

UNIVERSIDADE CANDIDO MENDES – UCAM  
PROGRAMA DE PÓS-GRADUAÇÃO EM PESQUISA OPERACIONAL E  
INTELIGÊNCIA COMPUTACIONAL

CURSO DE MESTRADO EM PESQUISA OPERACIONAL E INTELIGÊNCIA  
COMPUTACIONAL

Anderson Alex de Souza Sales

**APLICAÇÃO DO PROCESSO DE DESCOBERTA DE  
CONHECIMENTO EM UM SISTEMA DE SOLICITAÇÃO DE  
SERVIÇOS DE TI: UM ESTUDO DE CASO**

CAMPOS DOS GOYTACAZES/RJ

Dezembro de 2014

UNIVERSIDADE CANDIDO MENDES – UCAM  
PROGRAMA DE PÓS-GRADUAÇÃO EM PESQUISA OPERACIONAL E  
INTELIGÊNCIA COMPUTACIONAL

CURSO DE MESTRADO EM PESQUISA OPERACIONAL E INTELIGÊNCIA  
COMPUTACIONAL

Anderson Alex de Souza Sales

**APLICAÇÃO DO PROCESSO DE DESCOBERTA DE  
CONHECIMENTO EM UM SISTEMA DE SOLICITAÇÃO DE  
SERVIÇOS DE TI: UM ESTUDO DE CASO**

Dissertação apresentada ao Programa de Pós-Graduação em Pesquisa Operacional e Inteligência Computacional, da Universidade Candido Mendes – Campos/RJ, para obtenção do grau de MESTRE EM PESQUISA OPERACIONAL E INTELIGÊNCIA COMPUTACIONAL.

Orientadora: Prof.<sup>a</sup> Geórgia Regina Rodrigues Gomes, D.Sc.

Co-orientador: Prof. Eduardo Shimoda, D.Sc.

CAMPOS DOS GOYTACAZES/RJ

Dezembro de 2014

ANDERSON ALEX DE SOUZA SALES

**APLICAÇÃO DO PROCESSO DE DESCOBERTA DE CONHECIMENTO EM UM  
SISTEMA DE SOLICITAÇÃO DE SERVIÇOS DE TI: UM ESTUDO DE CASO**

Dissertação apresentada ao Programa de Pós-Graduação em Pesquisa Operacional e Inteligência Computacional, da Universidade Candido Mendes – Campos/RJ, para obtenção do grau de MESTRE EM PESQUISA OPERACIONAL E INTELIGÊNCIA COMPUTACIONAL.

Aprovado 19 de dezembro de 2014

---

Prof.<sup>a</sup> Geórgia Regina Rodrigues Gomes, D.Sc. - Orientadora  
Universidade Candido Mendes - Campos

---

Prof. Eduardo Shimoda, D.Sc. - Co-orientador  
Universidade Candido Mendes - Campos

---

Prof. Ítalo de Oliveira Matias, D.Sc.  
Universidade Candido Mendes - Campos

---

Prof. Breno Fabrício Terra Azevedo, D.Sc.  
Instituto Federal Fluminense - campus Campos-Centro

CAMPOS DOS GOYTACAZES, RJ

2014

Dedico este trabalho à minha família, em especial à minha mãe, e à Marcelle, lua da minha vida e eterna fonte de inspiração, que esteve ao meu lado em todas as etapas e soube compreender quando precisei me ausentar.

## AGRADECIMENTOS

Primeiramente a Deus, por toda benevolência e oportunidades a mim oferecidas.

À minha orientadora Geórgia, por sua paciência e sabedoria transmitidas não somente no desenvolvimento desta dissertação, mas também nas disciplinas as quais ministrou.

Ao meu co-orientador Shimoda, pelas palavras de incentivo e direcionamentos realizados.

Aos amigos de turma, em especial aos pertencentes do grupo "Talentation".

E ao Instituto Federal Fluminense pela disponibilização de bolsa de capacitação.

"A mente que se abre a uma nova ideia jamais  
volta ao seu tamanho original."

