logística e educação. Esta evolução, segundo Martins e Oliveira (2017), realça o valor do profissional da TI, que busca contribuir diretamente nas estratégias das organizações, resolvendo obstáculos de forma rápida e eficiente, desenvolvendo assim um perfil propício a ir além das soluções tecnológicas. A busca por profissionais da TI de acordo com Moreno, Cavazotte e Farias (2009), encontra-se em crescente devido a submissão das empresas nas rotinas de negócios apoiados às tecnologias, como também a transição tecnológica e a baixa inserção de profissionais no mercado tecnológico.

Segundo Lima (2017), o mercado de trabalho da TI é visto como uma área estratégica, de grande influência no mundo dos negócios. No entanto, a escassez de profissionais com conhecimento e preparação para atender as demandas das empresas, tornou-se um desafio a ser solucionado. Por conta disso, a busca por mão de obra qualificada passou a ser indispensável e bastante disputada perante as exigências do mercado de trabalho. A evolução da TI proporcionou o crescimento na demanda destes profissionais, retribuindo remunerações diferenciadas e possibilidade da aplicação do profissional em diferentes setores da área da TI. Para os que já se encontram no mercado de trabalho, a necessidade de constante inovação e velocidade na produção, exige que os profissionais estejam sempre alinhados com as últimas tecnologias e competências.

Segundo Zanardi (2014), conhecer o perfil do profissional da TI possibilita futuros direcionamentos de ações estratégicas dos cursos de tecnologia ofertados pelas IES, além de avaliar conhecimentos específicos, inerentes aos profissionais junto as necessidades do mercado de trabalho. A Universidade Federal do Ceará (UFC) de Quixadá-CE, destaca-se por ser um polo tecnológico no qual oferece diferentes cursos de TI, com o propósito de atender as necessidades qualitativas e quantitativas exigidas pelo mercado de trabalho, tendo como principal missão formar profissionais habilitados para aplicar conhecimentos adquiridos na ciência e na tecnologia, oferecendo-lhes a base teórica e prática exigida pelo mercado(UFC, 2008).

Diante desse cenário, a proposta deste trabalho foi identificar o perfil profissional dos egressos da UFC, caracterizando-se em uma pesquisa exploratória e descritiva, constituída de um questionário que busca identificar as respostas dos egressos diante do mercado de trabalho. Para análise dos resultados foram aplicadas as seguintes técnicas: análise descritiva e a técnica de MD

clusterização, no qual o algoritmo de agrupamento utilizado, foi o *K-means* junto a ferramenta *Weka*. Para auxiliar na exploração dos dados foi empregado as etapas do processo de descoberta de conhecimento KDD (*Knowledge discovery in databases*).

Este trabalho encontra-se organizado da seguinte forma: A Seção 2 apresenta os trabalhos relacionados. A seção 3 é apresentado a fundamentação teórica. A seção 4 descreve os materiais e procedimentos do trabalho. A seção 5 descreve os diferentes experimentos realizados e os resultados obtidos. A seção 6 são argumentados as principais comparações e conclusões deste trabalho. A seção 7 contém as considerações, limitações do trabalho e trabalhos futuros.

## **1.1 Objetivos**

### ***1.1.1 Objetivo Geral***

Identificar o perfil profissional dos egressos da Universidade Federal do Ceará Campus Quixadá-CE, a partir da atuação destes no mercado de TI.

### ***1.1.2 Objetivos específicos***

- a) Determinar as oportunidades de mercado para os egressos da UFC Campus Quixadá-CE;
- b) Realizar a mineração de dados para descoberta de conhecimento;
- c) Analisar os resultados obtidos na mineração de dados e traçar o perfil dos egressos.

## 2 TRABALHOS RELACIONADOS

Zanardi (2014) realizou uma pesquisa com relação aos perfis dos profissionais e das empresas de TI existentes no mercado de trabalho na cidade de Frederico Westphalen-RS. De forma, a se obter uma base sólida de conhecimento para construção de projetos e programas, visando a inserção de novos profissionais oriundos das universidades para o mercado de trabalho. Zanardi (2014) expõe que a execução do trabalho ocorreu pela grande demanda dos profissionais e a exigência de conhecimento dos perfis das empresas, direcionando-se diretamente para ações estratégicas dos cursos de TI, ofertados pelas IES da região.

Dessa forma, foi realizado nesse estudo uma pesquisa de caráter descritivo, aplicada por meio de um questionário semiestruturado para os gestores das empresas e outro questionário para os colaboradores. A partir dos dados coletados foi aplicado regras de associação na etapa de mineração de dados, para a identificação de conhecimentos e gerenciamento das informações. Sendo assim, este trabalho assemelha-se a Zanardi (2014), por buscar conhecer o perfil dos profissionais de TI, utilizando-se de um questionário para levantamento dos dados e para análise dos mesmos e além disso utilizando técnicas de MD. Por outro lado, difere-se por utilizar a técnica de clusterização na etapa de MD.

Bernadino e Gonçalves (2016), aplicaram métodos de clusterização e associação na MD, com o propósito de alcançar informações de uma base de dados educacional sobre os alunos de uma IES. Durante a realização das tarefas, realizou-se um processo de limpeza e normalização da base de dados, para a utilização dos algoritmos *K-Means* e *Apriori* auxiliado pela ferramenta *Weka*. Após a execução dos algoritmos, os resultados obtidos foram de fundamental importância para IES, pois foram identificadas informações pertinentes para serem utilizadas como estratégias de crescimento educacional. Dentre os algoritmos utilizados, o *K-Means* obteve retorno mais objetivos, sem variações. Já o algoritmo *Apriori* proporcionou uma melhor compreensão dos resultados por meio de regras. Deste modo, este trabalho assemelha-se a Bernadino e Gonçalves (2016), por utilizar o algoritmo *K-Means* e a ferramenta *Weka*. Como diferencial, este trabalho utilizou-se de um questionário para levantamento dos dados.

Menezes, Hora et al. (2014), realizou um estudo visando analisar os perfis dos alunos e egressos dos cursos técnicos do Instituto Federal Fluminense (IFF) que realizaram ou estavam em período de estágio. Entre os instrumentos utilizados, o *Google Forms* foi usado para desenvolvimento e aplicação do questionário. Para a MD, foi utilizado o algoritmo de classificação C4.5 com auxílio da ferramenta *Weka*. Dos resultados levantados durante a pesquisa,

apresentou-se duas alternativas para os estagiários perante o mercado de trabalho.

Das alternativas apresentadas, a primeira trata-se da realização de todas as atividades de forma correta como prevista. A segunda trata-se de exercer funções que incorpore conhecimento no setor em que se encontra durante o estágio. Os resultados mostram que o estágio, torna-se um caminho oportuno para os jovens no mercado de trabalho. Este trabalho, assemelha-se a Menezes, Hora et al. (2014), por trabalhar com informações referentes aos perfis dos egressos utilizando-se do *Google Forms* para produção e aplicação do questionário da pesquisa, além da ferramenta *Weka* também já mencionada. Este trabalho difere-se de Menezes, Hora et al. (2014) por usar o algoritmo de clusterização *K-Means*, no entanto para análise dos dados.

### **3 FUNDAMENTAÇÃO TEÓRICA**

#### **3.1 Mercado da Tecnologia da Informação**

A concorrência e busca por estabilidade, são questões eminentes no mercado de trabalho, que passa por um momento de transformação em seu âmbito tecnológico. Segundo Schuster (2008), os profissionais da TI devem se preparar com o propósito de aproveitar as melhores oportunidades, visando o crescimento profissional e levando em consideração a qualidade de desenvolvimento, além do aperfeiçoamento diário, buscando construir um novo perfil profissional.

Com a amplificação da tecnologia, a demanda dos cursos de TI cresce cada vez mais, formando profissionais com o propósito de atender a todas as exigências do mercado de trabalho. Segundo Janotík (2016), a situação empregatícia para os profissionais da TI tornou-se favorável, de modo que possam escolher as oportunidades com baixo grau de concorrência. Segundo Nunes (2016), pessoas atuam na TI sem nenhuma formação, buscando atender as necessidades das empresas que sofrem com a busca de profissionais qualificados. O mercado de trabalho, consiste na ideia de encontrar profissionais com capacitação para se desenvolver diante das constantes modificações.

Portanto, é importante e relevante conhecer o perfil profissional dos egressos em TI da UFC de Quixadá-CE, buscando explorar suas experiências e características no mercado de trabalho. O progresso do número de vagas e a queda de disponibilidade segundo Janotík (2016), prosseguirão durante os próximos anos, forçando as empresas a inovar, investindo maiores recursos para contratações, buscando novas abordagens para os candidatos no mercado de trabalho, em especial aqueles profissionais de ampla experiência e qualificação.

#### **3.2 Análise Descritiva**

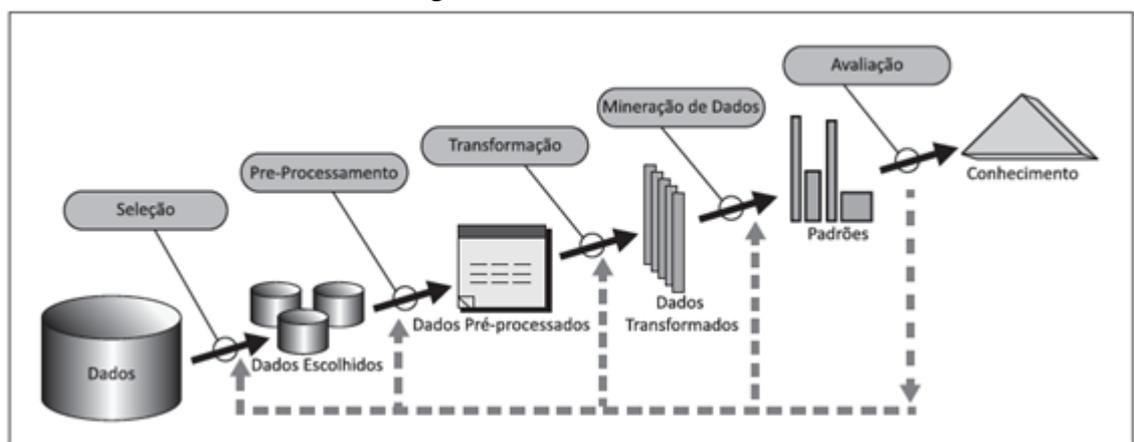
O levantamento de dados encontra-se em crescimento no cenário tecnológico atual, em particular com o surgimento dos computadores e softwares refinados. Quando coletados, estes dados podem ser investigados e analisados de diferentes formas. Segundo Reis e Reis (2002) a análise descritiva destaca-se entre as diferentes técnicas exploradas em estudos estatísticos, na qual tem por propósito de organizar e descrever valores da mesma natureza, permitindo uma visão global das modificações de valores de duas diferentes formas: gráficos e tabelas.

Durante o processo de análise, o início tem a intenção de obter informações que indicam referências que podem ser utilizadas durante a fase final, também conhecido como inferência estatística. Neste trabalho, esta análise retornou resultados para cada variável do questionário, com representações gráficas dando ênfase nas respostas obtidas em cada pergunta identificada.

### 3.3 KDD - Knowledge discovery in databases

O KDD também conhecido como processo de descoberta de conhecimento, tem o propósito de identificar conhecimento útil em dados desconhecidos. O KDD segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), consiste em um processo não trivial, responsável por identificar padrões nos dados de forma a serem válidos, potencialmente úteis e compreensíveis, com a finalidade de melhorar o entendimento de um problema ou auxiliar em uma tomada de decisão. O processo do KDD, tem como característica ser interativo e iterativo, além de cognitivo e exploratório, com passos que devem ser executados de forma sequencial (Figura 1). De modo que, no fim de cada etapa o resultado encontrado contribua para o passo seguinte, podendo sempre que necessário realizar novamente as etapas anteriores. As etapas do KDD são: seleção, pré-processamento e limpeza, transformação de dados, MD, interpretação e avaliação dos dados.

Figura 1 – Processo de KDD



Fonte – Retirado de Fayyad, Piatetsky-Shapiro e Smyth (1996)

### **3.3.1 Seleção**

A seleção dos dados é a primeira etapa no processo de KDD que provoca um impacto significativo no resultado final, pois é nesta fase, que são apurados os conjuntos de dados compreendendo seus atributos e observações que serão utilizados durante a análise. A responsabilidade pelo processo de separação e seleção, fica de acordo com a metodologia do especialista de domínio.

O processo de seleção segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), é bastante complexo, uma vez que os dados podem vir de uma série de fontes diferentes, possuindo diversos formatos, sendo um passo significativo que causa um grande impacto no processo.

### **3.3.2 Pré-Processamento e Limpeza**

No pré-processamento e limpeza, a qualidade dos dados determina a eficiência dos algoritmos utilizados na etapa de MD. Nesta etapa, ocorre a realização de operações básicas nas quais incluem remoção de ruídos, retirada de dados redundantes e inconsistentes, recuperação de dados incompletos e avaliação dos possíveis dados distintos ao conjunto, também conhecidos como (*outliers*).

Nesta etapa, segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), a contribuição do especialista de domínio é fundamental para utilização de métodos de transformação e redução do número de variáveis relacionadas durante o processo para assim, aprimorar o desempenho do algoritmo na análise. No decorrer do processo, também ocorre o tratamento dos dados ausentes (*missing values*), que são encontrados no decorrer do processo, gerando ausência de valores em algumas variáveis. Existe também, os dados discrepantes que possuem valores distintos dos demais registros, que podem ser tratados segundo Ribeiro e Hoppen (2017), de três diferentes formas: exclusão, análise separada, ou utilização de métodos de clusterização. Por fim, os dados derivados consistem, em dados obtidos de combinações e transformações de outros dados.

### **3.3.3 Transformação de Dados**

A etapa transformação, corresponde ao passo precedente a MD. Após a seleção, limpeza e pré-processamento, os dados obtidos são organizados e formatados devidamente, para que possam ser aplicados posteriormente os algoritmos na etapa de MD.

### 3.3.4 *Mineração de Dados*

Desde o surgimento dos sistemas computacionais, um dos principais objetivos das organizações tem consistido no armazenamento de dados. Nas últimas décadas essa tendência ficou ainda mais evidente com a queda nos custos para a aquisição de hardware, tornando possível armazenar quantidades cada vez maiores de dados. Assim, novas e mais complexas estruturas de armazenamento foram desenvolvidas, tais como: banco de dados, *data warehouses*, bibliotecas virtuais (CAMILO; SILVA, 2009).

Diante disso, ferramentas e técnicas foram desenvolvidas para trabalhar com esse volume de dados que crescem diariamente. Nessa perspectiva surge a MD, que consiste em um processo minucioso utilizado na exploração de uma grande quantidade de dados. A MD segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), tem o objetivo de descobrir um padrão consistente de informações, detectando relacionamentos entre suas variáveis e aplicando os padrões encontrados a novos subconjuntos de dados relacionados criados.

### 3.3.5 *Avaliação dos Dados*

Na etapa de avaliação dos dados, regras são determinadas de forma que o conhecimento obtido por essas técnicas de MD sejam alcançadas como objetivo final (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). Durante a interpretação poderá surgir padrões, relacionamentos e descoberta de novos dados aptos a serem utilizados durante a pesquisa. Em casos de resultados obtidos como não satisfatórios, o processo pode ser retornado para qualquer estágio anterior da MD, conforme é verificado na Figura 1.

## 3.4 **Clusterização**

Clusterização de dados ou análise de agrupamentos, trata-se de uma metodologia utilizada na MD, que tem por finalidade agrupar de forma automática e não supervisionada um conjunto de elementos obedecendo alguma similaridade, existindo assim  $n$  casos bases separados em  $k$  grupos ou *clusters*. No geral, as medidas de similaridade mais utilizadas, consiste nas medidas de distancias habituais como por exemplo, a distância euclidiana. Dentre os diferentes tipos de algoritmos de clusterização, neste trabalho, foi utilizado o algoritmo *K-Means* para o processo de MD.

O algoritmo *K-Means* é um algoritmo de clusterização simples e eficiente, no qual

fornece uma classificação baseando-se em análises e comparações dos resultados alcançados, classificando de forma automática sem nenhuma pré-classificação existente. Este algoritmo, caracteriza-se por não utilizar um supervisor para definir possíveis modelos, dependendo diretamente de um parâmetro específico ( $k$ =número de *cluster*), no qual é definido pelo usuário. Por conta desta definição, problemas costumam ser identificados, pois normalmente não se sabe qual o número de *clusters* existem por suposição. A seguir Pichiliani (2017), detalha quais são as etapas presentes no algoritmo *K-Means*.

**1. Estabelece valores para os centroides.**

No início os  $k$  centroides recebem seus valores iniciais. Neste processo, é importante colocar todos os pontos em um centroide para que o algoritmo possa ter seu desenvolvimento iniciado.

**2. Gerar uma matriz de distância entre cada ponto e os centroides.**

No passo seguinte, a distância entre cada ponto e o centroide é calculada. dependendo diretamente do número de  $N$  pontos e  $K$  centroides, calculando assim  $N \times K$ .

**3. Define cada ponto nas classes de acordo com a distância do centroide.**

Após a realização dos cálculos, os pontos são classificados de acordo com a sua distância para os centroides das classes. O centroide que estiver mais próximo do ponto vai agregá-lo, de modo que o ponto pertença a classe representada pelo centroide. O processo é finalizado a partir do momento em que nenhum ponto altera a classe.

**4. Recalcular o centroide de cada grupo.**

Neste passo, as coordenadas dos centroides são refinadas. As classes que possuem mais de um ponto, um novo valor de centroide é calculado fazendo assim, a média de cada atributo pertencente a cada classe.

**5. Realiza a repetição dos passos 2 e 3 até a convergência.**

Como passo final, o algoritmo retorna para o passo 2, seguindo novamente a ordem dos passos, refinando o cálculo das coordenadas dos centroides. Desta forma, finaliza-se a classificação que coloca cada ponto em apenas uma classe, tornando assim, esse algoritmo classificado como (*hard clustering*).

Além disso, o algoritmo *K-Means* possui pontos de vantagem e desvantagem. Como vantagens, este algoritmo destaca-se por ser relativamente escalável e eficaz. Para Linden (2009), o *K-Means* é um algoritmo rápido, com poucas convergências e de configuração estável. A posição inicial do centroide pode variar, permitindo desta forma determinar condições de

dependência. Como desvantagem, antes da aplicação deve-se especificar o número de *cluster* para o algoritmo *K-Means*, de forma que todos os seus objetos ou dados estejam presentes em algum grupo.

### 3.5 Ferramenta de MD

Para auxiliar no processo de MD, a ferramenta escolhida para este trabalho, foi o *Weka* (*Waikato Environment for Knowledge Analysis*). O *Weka* é uma ferramenta composta por uma coleção de algoritmos de aprendizado de máquina, incluindo ferramentas de pré-processamento, utilizadas na etapa de MD. Essa ferramenta foi desenvolvida na Universidade de *Waikato* na Nova Zelândia, onde todo o seu código é escrito em *Java*, encontrando-se disponível na *web*.