(Albert Einstein)

## RESUMO

### APLICAÇÃO DO PROCESSO DE DESCOBERTA DE CONHECIMENTO EM UM SISTEMA DE SOLICITAÇÃO DE SERVIÇOS DE TI: UM ESTUDO DE CASO

O estudo desenvolvido objetiva empregar o processo de descoberta de conhecimento em bases de dados aos registros do sistema utilizado para requisitar serviços de Tecnologia da Informação que se encontra instalado nas dependências da direção de TI da reitoria de uma instituição federal de ensino, e se propõe ainda a inferir particularidades de todas as solicitações constantes na amostra que possam subsidiar processos de melhoria na administração desses serviços. A base de dados selecionada fora proveniente do sistema osTicket® na versão 1.6 ST, e a partir da mesma foram definidos os atributos e um modelo de categorias de tratamento de serviços de TI que fora aplicado na classificação das ordens de serviço averiguadas. Posteriormente, os atributos procedentes dessas ordens foram submetidos, através da ferramenta WEKA®, à execução do algoritmo de mineração de dados Apriori para geração de regras de associação, e de acordo com os resultados angariados foi exequível engendrar novos conhecimentos sobre as características dos serviços solicitados, evidenciando aos gestores de Tecnologia da Informação novos subsídios na tomada de decisões estratégicas visando ao aperfeiçoamento continuado da oferta e atendimento de serviços de sua responsabilidade.

**PALAVRAS-CHAVE:** Descoberta de Conhecimento em Bases de Dados; Tecnologia da Informação; WEKA®; Mineração de Dados; Regras de Associação.

## ABSTRACT

### PROCESS APPLICATION OF KNOWLEDGE DISCOVERY IN A IT SERVICES REQUEST SYSTEM: A CASE STUDY

The study developed objectives to employ the process of knowledge discovery in databases to the system used for ordering Information Technology services that is installed on the facilities of the IT direction of a rectory of a federal educational institution, and proposes yet to infer characteristics of all the solicitations in the sample data that can support improvement processes in the management of these services. The selected database was obtained from the system osTicket® in the version 1.6 ST, and from it was defined the attributes and a model of categories treatment of IT services that had applied on the classification of the service orders investigated. Later, the derived attributes of such orders were submitted, through the WEKA© tool, to the execution of the data mining algorithm Apriori to generate association rules, and according to the results raised was feasible engender new knowledgement about the characteristics of the requested services, evidencing to the Information Technology managers new subsidies in making strategic decisions aimed at continuous improvement of supply and customer service of your responsibility.

**KEYWORDS:** Knowledge Discovery in Databases; Information Technology; WEKA©; Data Mining; Association Rules.

## LISTA DE EQUAÇÕES

Equação 1 - Cálculo do grau de confiança. ....	36
Equação 2 - Cálculo do grau de suporte. ....	36

## LISTA DE FIGURAS

Figura 1- Distinção entre governança de TI e gerenciamento da TI. ....	20
Figura 2 - Evolução histórica da ITIL®. ....	21
Figura 3 - Ciclo de vida da ITIL®V3. ....	22
Figura 4 - Etapas do KDD. ....	26
Figura 5 - Modelo de classificador. ....	30
Figura 6 - Exemplo de Clustering. ....	32
Figura 7 - Representação de uma Árvore de Decisão. ....	34
Figura 8 - Rede Neural Artificial. ....	35
Figura 9 - Regras de associação obtidas pelo algoritmo Apriori. ....	36
Figura 10 - Algoritmo Apriori. ....	38
Figura 11 - Função apriori-gen do algoritmo Apriori. ....	38
Figura 12 - Representação de um arquivo arff. ....	41
Figura 13 - Tela de informações do sistema osTicket© na versão 1.8.1.2. ....	43
Figura 14 - Página de solicitação de serviços via web. ....	44
Figura 15 - Tabelas do sistema osTicket©. ....	45
Figura 16 - Consulta SQL realizada no banco de dados do sistema osTicket©. ....	46
Figura 17 - Tela que exhibe a exportação das colunas da tabela ost_ticket. ....	47

## LISTA DE GRÁFICOS

Gráfico 1- Distribuição percentual do atributo “meio_de_abertura” .....	48
Gráfico 2 – Distribuição percentual do atributo “campus_de_origem” .....	49
Gráfico 3 - Distribuição percentual das solicitações pelo seu tempo de resolução.....	50
Gráfico 4 - Distribuição percentual do atributo “tipo_de_chamado”.....	57