Figura 2 – Tela inicial da ferramenta *Weka*

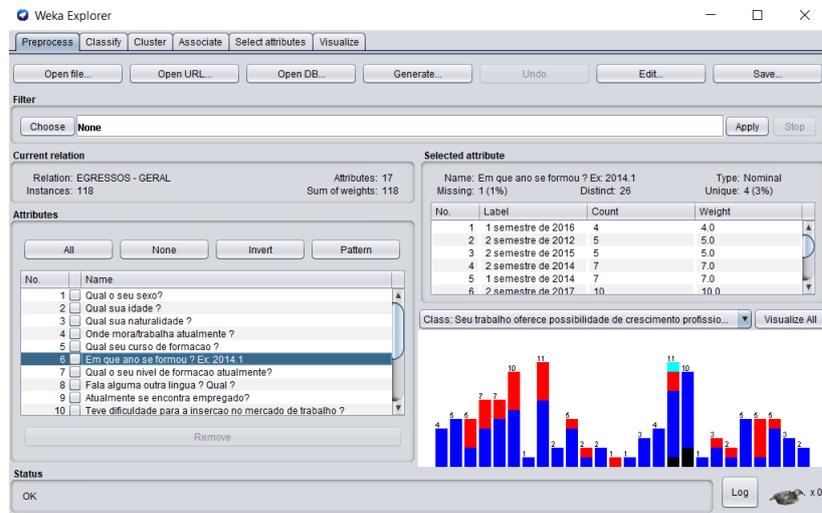


Fonte – Autor

A ferramenta possui uma interface gráfica amigável como mostra a Figura 2 e Figura 3, composta por um conjunto de algoritmos que fornecem relatórios com dados analíticos e estatísticos do domínio minerado. Grande parte de seus recursos, é acessível via, Interface Gráfica do Usuário (GUI) sendo que os demais podem ser utilizados programaticamente por meio de *APIs*. A documentação encontra-se disponibilizada *online*, juntamente ao código fonte. Por ser escrito em *Java*, o código pode ser rodado em diferentes plataformas, conferindo uma boa portabilidade ao *software*. Uma limitação da ferramenta é a escalabilidade, uma vez que as versões atuais limitam o volume de dados a ser manipulado à dimensão de memória principal

(SILVA, 2004).

Figura 3 – Tela de exploração do *Weka*



Fonte – Autor

O *Weka* possui uma interface gráfica que oferece ferramentas para o usuário por meio de janelas e elementos presente na aplicação. As linhas de comandos são utilizadas para melhorar a agilidade nos processos, de forma repetitivas ou para acesso direto de determinada funcionalidade. O acesso por *API*, é normalmente utilizado por desenvolvedores, sendo uma forma mais prática para utilização das funcionalidades já implementadas (SILVA, 2004).

## 4 MATERIAIS E MÉTODOS

Este trabalho trata-se de uma pesquisa de caráter exploratória e descritiva, que teve por finalidade identificar o perfil profissional dos egressos da UFC de Quixadá-CE, analisando áreas de atuações e atividades no mercado de trabalho.

Antes da aplicação do algoritmo de MD, foram necessárias a realização de algumas tarefas de pré-processamento como, transformação, limpeza e integração de variáveis. Em princípio, todos os dados disponíveis foram integrados em um único grupo, identificando-se um número elevado de dados discrepantes *outliers* e dados faltantes *missing values*. Com o propósito de uma melhor organização e adequação dos dados, o Quadro 2 apresenta as perguntas que contêm *outliers*, *missing values* e outros problemas, que conseqüentemente passaram por modificações. Com a intenção de não perder estes dados, o tratamento utilizado foi o método de clusterização, com o algoritmo de aproximação *K-Means*. Segundo Ribeiro e Hoppen (2017), a utilização deste método auxilia no encontro de um novo valor, para correção e atribuição a estes *outliers*. Para os dados faltantes o tratamento utilizado foi definido como "Não Identificado".

Neste trabalho, a metodologia empregada trata-se de uma pesquisa de levantamento, que tem como objetivo produzir estatísticas, isto é, descrições quantitativas e numéricas sobre alguns aspectos de uma população. A principal forma de coleta dos dados é realizada por meio de perguntas, cujas respostas constituem os dados a serem analisados (FOWLER, 2011).

Nas subseções seguintes, são apresentados os passos realizados neste estudo. Desenvolvimento do questionário, aplicado com os egressos da UFC de Quixadá-CE, bem como a realização do pré-teste que tem por finalidade analisar a viabilidade das questões do questionário e os possíveis erros, além de detalhar a forma que foi realizada a aplicação. Em seguida é detalhada a análise descritiva. Por fim, são apresentadas as atividades realizadas para cada etapa do KDD.

### 4.1 Desenvolvimento e pré-teste do questionário

Segundo Gil (2008), o questionário pode ser definido como um instrumento de investigação constituído por uma quantidade de questões que tem o objetivo de adquirir conhecimento de opiniões, crenças, sentimentos, interesses, expectativas dentre outros. Portanto, foi realizada uma pesquisa com os egressos da UFC campus Quixadá-CE, no formato de aplicação de questionário online, na qual a secretaria acadêmica da UFC de Quixadá-CE

disponibilizou o endereço eletrônico dos respectivos egressos da instituição, para a comunicação e realização da pesquisa.

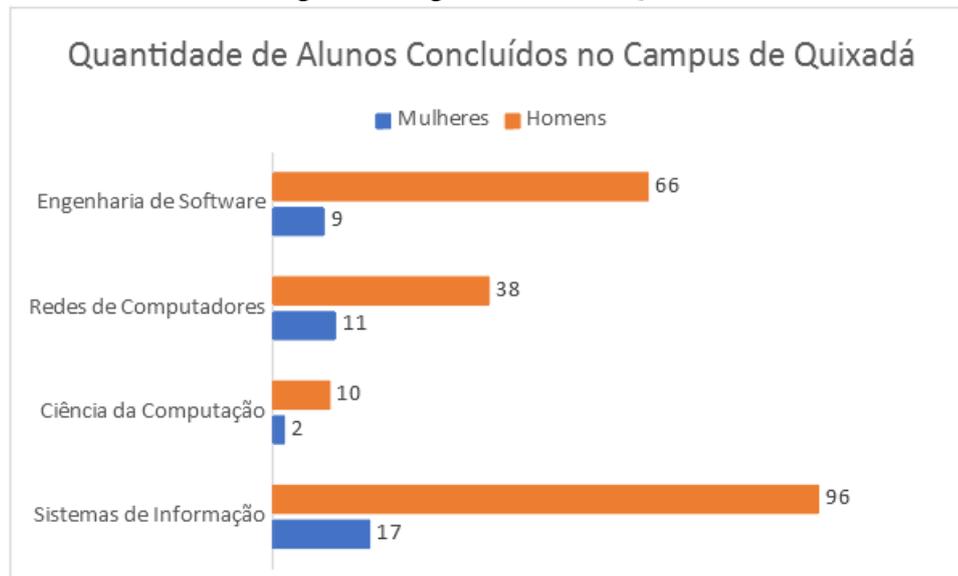
O questionário (Apêndice A) foi elaborado com base no modelo desenvolvido por Ieger e Bridi (2014). Dada a dificuldade de acesso presencial com os egressos da universidade, a utilização de questionário online foi necessária, para adquirir os dados deste trabalho. Portanto, foi utilizada a ferramenta *Google Forms* para envio do questionário, bem como para facilitar o armazenamento e posteriormente exportação da base de dados.

Para Chagas (2000), uma vez desenvolvido, o questionário deve ser aplicado com os potenciais participantes do trabalho com finalidade de pré-teste, como forma de prevê antecipadamente possíveis problemas e dúvidas que podem surgir durante sua aplicação. O pré-teste do questionário neste caso foi realizado com três egressos da UFC de Quixadá-CE, respondido de forma *online*. Durante a aplicação do pré-teste do questionário, os egressos retornaram como *feedbacks*, uma melhor formulação das questões, de forma a contribuir com o melhoramento do questionário, permitindo uma aplicação sem inconsistências e assim atendendo o objetivo do trabalho.

## **4.2 Aplicação do Questionário**

A aplicação do questionário trata-se de uma etapa minuciosa, pois durante a execução deste artefato é de fundamental importância que os entrevistados tenha conhecimento do que se trata a pesquisa. Todas as correções foram realizadas após aplicação do pré-teste, obtendo o questionário final que foi aplicado em uma população de 249 egressos como mostra a Figura 4, do dia 16 de julho á 01 de agosto de 2018.

Figura 4 – Egressos - UFC Quixadá



Fonte – O autor

Os egressos, formaram-se em pelo menos um dos quatro cursos da UFC que são: Sistemas de Informação (SI), Ciência da Computação (CC), Engenharia de Software (ES) e Redes de Computadores (RC). Considerando este universo, de acordo com o cálculo da calculadora amostral de Santos (2014), buscou-se obter um nível de confiança de 95% e erro amostral de aproximadamente 7%, com um valor mínimo de retorno das respostas, correspondente a 28,11% dos egressos (70 egressos). Esse valor foi obtido com base na seguinte fórmula:

$$n = \frac{N \cdot Z^2 \cdot p \cdot (1 - p)}{Z^2 \cdot p \cdot (1 - p) + e^2 \cdot (N - 1)}$$

Equação 1:

n - Amostra calculada

N - População

Z - Variável Normal padronizada associada ao nível de confiança

p - Verdadeira probabilidade do evento

e - Erro amostral

O questionário foi enviado para todos os egressos da UFC de Quixadá-CE, alcançando 118 das 249 respostas possíveis, um valor além das expectativas considerando o cálculo amostral. Para a análise dos dados, foram aplicadas as técnicas de análise descritiva e a clusterização com o algoritmo *K-Means*, utilizando a ferramenta *Weka* para execução da clusterização.

### 4.3 Aplicação do KDD

Neste tópicos são relatadas as etapas do processo de descoberto de conhecimento, seguindo a sequência de passos: Seleção do dados, pré-processamento e limpeza, transformação, MD e interpretação dos dados.

#### 4.3.1 Seleção dos Dados

Durante a seleção foram escolhidos todos dados dos egressos que participaram da pesquisa. Após selecionados, as variáveis “Qual a natureza da organização que trabalha atualmente?” e “Quais tecnologias trabalha atualmente?” foram identificados com problemas como, falta de padronização (*outliers*) e ausência de respostas. Por conta de um grande número de *missing values*, foram realizadas as exclusões destas duas variáveis, finalizando em um total de 17 como pode-se observar no Quadro 1. Assim, estas variáveis selecionadas passaram para as etapas seguintes do processo de descoberta de conhecimento.

Quadro 1 – Variáveis selecionadas selecionadas

Variáveis
Qual o seu gênero?
Qual sua idade?
Qual sua naturalidade?
Onde mora/trabalha atualmente?
Qual seu curso de formação?
Em que ano se formou?
Qual seu nível de formação atual?
Possui fluência em alguma outra língua?
Encontra-se empregado?
Teve dificuldade para a inserção no mercado de trabalho?
Em qual setor atual?
Qual área trabalha atualmente?
Qual sua remuneração?
Qual sua satisfação com sua atual atividade?
Trabalhou ou trabalha em uma área no qual não se identifica?
Quanto tempo levou para se inserir no mercado de trabalho?
Seu trabalho oferece possibilidade de crescimento profissional ou plano de carreira?