## LISTA DE SIGLAS E ABREVIATURAS

CCTA – *Central Computer and Telecommunication Agency*

DCBD – *Descoberta de Conhecimento em Bases de Dados*

DM – *Data Mining*

GSTI – *Gestão de Serviços de Tecnologia da Informação*

ITIL – *Information Technology Infrastructure Library*

itSMF - *Information Technology Service Management Forum*

KDD – *Knowledge Discovery in Databases*

MD – *Mineração de Dados*

OGC - *Office of Government Commerce*

TI – *Tecnologia da Informação*

## LISTA DE TABELAS

Tabela 1- Estágios do ciclo de vida e seus processos. ....	23
Tabela 2 - Modelo de categorias de tratamento de solicitações de serviços de TI. ....	52
Tabela 3 - Regras de associação dos atributos “categoria” → “tipo_de_chamado”. ....	59
Tabela 4 - Regras de associação dos atributos “categoria” → “campus_de_origem”.....	60
Tabela 5 - Regras de associação dos atributos “categoria” → “meio_de_abertura”.....	61
Tabela 6 - Regras de associação dos atributos “categoria” → “tempo_de_resolucao”.....	62
Tabela 7 - Regras de associação dos atributos “tempo_de_resolucao” → “tipo_de_chamado”. .....	64
Tabela 8 - Correlação dos atributos “categoria”, “subcategoria”, “objeto_central” → “tipo_de_chamado”. ....	65
Tabela 9 - Correlação dos atributos “categoria”, “subcategoria”, “objeto_central” → “campus_de_origem”. ....	66
Tabela 10 - Correlação dos atributos “categoria”, “subcategoria”, “objeto_central” → “meio_de_abertura”. ....	67
Tabela 11 - Correlação dos atributos “categoria”, “subcategoria”, “objeto_central” → “tempo_de_resoucao”. ....	68

## Sumário

<b>1. INTRODUÇÃO.....</b>	<b>16</b>
<b>1.1. OBJETIVOS .....</b>	<b>17</b>
1.1.1. Objetivos Específicos .....	17
<b>1.2. JUSTIFICATIVA .....</b>	<b>18</b>
<b>1.3. ESTRUTURA DO TRABALHO .....</b>	<b>18</b>
<b>2. REVISÃO DA LITERATURA.....</b>	<b>19</b>
<b>2.1. GOVERNANÇA DE TI.....</b>	<b>19</b>
<b>2.2. ITIL®.....</b>	<b>20</b>
2.1.1 Estágio do ciclo de vida Operação de Serviço .....	24
<b>2.3. DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS.....</b>	<b>25</b>
2.3.1 Etapas do Processo de Descoberta de Conhecimento .....	26
2.3.1.1 Etapa de Seleção dos Dados .....	26
2.3.1.2 Etapa de Pré-Processamento.....	27
2.3.1.3 Etapa de Transformação dos Dados.....	27
2.3.1.4 Etapa de Mineração dos Dados .....	27
2.3.2 Aspectos relacionados à Mineração de Dados.....	28
2.3.2.1 Tarefas de Mineração de Dados.....	29
2.3.2.1.1 Tarefa de Classificação .....	30
2.3.2.1.2 Tarefa de Regressão .....	31
2.3.2.1.3 Tarefa de Clusterização.....	31
2.3.2.1.4 Tarefa de Associação .....	32
2.3.2.1.5 Tarefa de Sumarização.....	33
2.3.2.2 Métodos de Mineração de Dados .....	33
2.3.2.3 O algoritmo Apriori .....	37
2.3.2.4 O <i>software</i> de Mineração de Dados WEKA©.....	39
<b>3. METODOLOGIA .....</b>	<b>42</b>
<b>3.1. O SISTEMA DE SOLICITAÇÃO DE SERVIÇOS DE TI.....</b>	<b>42</b>
<b>3.2. A SELEÇÃO DOS DADOS .....</b>	<b>44</b>
<b>3.3. ETAPAS DE PRÉ-PROCESSAMENTO E TRANSFORMAÇÃO DOS DADOS</b>	
<b>48</b>	
3.3.1 Processo de criação do Modelo de Categorias de Tratamento de Serviços de TI .....	50

3.3.2	Processo de classificação das solicitações de serviços segundo o modelo criado.....	55
3.4.	<b>ETAPA DE MINERAÇÃO DE DADOS.....</b>	<b>58</b>
4.	<b>RESULTADOS .....</b>	<b>59</b>
5.	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>70</b>
5.1.	<b>CONTRIBUIÇÕES .....</b>	<b>71</b>
5.2.	<b>TRABALHOS FUTUROS .....</b>	<b>71</b>
6.	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>73</b>

## 1. INTRODUÇÃO

Atualmente, o conhecimento embasa-se num bem de extrema valia tanto para empresas privadas quanto públicas, e a forma como ocorrem seus processos de tratamento e gestão impactam diretamente no efetivo alcance do sucesso e de metas estabelecidas. Com o referido valor vital assumido pelo mesmo e a dependência cada vez maior de sistemas de informação, as organizações deixaram de focar essencialmente nos seus bens tangíveis (físicos) ou mensuráveis (financeiros) e a considerar o diferencial competitivo que o conhecimento pode oportunizar (ESTEVEES, 2012; LOPES, 2012).

Paralelamente a esse quadro, constata-se um aumento exponencial do volume de dados armazenados pelas empresas, independente de seu tamanho e área de atuação, nos seus centros de processamento de dados. Tal crescimento pode ser creditado a diversos fatores, como a manutenção do histórico de registros de operações e serviços, a variabilidade de sistemas utilizados, a complexidade de transações financeiras realizadas, etc (CORRÊA, 2007).

Entretanto, extrair conhecimento por intermédio de técnicas de exploração tradicionais em tais repositórios de dados é uma tarefa humanamente inconcebível, seja devido ao hercúleo esforço empregado ou ao tempo e recursos necessários para se concluir a tarefa de análise (CAMILO; SILVA, 2009).

Diante do contexto apresentado, ressalta-se a utilização de técnicas e ferramentas computacionais atreladas ao processamento e avaliação de extensos volumes de dados com o intuito de suprir as deficiências supracitadas e de viabilizar a obtenção de informações e padrões desconhecidos. Dessarte, a linha de pesquisa *Knowledge Discovery in Databases*, conceituada elementarmente como o processo de descoberta de informações de alto nível

(conhecimento) a partir de vastos conjuntos de dados de baixo nível, consiste numa das metodologias que auxiliam a tomada de decisões estratégicas com base em fatos inéditos revelados pelo processo citado (FAYYAD, 1996; CARDOSO; MACHADO, 2008).

## **1.1. OBJETIVOS**

A presente pesquisa propõe-se a aplicar o processo de descoberta de conhecimento em bases de dados aos tíquetes de um sistema de solicitação de serviços de Tecnologia da Informação instalado no *data center* da reitoria de uma instituição federal de ensino visando a depreender as características pertinentes dos serviços de TI requisitados a fim de utilizá-las no aprimoramento de ações e procedimentos relacionados à sua administração e na identificação de demandas atendidas e, porventura, não atendidas.

### **1.1.1. Objetivos Específicos**

Além do objetivo supracitado, busca-se favorecer a obtenção dos seguintes propósitos secundários:

- Demonstrar os conceitos teóricos atrelados ao processo de descoberta de conhecimento em bases de dados discutidos segundo as proposições de Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth e Ramasamy Uthurusamy;
- Elaborar um modelo de categorias de tratamento de solicitações de serviços de TI proveniente dos tíquetes existentes na amostra utilizada;
- Permitir aos gestores de TI da instituição de ensino a estabelecerem as melhores diretrizes por meio do conhecimento obtido no processo inicial da implantação de modelos associados à gestão de serviços de TI;
- Corroborar o modelo de categorias de tratamento criado mediante seu emprego na pesquisa em questão.

## 1.2. JUSTIFICATIVA

Este estudo constitui-se no primeiro de seu âmbito na instituição de ensino para avaliar o que TI oferece e concede suporte e, conseqüentemente, destaca-se a inexistência de informações que apontem as especificidades do ambiente operacional e de parâmetros que indiquem se a TI está alinhada a um processo de uso racional de seus recursos e de melhoria contínua de seus procedimentos e serviços.