Fonte – Autor

### 4.3.2 Pré-Processamento e Transformação

Quadro 2 – Variáveis e verificação de problemas

Variáveis	Outliers	Missing Values	Outros Problemas
<b>Qual o seu gênero?</b>	Não	Não	Não
<b>Qual a sua idade?</b>	Não	Não	Não
<b>Qual a sua naturalidade?</b>	Não	Sim	Não
<b>Onde mora/trabalha atualmente?</b>	Não	Sim	Não
<b>Qual o seu curso de formação?</b>	Não	Não	Não
<b>Qual seu ano de formação?</b>	Sim	Não	Sim
<b>Qual o seu nível de formação atual?</b>	Não	Não	Não
<b>Possui fluência em alguma outra língua?</b>	Sim	Não	Não
<b>Encontra-se empregado?</b>	Não	Não	Não
<b>Teve dificuldade para a inserção no mercado de trabalho?</b>	Não	Não	Não
<b>Em qual setor atual?</b>	Não	Não	Não
<b>Qual área trabalha atualmente?</b>	Não	Não	Não
<b>Qual sua remuneração?</b>	Não	Não	Não
<b>Qual sua satisfação com sua atual atividade?</b>	Não	Não	Não
<b>Trabalhou ou trabalha em uma área no qual não se identifica?</b>	Não	Não	Sim
<b>Quanto tempo levou para se inserir no mercado de trabalho?</b>	Não	Não	Não
<b>Seu trabalho oferece possibilidade de crescimento profissional?</b>	Não	Não	Sim

Fonte – Autor

Todas as variáveis selecionadas do questionário sofreram transformações para uma melhor identificação, de forma a se adequar a um modelo possível de ser manipulado na ferramenta *weka*. O Quadro 3 revela o resultado da transformação e como foram reformuladas, por exemplo, a variável "Qual o seu gênero ?", foi reformulada para "sexo", ou seja, condizendo de forma explícita com o objetivo da questão, e assim todas as outras passaram por esse processo seguindo o modelo de raciocínio do exemplo anterior.

Quadro 3 – Variáveis do questionário e as respectivas transformações

Qual o seu gênero?	<b>sexo</b>
Qual sua idade?	<b>idade</b>
Qual sua naturalidade?	<b>naturalidade</b>
Onde mora/trabalha atualmente?	<b>mora_trabalha</b>
Qual seu curso de formação?	<b>curso</b>
Em que ano se formou?	<b>ano_formacao</b>
Qual seu nível de formação atual?	<b>nível_formacao</b>
Possui fluência em alguma outra língua?	<b>idiomas</b>
Encontra-se empregado?	<b>empregado</b>
Teve dificuldade para a inserção no mercado de trabalho?	<b>dificuldade_inserção</b>
Em qual setor atual?	<b>setor_trabalho</b>
Qual área trabalha atualmente?	<b>area_trabalho</b>
Qual sua remuneração?	<b>remuneracao</b>
Qual sua satisfação com sua atual atividade?	<b>satisfacao_ativ</b>
Trabalhou ou trabalha em uma area no qual não se identifica?	<b>identificacao_area</b>
Quanto tempo levou para se inserir no mercado de trabalho?	<b>insercao_mercado</b>
Seu trabalho oferece possibilidade de crescimento profissional ou plano de carreira?	<b>crescimento_carreira</b>

Fonte – Autor

Quadro 4 – Antes e depois das alternativas com *missing values* ou *outliers*

Variáveis	Respostas dos Egressos	Exemplo da Transformação
<b>naturalidade</b>	Quixadá-CE	Quixadá/CE
<b>mora_trabalha</b>	Quixadá-CE	Quixadá/CE
<b>idiomas</b>	Inglês,Espanhol	Inglês/Espanhol
<b>area_trabalho</b>	Desenvolvimento,Suporte Técnico	Desenvolvimento/Suporte Técnico
<b>ano_formacao</b>	2014.2	2 semestre de 2014

Fonte – Autor

O Quadro 4, apresenta as respostas de algumas variáveis que precisaram ser reformuladas para melhor compreensão das informações, além da padronização de todas as variáveis para um mesmo tipo de dado. A variável naturalidade, recebeu respostas como "Quixadá-CE", no qual após a transformação foi remodelada para "Quixadá/CE", garantindo a integridade da resposta. Na variável "Em que ano se formou?", as respostas tornavam o formato da variável irreconhecível pela ferramenta *weka*. Após transformações, respostas como "2014.2" passaram a ser representadas como "2 semestre de 2014", permitindo que este formato fosse reconhecido com o mesmo tipo de dados das outras variáveis do trabalho que não sofreram transformações, garantindo ainda a integridade da resposta.

### **4.3.3 Mineração dos Dados**

Na fase de MD, os dados foram organizados em um único arquivo para a análise geral. Durante a MD, foi utilizada a ferramenta *Weka*, com o algoritmo *K-Means*, considerando 95% do percentual dos dados. Devido a amostra ser pequena, optou-se pela definição de apenas 2 *clusters* para análise e avaliação dos resultados. O algoritmo *K-Means*, permitiu a associação de cada ponto ao *cluster* mais próximo, redefinindo suas posições conforme a média dos pontos associados durante seu processo. Todos os resultados obtidos são descritos na Seção 5.

### **4.3.4 Avaliação dos Dados**

Nesta etapa, buscou-se analisar as informações obtidas na etapa de MD de forma a responder os objetivos traçados neste estudo, averiguando as diferenças de oportunidades, salários, grau de instrução, além de outras informações. Lembrando que foi considerando um único cenário geral, com a base de dados completa sem que houvesse separação de dados, possibilitando mostrar um contexto geral do perfil dos profissionais egressos da UFC de Quixadá-CE.

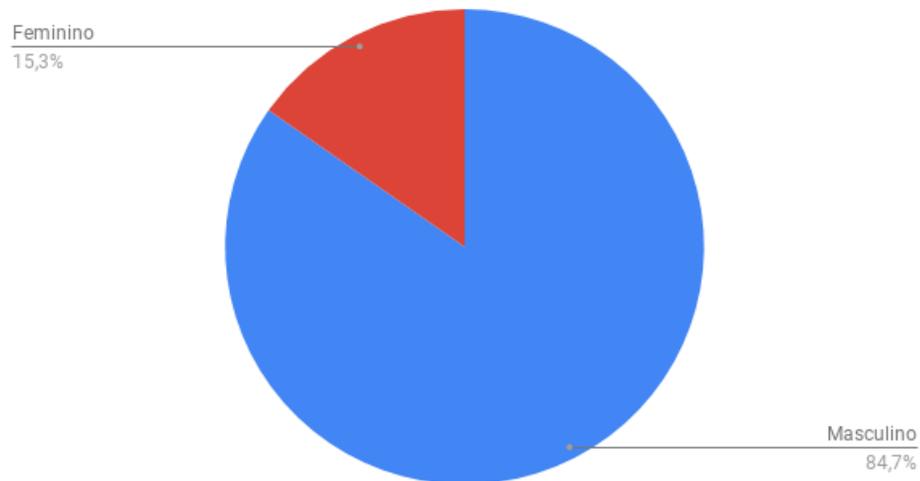
## 5 RESULTADOS

Com o propósito de buscar responder os objetivos deste trabalho, nesta seção são apresentados os resultados da análise descritiva, e da MD, após a aplicação do algoritmo *K-Means*. Destacam-se que os dados utilizados neste estudos, são referentes aos egressos que concluíram seus respectivos cursos até o período de 2017.2.

### 5.1 Resultados da Análise Descritiva

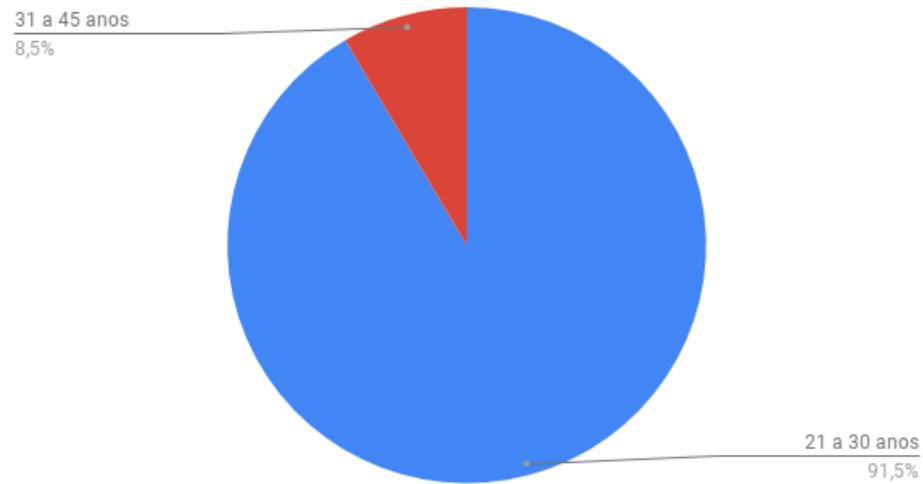
Analisando a Figura 5, verifica-se um detalhamento por sexo dos respondentes, destacando a grande diferença do sexo masculino em detrimento ao sexo feminino, uma vez que aproximadamente 85% são do sexo masculino e apenas cerca de 15% são do sexo oposto. As faixas etárias referentes as idades dos egressos apresenta-se na Figura 6, na qual 91,5% tem entre 21 e 30 anos de idade e apenas 8,5% tem entre 31 e 45 anos. Nenhum dos alunos se encaixaram nos demais intervalos de idade.