Outros elementos fatores que estimulam a realização da presente pesquisa consistem na contribuição a se proporcionar devido a ínfima ocorrência de estudos associando serviços de TI e a aplicação do processo de KDD, e a tentativa de se criar uma abordagem diferenciada que delimite um ponto de partida e os caminhos disponíveis que favoreçam a implantação de modelos associados a uma apropriada gestão de serviços de TI totalmente aderentes às características e requisitos da instituição de ensino.

## 1.3. ESTRUTURA DO TRABALHO

Esta dissertação encontra-se disposta da seguinte maneira:

- Capítulo 2 - aborda as fundamentações teóricas sobre Governança de TI; do *framework* ITIL®, com a apresentação dos aspectos de seu ciclo de vida do serviço; e o processo de descoberta de conhecimento em bases de dados, com a explicitação de suas etapas, dos conceitos sobre *data mining*, suas tarefas, métodos, e do algoritmo e *software* empregados na mineração;
- Capítulo 3 - expõe detalhes do sistema de solicitação de serviços de TI; das técnicas empregadas na obtenção da amostra; da criação do modelo de categorias de tratamento de solicitações de serviços de TI e geração dos atributos;
- Capítulo 4 - apresenta a aplicação do algoritmo de mineração de dados Apriori para extração de conhecimento sobre os atributos selecionados e os resultados inferidos;
- Capítulo 5 - expressa as conclusões desta pesquisa bem como suas contribuições e sugestões de trabalhos futuros.

## 2. REVISÃO DA LITERATURA

Dentre os vários modelos ou *frameworks* existentes com o objetivo concreto de aperfeiçoar e qualificar a gestão de serviços de TI das organizações, tratar-se-á neste trabalho o *framework* ITIL®. Ressalta-se que o objetivo das duas seções a seguir não é de aprofundar todos os conceitos inerentes à governança de TI ou aos livros da ITIL® e seus processos, visto que, nesta pesquisa, ambos foram utilizados como complementos para uma melhor compreensão da abordagem adotada.

### 2.1. GOVERNANÇA DE TI

A governança de TI pode ser fundamentada basicamente como uma estrutura de processos, mecanismos e relacionamentos que pode ser implementada visando ao planejamento, organização e monitoramento das estratégias, recursos e atividades de TI. Sua orientação deve garantir, ou pelo menos maximizar, o alinhamento de tudo o que é concernente à TI aos objetivos e metas estratégicas de uma organização ou, em se tratando do setor governamental, aos objetivos da administração pública, suas atividades finalísticas e seus usuários (SETHIBE; CAMPBELL; MCDONALD, 2007; CEPIK; CANABARRO, 2010).

Contudo, é necessário salientar a distinção existente entre governança de TI e gerenciamento da TI, pois a primeira compreende um sentido mais amplo, pois suas ações buscam efetivar o direcionamento das ações TI às demandas presentes e futuras dos negócios

de uma organização, enquanto o gerenciamento da TI procura garantir uma eficiente e eficaz disponibilização de produtos e serviços de TI (PETERSON, 2004; LUNARDI, 2008). A figura 1, apresentada abaixo, corrobora e auxilia elucidar a diferenciação acima abordada:

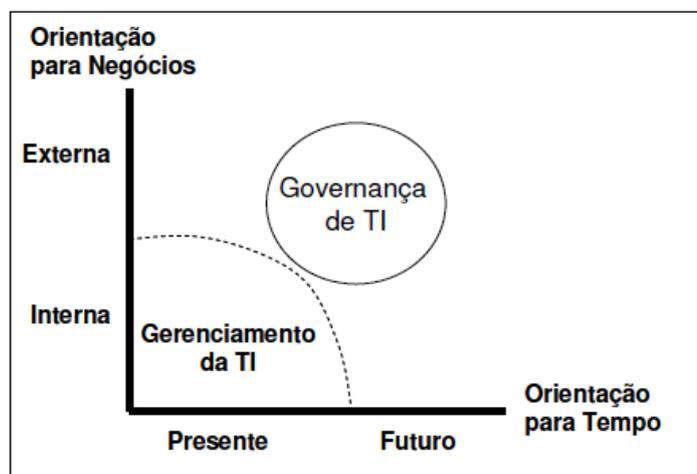


Figura 1- Distinção entre governança de TI e gerenciamento da TI.

Fonte: Peterson (2004).

Portanto, expressa-se que a governança de TI possui uma conotação mais abrangente, com uma visão que incorpora o tratamento da importância da informação e dos negócios da TI e ao mesmo tempo acrescenta a necessidade de abordar não somente sua estrutura, mas também os processos que farão com que os produtos e serviços sejam alcançados e que estejam alinhados ao objetivo estratégico do negócio (PETERSON, 2004; LUNARDI, 2008).

## 2.2. ITIL®

A ITIL® consiste num projeto inicialmente desenvolvido pelo CCTA na década de 1980 quando o governo britânico verificou que o nível de qualidade dos serviços de TI prestados não era eficiente. Dessa forma, fora requisitada a criação de uma abordagem que propiciasse um gerenciamento eficiente e responsável dos recursos de TI e extensivo a empresas com particularidades técnicas e de negócio distintas. Anos mais tarde, em abril de 2001, a CCTA foi integrada ao OGC, organismo responsável atualmente pela divulgação e evolução da ITIL® (FERNANDES; ABREU, 2012). Abaixo, a figura 2 apresenta a evolução do *framework*:

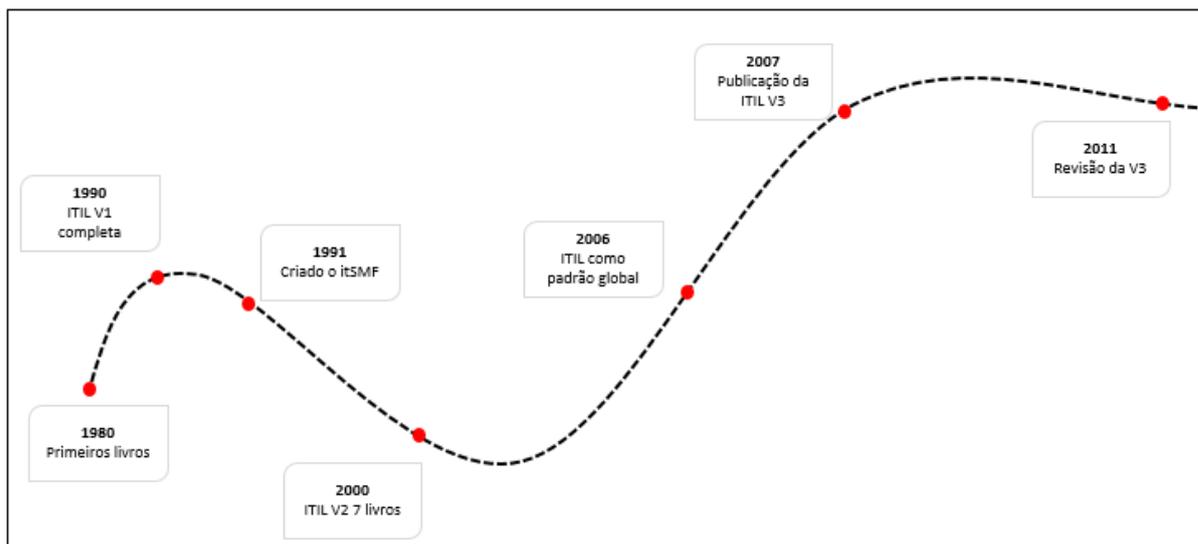


Figura 2 - Evolução histórica da ITIL®.

Fonte: Adaptado de Freitas (2013).

A ITIL®, na sua versão inicial, constituía-se numa biblioteca de 31 volumes que abrangia todos os aspetos sobre o fornecimento de serviços TI, e desde o seu desenvolvimento houve uma evolução de sua documentação por meio de duas atualizações. Em 2001, a primeira delas, foi lançada sua segunda versão e o número de livros foi reduzido para sete, o que a auxiliou a ser reconhecida mundialmente e adotada em larga escala (CARTLIDGE *et al.*, 2007a).