Figura 5 – Gênero dos egressos



Fonte: Autor

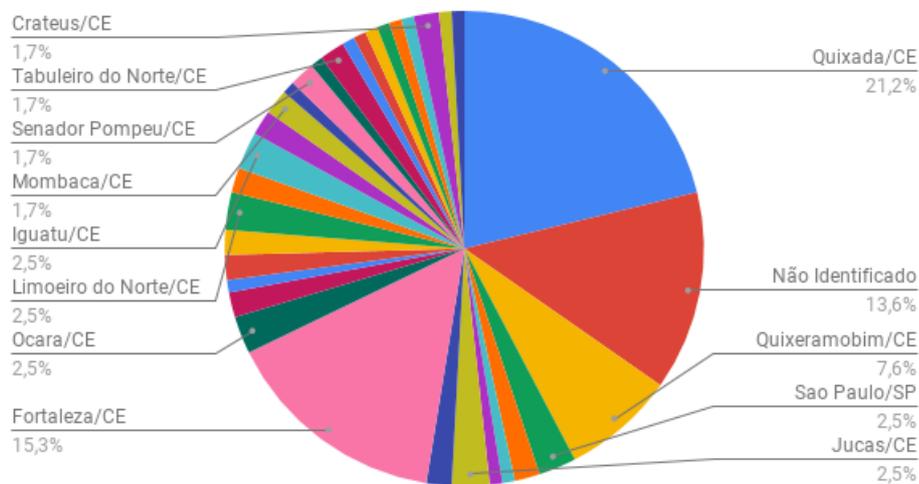
Figura 6 – Idade dos egressos



Fonte: Autor

As cidades de origem dos egressos são retratadas na Figura 7, destacando as cidades de Quixadá-CE e Fortaleza-CE, que foram as duas cidades que tiveram os maiores percentuais, com 21,2% e 15,3%, respectivamente. Fora essas duas cidades e com exceção da cidade de Quixeramobim-CE, que obteve aproximados 8%, os demais municípios citados apresentaram percentuais com menos de 3% de respondentes, e por fim, verifica-se também que 13,6% dos egressos não identificaram suas respectivas origens.

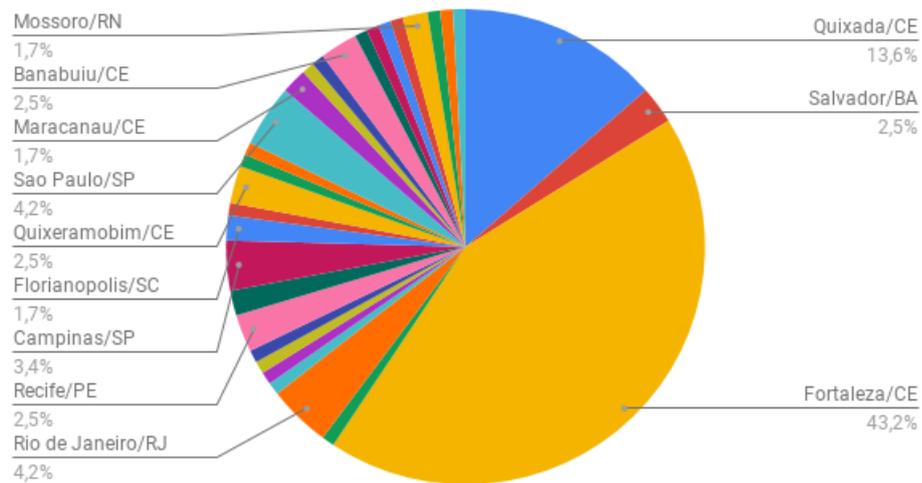
Figura 7 – Naturalidade dos egressos



Fonte: Autor

As cidades nas quais os egressos moram e trabalham atualmente, são destacadas na Figura 8, constatando-se que as cidades de Fortaleza-CE e Quixadá-CE, obtiveram os maiores percentuais com 43,2% e 13,6%, respectivamente. Além dessas duas cidades, as demais mencionadas tiveram percentuais inferiores a 5% de respondentes.

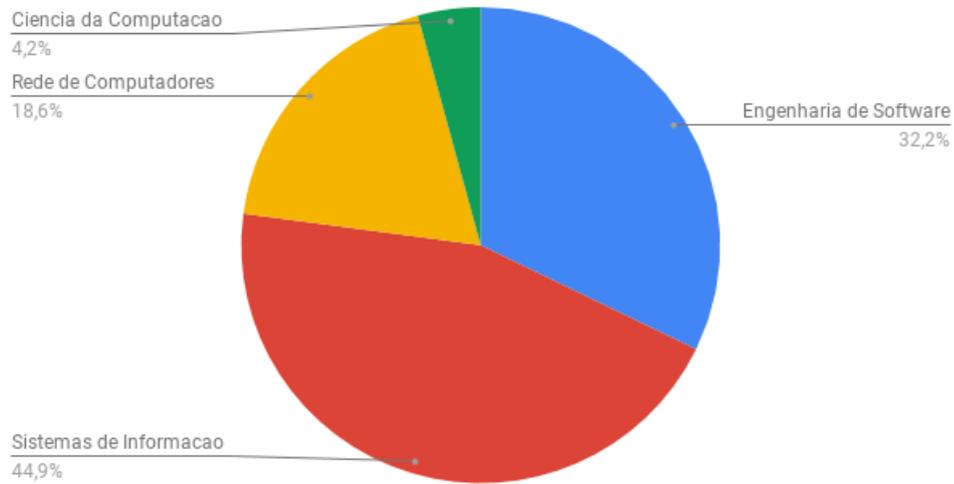
Figura 8 – Cidade atual onde reside e trabalha



Fonte: Autor

Os cursos de formação dos egressos são ressaltados na Figura 9. Observa-se nos resultados que com 44,9% o curso de SI e com 33,2% o curso de ES são os que possuem maior número de egressos, seguido dos cursos de RC com 18,6% e CC com apenas 4,2%.

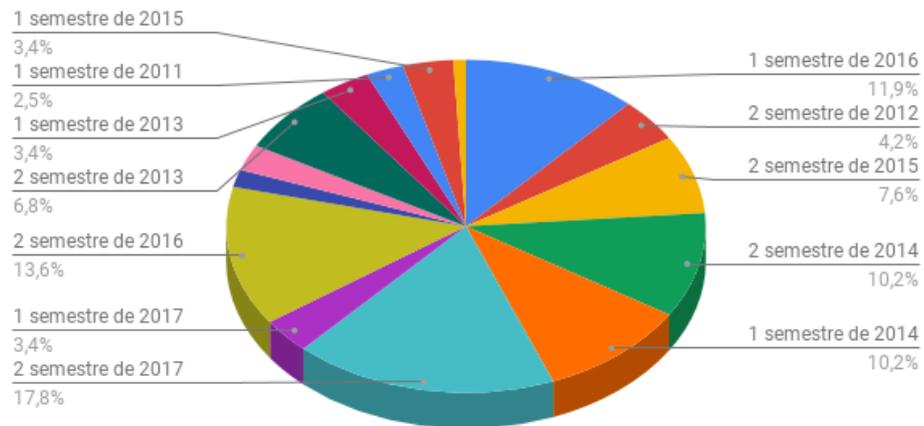
Figura 9 – Formação acadêmica dos egressos



Fonte: Autor

O período de formação dos egressos são detalhados na Figura 10, constatando-se que em 2017.2 e 2016.2, foram os períodos que tiveram os maiores números de concludentes, levando em consideração apenas aqueles que participaram desta pesquisa, com 17,8% e 13,6% respectivamente. Além disso, em 2016.1 o percentual foi de 11,6%, concluindo o ano de 2016 como o de maior número de formandos. No ano de 2014 formaram-se respectivamente 20,4% dos respondentes, 10,2% correspondem a 2014.1 e 10,2% a 2014.2. Os demais períodos tiveram percentuais inferiores a 8% dos respondentes.

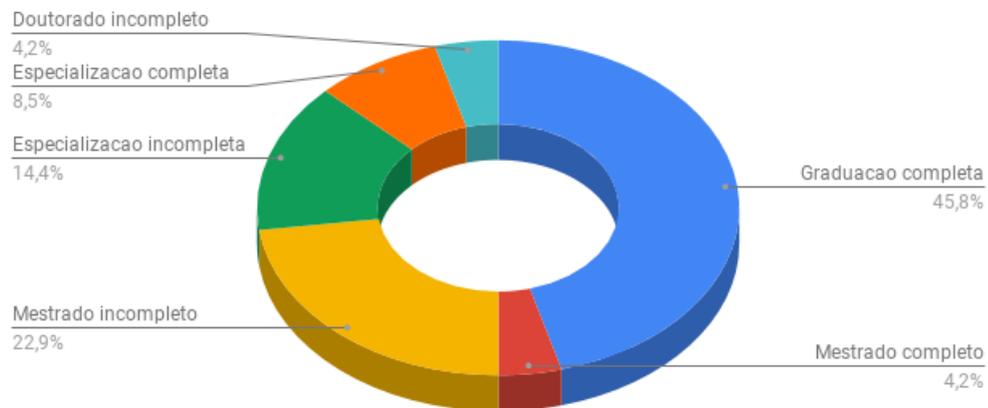
Figura 10 – Ano de formação dos egressos



Fonte: Autor

O atual nível de formação dos egressos são apresentados na Figura 11, identificando que 45,8% possuem apenas graduação completa, 27,1% seguiram para mestrado, no qual 4,2% dos respondentes já concluíram e 22,9% ainda estão em processo de formação. Ressalta-se que 22,9% optaram por uma especialização, visto que apenas 8,5% concluíram suas especializações e 14,4% ainda não finalizaram. Por fim, 4,2% dos egressos revelaram estar com doutorado incompleto.

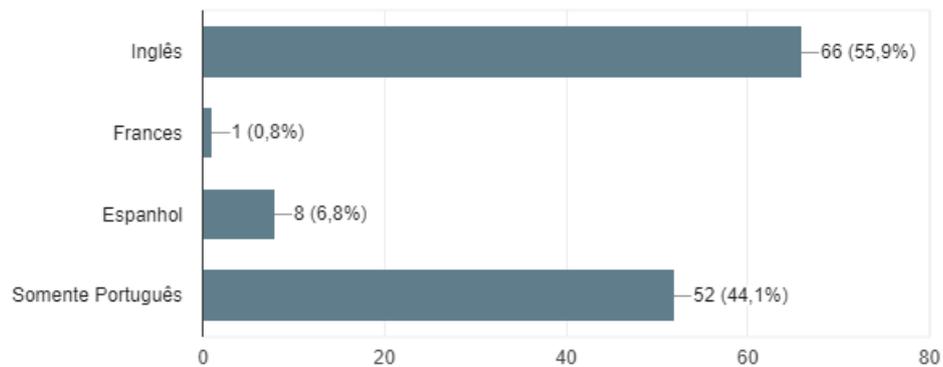
Figura 11 – Atual nível de formação dos egressos



Fonte: Autor

Os idiomas que os egressos possuem fluência são exibidos na Figura 12, no qual 55,9% declararam fluência no inglês e 44,1% em Português. Os demais resultados mencionados atingiram percentuais inferiores a 7%. Por se tratar de uma questão não estruturada alguns dos egressos responderam mais de uma alternativa.

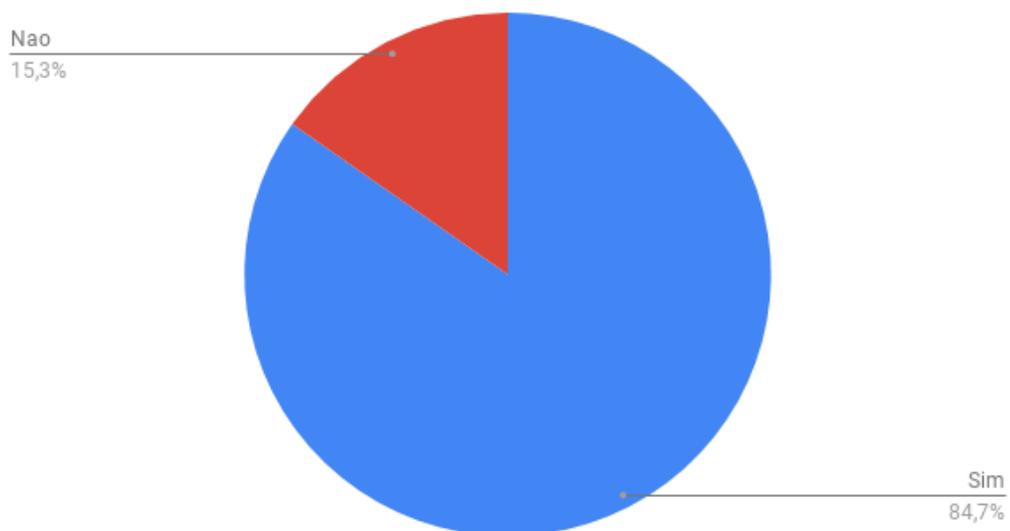
Figura 12 – Idiomas fluentes



Fonte: Autor

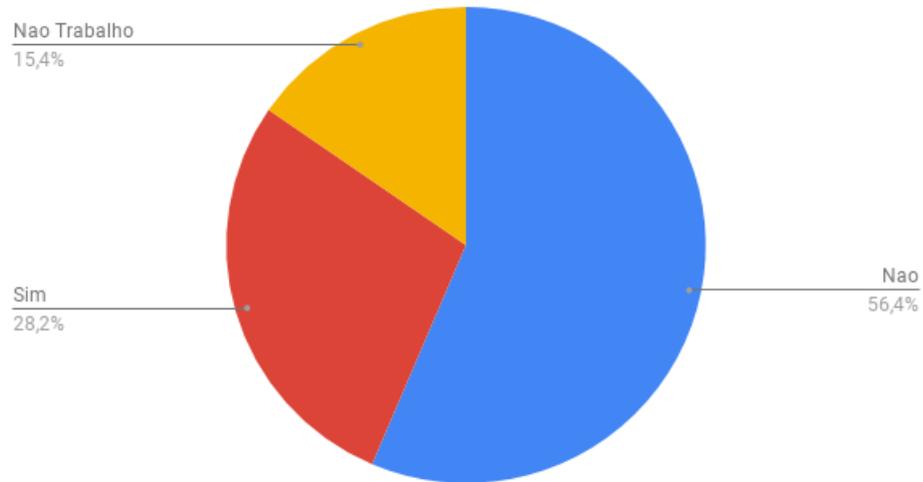
A atual situação empregatícia dos egressos são retratadas na Figura 13. Os resultados mostram que aproximadamente 85% dos respondentes no momento do estudo encontravam-se empregados e apenas 15% estão em situação distinta. As dificuldades encontradas pelos egressos para se inserir no mercado de trabalho, são vistos na Figura 14, constatando que 55,4% não tiveram dificuldades e 28,2% afirmaram o contrário, porém 15,4% dos respondentes declaram não está trabalhando, de modo que não foi possível definir se estes tiveram ou não dificuldades para se inserir no mercado de trabalho.

Figura 13 – Atual situação empregatícia dos egressos



Fonte: Autor

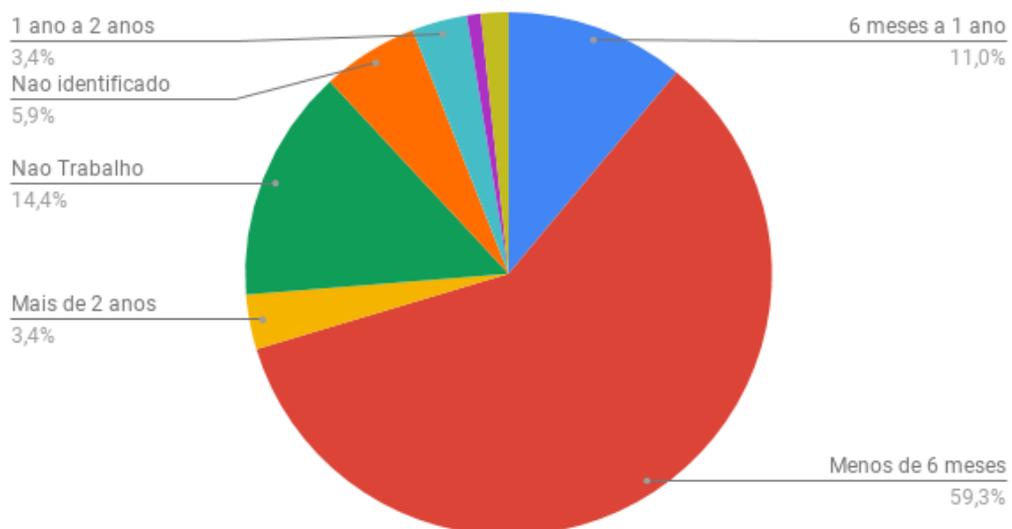
Figura 14 – Dificuldades para inserção no mercado de trabalho



Fonte: Autor

Os intervalos de tempo que os egressos levaram para se inserir no mercado de trabalho são exibidos na Figura 15, de modo que 59,3% levaram menos de 6 meses, por outro lado, 11% levaram de 6 meses a 1 ano. Além disso, 14,4% declaram não está trabalhando no momento do estudo. Os demais resultados mencionadas tiveram percentuais com menos de 6% dos respondentes.

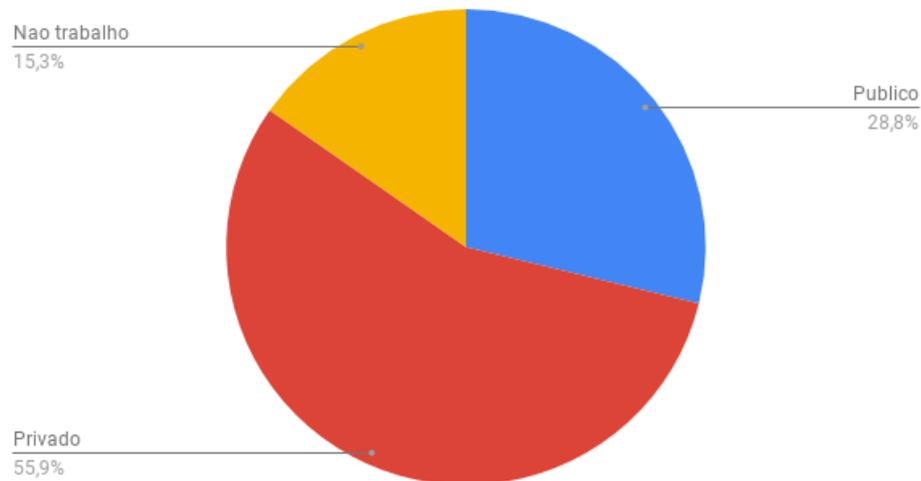
Figura 15 – Tempo para inserção no mercado de trabalho



Fonte: Autor

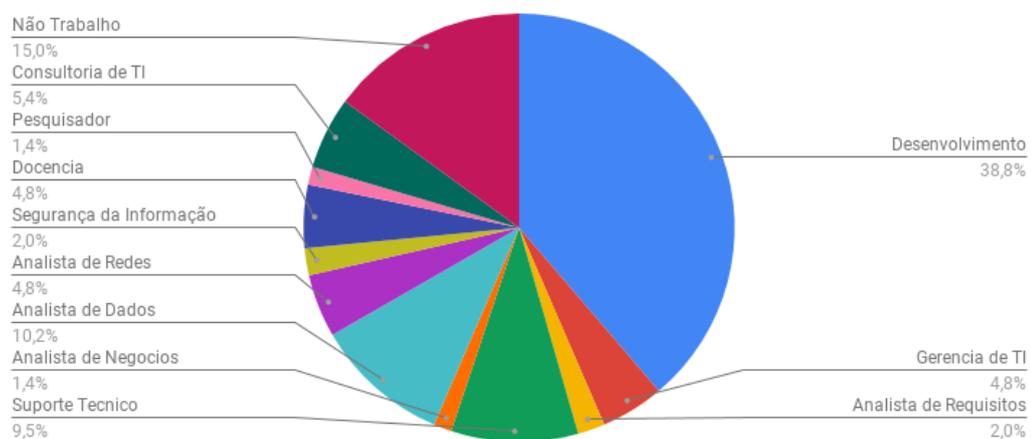
Um balanço dos setores de atuação dos egressos no mercado de trabalho são apresentados na Figura 16, indicando que 55,9% encontra-se atuando no setor privado e 28,8% no setor público, identificando-se também que 15,3% declararam que não trabalham. Por outro lado, as áreas de atuações são denotadas na Figura 17, em que pode-se ver destaques em áreas como Desenvolvimento, Análise de Dados e Suporte Técnico, com 38,8%, 10,2% e 9,5% de respondentes, respectivamente. Além disso, 15% revelaram não está trabalhando no momento do estudo. Por fim, as demais áreas mencionadas tiveram valores inferiores a 6% das respostas obtidas.