Sua última versão, lançada em maio de 2007 e revisada em julho de 2011, possuía cinco livros, cada um correspondente a um estágio do ciclo de vida do serviço, conforme exposto na figura 3 (CARTLIDGE *et al.*, 2007a):

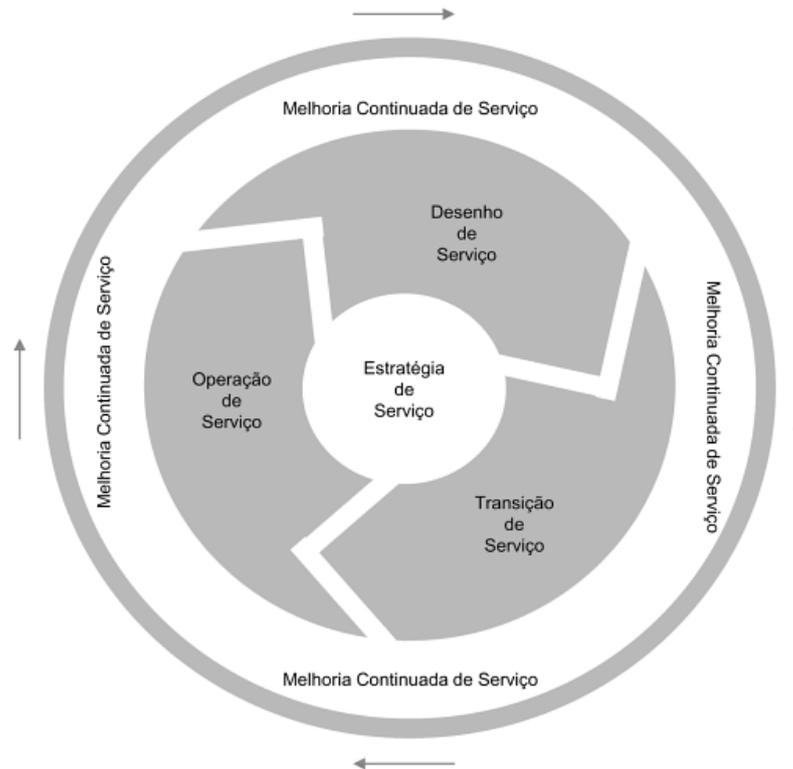


Figura 3 - Ciclo de vida da ITIL®V3.

Fonte: Adaptado de Cartlidge *et al.* (2012b).

No que concerne o ciclo de vida, a Estratégia de Serviço caracteriza-se por ser seu estágio inicial. O mesmo fornece orientações sobre como projetar, desenvolver e implementar a gestão de serviço não somente como uma tarefa da organização, mas sim como um ativo estratégico. Também visa a estabelecer quem são os clientes da TI, os serviços que são oferecidos e os recursos e requisitos necessários para desenvolvê-los e cumpri-los com sucesso (ARRAJ, 2013).

O propósito do Desenho de Serviço é o de assegurar que os serviços novos ou modificados sejam projetados para atender as exigências do negócio. Este livro é o estágio do ciclo de vida que transforma uma nova exigência da Estratégia de Serviço num projeto para atender os objetivos de negócios (ARRAJ, 2013).

A finalidade da Transição de Serviço é de garantir que os serviços novos ou modificados atendam às expectativas da organização como documentada na Estratégia de Serviço e Desenho de Serviço. Algumas das principais atividades desta etapa incluem gerenciamento de mudança, gerenciamento de conhecimento e a garantia de que o valor esperado pelo negócio seja entregue (FERNANDES; ABREU, 2012).

O objetivo da Operação de Serviço é de entregar os níveis de serviço acordados com

os usuários e clientes e de administrar as aplicações, tecnologias e infraestrutura que suportam a entrega dos serviços (CARTLIDGE *et al.*, 2012b). Esta fase do ciclo de vida será mais abordada adiante, uma vez que foi utilizada como subsídio para a pesquisa apresentada.

A Melhoria Continuada do Serviço congrega princípios, práticas e métodos de gerenciamento de qualidade para aprimorar cada estágio do ciclo de vida, as tecnologias empregadas e a eficiência e eficácia dos processos utilizados na gestão global de serviços (CARTLIDGE *et al.*, 2012b; ARRAJ, 2013).