Figura 16 – Setor de atuação dos egressos



Fonte: Autor

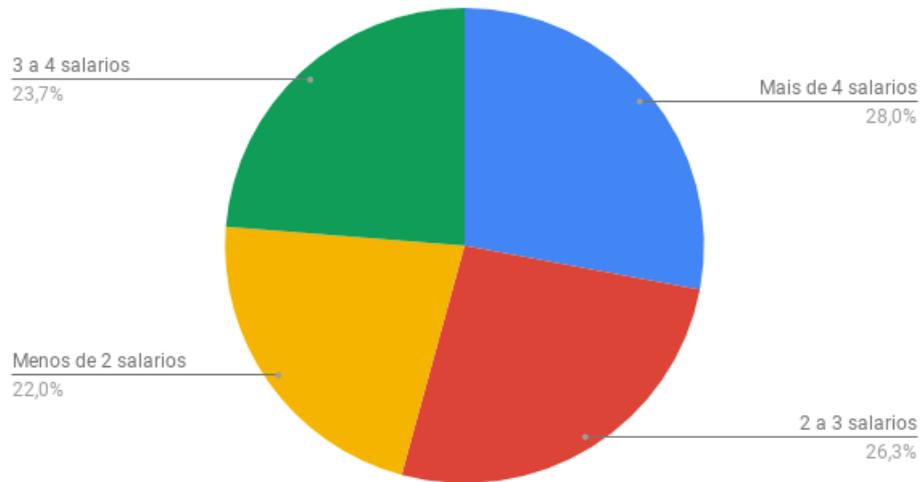
Figura 17 – Área de atuação de atuação dos egressos



Fonte: Autor

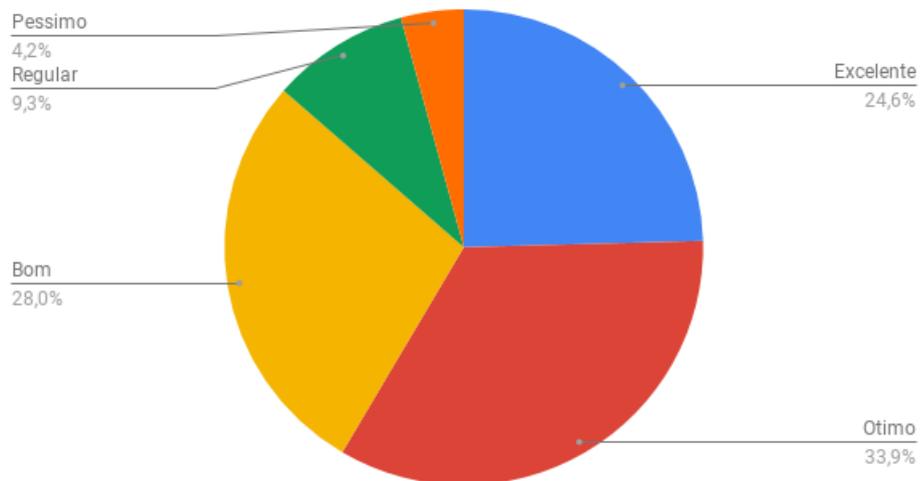
Os resultados referentes às remunerações dos egressos são retratadas na Figura 18, mostrando que 28% ganham mais de 4 salários, 23,7% entre 3 e 4 salários, 26,3% entre 2 e 3 salários, por fim 22,0% com menos de 2 salários. A satisfação dos egressos com suas atuais atividades são refletidos na Figura 19, destacando os níveis de satisfação Ótimo, Bom e Excelente que tiveram os maiores percentuais com 33,9%, 28,0% e 24,6% respectivamente. Por outro lado, apenas 9,3% classificaram como Regular, e 4,2% como Pésimo.

Figura 18 – Remuneração dos egressos



Fonte: Autor

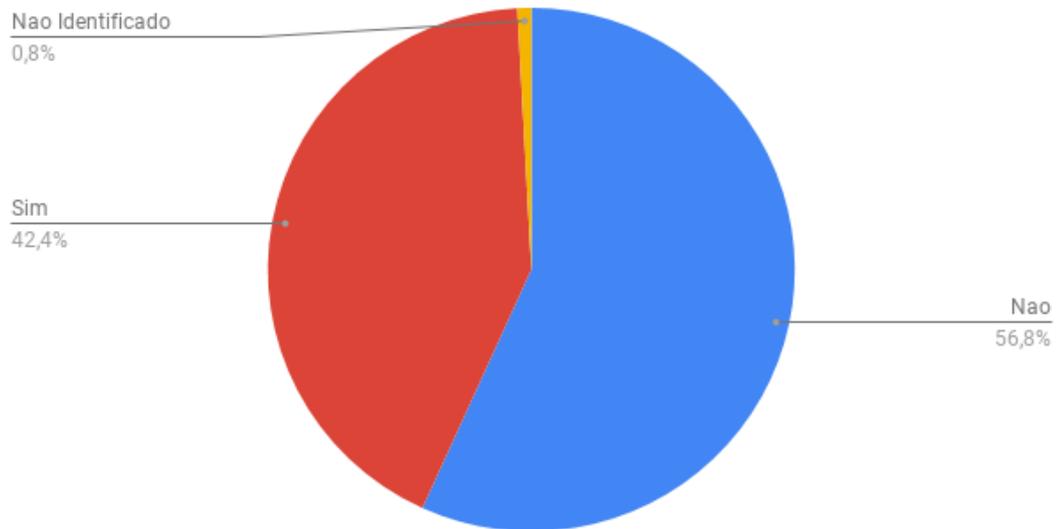
Figura 19 – Nível de satisfação da atual atividade



Fonte: Autor

Os egressos revelaram que já atuaram em algumas áreas nos quais não se identificavam. Os resultados alcançados são detalhados na Figura 20, mostrando que, 56,8% jamais trabalhou em algum setor sem identificação, já 42,4% responderam que sim. Por fim, 0,8% das respostas referentes a esta pergunta não foi possível serem identificadas.

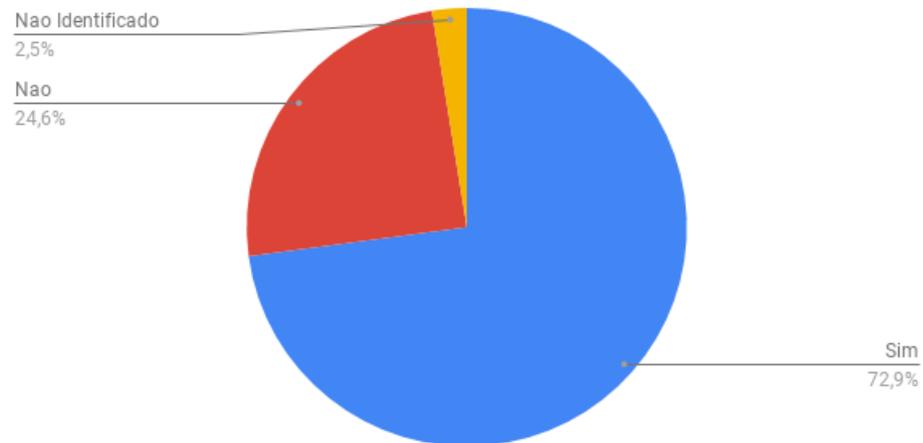
Figura 20 – Identificação na área que trabalha



Fonte: Autor

Por fim, na Figura 21 observa-se os resultados das possibilidades de crescimento dos egressos diante do seu atual trabalho, no qual 72,9% declararam existir possibilidade de desenvolvimento e 24,5% responderam de forma contraria. No fim, 2,5% das respostas referentes a esta pergunta não foram possíveis de serem identificadas.

Figura 21 – Possibilidade de crescimento profissional



Fonte: Autor

## 5.2 Análise de Clusterização

Para representação dos resultados da análise da técnica de clusterização, foram definidos dois *clusters*, levando em consideração 98% do percentual de explicação dos dados definido na ferramenta *Weka*. Na Figura 22 são detalhados a representação dos *clusters* encontrados:

O *Cluster 0*, representa os egressos do sexo masculino com idade entre 21 e 30 anos, naturais da cidade de Quixadá-CE, morando atualmente em Fortaleza-CE, graduados no período de 2016.1 no curso de SI, com apenas graduação completa, possuindo fluência no inglês, encontrando-se no momento da pesquisa, empregado. Além disso, estes egressos não sentiram dificuldades de inserção no mercado de trabalho, levando menos de 6 meses para conseguir o emprego. Atualmente, encontram-se trabalhando no setor privado na área de desenvolvimento, com remuneração superior a 4 salários mensais. Por fim, os egressos declaram-se satisfeitos com a atual atividade, mostrando identificação em sua área de atuação, com garantia de crescimento profissional.

Figura 22 – Resultado geral

Attribute	Cluster#	
	0 (82.0)	1 (33.0)
sexo	Masculino	Masculino
idade	21 a 30 anos	21 a 30 anos
naturalidade	Quixada/CE	Fortaleza/CE
mora_trabalha	Fortaleza/CE	Fortaleza/CE
curso	Sistemas de Informacao	Engenharia de Software
ano_formacao	1 semestre de 2016	2 semestre de 2017
nivel_formacao	Graduacao completa	Graduacao completa
idiomas	Inglês	Inglês
empregado	Sim	Sim
dificuldade_insercao	Nao	Sim
setor_trabalho	Privado	Privado
area_trabalho	Desenvolvimento	Desenvolvimento
remuneracao	Mais de 4 salarios	2 a 3 salarios
satisfacao_ativ	Otimo	Bom
identificacao_area	Sim	Sim
inserir_mercado	Menos de 6 meses	6 meses a 1 ano
crescimento_carreira	Sim	Nao

Fonte: Autor

O *Cluster* 1 representa os egressos do sexo masculino com idade entre 21 e 30 anos, natural e morando atualmente na cidade de Fortaleza-CE, graduados no período de 2016.2 no curso de ES, com apenas graduação completa, com fluência no inglês, encontrando-se no momento da pesquisa, empregado. Diferente do primeiro *cluster*, os egressos encontraram dificuldades para inserção no mercado de trabalho, demorando em média de 6 meses a 1 ano. Atualmente encontram-se trabalhando no setor privado na área de desenvolvimento, com remuneração superior a 4 salários mensais. Por fim, os egressos declaram-se satisfeitos com a atual atividade, mostrando identificação com a área de atuação, mesmo sem garantia de crescimento profissional.

## 6 DISCUSSÃO

Após a descoberta e análise dos perfis, dentre os resultados obtidos, alguns pontos podem ser destacados sobre as variáveis durante a análise descritiva e clusterização. A seguir, são especificados os principais resultados obtidos, em que é realçada algumas diferenças entre os egressos por curso.

- Existe uma forte tendência de quem é formado em SI, ter salários melhores do que os formados em ES.
- Os formados em SI são predominantemente naturais da cidade de Quixadá-CE, em contrapartida os formados em ES são da cidade de Fortaleza-CE.
- Os formados em SI, tem mais facilidade para ingressar no mercado de trabalho, quando comparado com os formados de ES.
- Os formados em SI, se sentem mais satisfeitos nas atuais atividades do que os formados em ES.
- Os formados em SI em seus atuais empregos possuem forte tendência de crescimento profissional, diferindo-se dos formados em ES.

A principal conclusão a cerca das análises, conclui-se que os formados do curso de SI tem melhores oportunidades, conseguem melhores salários, tem mais chances de crescimento em detrimento dos formados dos demais cursos.

## 7 CONSIDERAÇÕES FINAIS

Conhecer a atual situação do mercado de trabalho na área da TI, foi um ponto a ser analisado e utilizado como motivação para o desenvolvimento deste trabalho. Além disso, os resultados alcançados trouxeram informações importantes dos egressos, que podem ser utilizadas pela própria universidade para futuras ações estratégicas.

A partir das análises, foi possível conhecer sobre o perfil profissional dos egressos da UFC de Quixadá-CE. Com a análise descritiva, identificou-se as proporções de algumas características dos egressos e com a técnica de clusterização, foram traçados os perfis referentes aos egressos da UFC.

As limitações encontradas neste trabalho, se deu por conta do baixo número de observações dos egressos, pela utilização de um único algoritmo de MD e uma única ferramenta, o que impossibilitou a comparação de resultados. Como trabalhos futuros, sugere-se a utilização de mais variáveis, incluindo os novos cursos presente na UFC de Quixadá-CE, assim como o desenvolvimento de um questionário com perguntas mais específicas. Outro ponto, seria a avaliação e comparação de outros algoritmos e ferramentas de MD, buscando encontrar aqueles que retornem melhores resultados.

## REFERÊNCIAS

- BERNADINO, G. d. S.; GONÇALVES, A. L. G. Mineração de dados aplicada no contexto educacional. **Tecnologias da informação e comunicação: Pesquisa e Inovação**, p. 8–18, 2016. Disponível em: <<https://publicacoes.rexlab.ufsc.br/index.php/ticpesquisaeinovacao/article/view/48>>. Acesso em: 19 setembro 2018.
- CAMILO, C. O.; SILVA, J. C. d. Mineração de dados: conceitos, tarefas, métodos e ferramentas. **Universidade Federal de Goiás (UFG)**, p. 1–29, 2009.
- CHAGAS, A. T. R. O questionário na pesquisa científica. **Administração on line**, v. 1, n. 1, 2000.
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, p. 37–54, 03 1996.
- FOWLER, F. J. J. **Pesquisa de levantamento-4**. Porto Alegre, Penso Editora, 2011.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. [S.l.]: 6. ed. Editora Atlas SA, 2008.
- IEGER, E. M.; BRIDI, M. A. Profissionais de ti: perfil e o desafio da qualificação permanente. **Revista da ABET**, 2014.
- JANOTÍK, T. How to hire the best talent in the area of information technology. In: IEEE. **Emerging eLearning Technologies and Applications (ICETA), 2016 International Conference on**. [S.l.], 2016. p. 133–139.
- LIMA, J. G. d. **Um estudo sobre o perfil profissional dos egressos do curso de sistemas de informação do centro de informática da UFPE**. 2017. 55 p. Monografia (Bacharel em Sistemas de Informação), UFPE (Universidade Federal de Pernambuco), Recife, Brazil.
- LINDEN, R. Técnicas de agrupamento. **Revista de Sistemas de Informação da FSMA**, v. 4, p. 18–36, 2009.
- MARTINS, B. V.; OLIVEIRA, S. R. d. Qualificação profissional, mercado de trabalho e mobilidade social: cursos superiores de tecnologia. **Sociedade, Contabilidade e Gestão**, v. 12, n. 2, 2017.
- MENEZES, T. de P.; HORA, H. R. M. da et al. Análise do perfil de aluno e egresso de cursos técnicos por meio de data mining: estudo de caso no instituto federal fluminense. **Tear: Revista de Educação, Ciência e Tecnologia**, v. 3, n. 1, 2014.
- MORENO, V.; CAVAZOTTE, F.; FARIAS, E. de. Novos desafios para o profissional de ti: Estudo de caso de uma empresa de prestação de serviços de tecnologia da informação. **Revista de Gestão da Tecnologia e Sistemas de Informação**, v. 6, p. 437–462, 2009.
- NUNES, M. S. M. **A Inserção dos egressos dos cursos de graduação na área de tecnologia da informação dos campi de interior da UFC no mercado regional**. 2016. 114 p. Dissertação (Mestrado Profissional em Políticas Públicas e Gestão da Educação Superior), UFC (Universidade Federal do Ceará), Fortaleza, Brazil.

PICHILIANI, M. **Data mining na prática: algoritmo K-Means**. 2017. Disponível em: <<https://www.devmedia.com.br/data-mining-na-pratica-algoritmo-k-means/4584>>. Acesso em: 24 abril 2018.

REIS, E. A.; REIS, I. A. Análise descritiva de dados. **Síntese numérica estatística**, 2002.

RIBEIRO, W. P.; HOPPEN, J. **Outliers, o que são e como tratá-los em uma análise de dados?** 2017. Disponível em: <<https://www.aquare.la/o-que-sao-outliers-e-como-trata-los-em-uma-analise-de-dados/>>. Acesso em: 03 outubro 2018.

SANTOS, G. E. d. O. **Cálculo amostral: calculadora on-line**. 2014. Disponível em: <<https://www.publicacoesdeturismo.com.br/calculoamostral/>>. Acesso em: 21 Maio 2017.

SCHUSTER, M. E. **Mercado de trabalho de tecnologia da informação: o perfil dos profissionais demandado**. 59 f. Trabalho de Conclusão de Curso (Monografia)–Instituto de Filosofia e Ciências Humanas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008.

SILVA, M. P. d. S. Mineração de dados: conceitos, aplicações e experimentos com weka. **Livro da Escola Regional de Informática Rio de Janeiro-Espírito Santo. Porto Alegre: Sociedade Brasileira de Computação**, v. 1, p. 1–20, 2004.

UFC. **Universidade Federal do Ceara. Projeto Político Pedagógico do Curso Sistemas de Informação**. [S.l.], 2008.

ZANARDI, F. Perfil dos profissionais e das empresas de tecnologia da informação (TI) da cidade de Frederico Westphalen-RS. **Anais do EATI - Encontro Anual de Tecnologia da Informação e Semana Acadêmica de Tecnologia da Informação**, v. 1, p. 389–392, 2014.

**APÊNDICE A – Perfil dos Egressos da UFC-Quixadá no Mercado de TI****1 - Qual seu sexo?**

- A.  Feminino
- B.  Masculino

**2 - Qual sua idade?**

- A.  Menor de 20 anos
- B.  21 a 30 anos
- C.  31 a 45 anos
- D.  Acima de 45 anos

**3 - Qual sua naturalidade?**

R-

**4 - Onde mora/trabalha atualmente?**

R-

**5 - Qual seu curso de formação?**

- A.  Sistemas de Informação
- B.  Ciências da Computação
- C.  Engenharia de Software
- D.  Redes de Computadores

**6 - Qual seu ano de formação?**

R-

**7 - Qual seu nível de formação atual?**

- A.  Graduação Completa
- B.  Especialização Incompleta
- C.  Especialização Completa
- D.  Mestrado Incompleto
- E.  Mestrado Completo
- F.  Doutorado Incompleto
- G.  Doutorado Completo
- H.  Outro

**8 - Possui fluência em alguma outra língua?**

- A.  Inglês
- B.  Espanhol

C. ( ) Francês

D. ( ) Somente Português

**9 - Atualmente se encontra empregado?**

A. ( ) Sim

B. ( ) Não

**10 - Teve dificuldades para inserção no mercado de trabalho?**

A. ( ) Sim

B. ( ) Não

**11 - Em qual setor você atua?**

A. ( ) Público

B. ( ) Privado

C. ( ) Não trabalho

**12 - Qual área trabalha atualmente?**

A. ( ) Suporte Técnico

B. ( ) Desenvolvimento

C. ( ) Análise de Dados

D. ( ) Gerente de TI

E. ( ) Segurança de Informação

F. ( ) Redes de Computadores

G. ( ) Consultoria

H. ( ) Docência

I. ( ) Não trabalho

J. ( ) Outros

**13 - Qual a sua atual remuneração?**

A. ( ) Menos de dois salários

B. ( ) Entre dois e três salários

C. ( ) Entre três e quatro salários

D. ( ) Mais de 4 salários.

**14- Qual sua satisfação com sua atual atividade?**

A. ( ) Péssimo

B. ( ) Regular

C. ( ) Bom

D. ( ) Ótimo

D. ( ) Excelente

**15 - Trabalhou ou trabalha numa função na área de TI na qual não se identifica?**

A. ( ) Sim

B. ( ) Não

**16 - Quanto tempo levou para se inserir no mercado de trabalho ?**

A. ( ) Menos de 6 meses

B. ( ) 6 meses a 1 ano

C. ( ) 1 ano a 2 anos

D. ( ) Mais de 2 anos

E. ( ) Não trabalho

F. ( ) Outros

**17 - Seu trabalho oferece possibilidade de crescimento profissional ou plano de carreira?**

A. ( ) Sim

B. ( ) Não