Cada fase do ciclo de vida possui processos correlacionados, e na tabela 1, exibida abaixo, são representadas tais associações (FERNANDES; ABREU, 2012).

Tabela 1- Estágios do ciclo de vida e seus processos.

Estágio do ciclo de vida de serviço	Processos
Estratégia de Serviço	Gerenciamento de estratégia Gerenciamento de Portfólio de Serviço Gerenciamento Financeiro Gerenciamento da Demanda Gerenciamento do Relacionamento com o Negócio
Desenho de Serviço	Coordenação do Desenho Gerenciamento de Catálogo de Serviço Gerenciamento de Nível de Serviço Gerenciamento da Disponibilidade Gerenciamento da Capacidade Gerenciamento da Continuidade do Serviço de TI Gerenciamento de Segurança da Informação Gerenciamento de Fornecedor
Transição de Serviço	Planejamento e Suporte da Transição Gerenciamento de Mudança Gerenciamento de Configuração e Ativo de Serviço Gerenciamento de Liberação e Implantação Validação e Teste de Serviço Avaliação da Mudança Gerenciamento de Conhecimento
Operação de Serviço	Gerenciamento de Evento Gerenciamento de Incidente Gerenciamento de Requisição Gerenciamento de Problema Gerenciamento de Acesso
Melhoria Continuada de Serviço	Processo de Melhoria em 7 Etapas Mensuração de Serviço

### 2.1.1 Estágio do ciclo de vida Operação de Serviço

A Operação de Serviço pode ser vista, analogicamente, como a fábrica de TI. Isto induz a uma implicação de maior foco nas atividades rotineiras e na infraestrutura que são utilizadas na prestação de serviços. Este estágio é imprescindível no ciclo de vida, uma vez que equívocos no controle e gerenciamento de atividades operacionais podem, eventualmente, comprometer a disponibilidade de um serviço. Além disso, o mesmo é composto por cinco processos: Gerenciamento de Evento, Gerenciamento de Incidente, Gerenciamento de Requisição, Gerenciamento de Problema e Gerenciamento de Acesso (OGC, 2007; FERNANDES; ABREU, 2012).

Neste ciclo, existem alguns conceitos importantes e suas definições são abordadas abaixo (CARTLIDGE *et al.*, 2012b):

- **Requisição de serviço:** é um pedido formal de um usuário de algo a ser fornecido;
- **Evento:** é um *status report* criado por um serviço, item de configuração ou ferramenta de monitoramento causado pela alteração no desempenho da infraestrutura ou de entrega de serviço;
- **Incidente:** interrupção inesperada ou redução da qualidade de um serviço de TI. Pode ser uma falha de um item de configuração que ainda não tenha impactado o serviço, mas que possa ocasioná-lo posteriormente;
- **Problema:** é a denominação da causa de um ou mais incidentes;
- **Solução de contorno:** medida adotada para resolver um incidente ou problema de forma temporária ou paliativa enquanto a resolução completa não for aplicada;
- **Erro conhecido:** é um problema que tem a causa raiz documentada e uma solução de contorno criada;
- **Base de Erros Conhecidos:** é um repositório centralizado de registros de erros conhecidos.

### 2.3. DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

O contínuo e rápido avanços pertinentes às diversas áreas de TI bem como a perceptível redução de custos de dispositivos de armazenamento em massa propiciam que expressivas quantidades de dados sejam neles inseridas. Com isso, manipular, processar e interpretar as informações contidas nessas bases de forma otimizada caracterizam-se como procedimentos vitais na técnica de tomada de decisão (ASSMANN; FROZZA; KIPPER, 2012).

Todavia, o modo tradicional de transformação de dados em conhecimento, baseado no processamento manual de todas as informações por especialistas que geram diversos relatórios a serem analisados, esbarra numa das consequências da chamada era da informação: a sobrecarga de dados (FAYYAD *et al.*, 1996). A exploração desses extensos volumes de dados pelo ser humano é impraticável sem o emprego de subsídios computacionais, e para isso é fundamental a utilização de ferramentas que assistam de maneira automática e inteligente a avaliação, interpretação e relacionamento dos dados de tais bases para selecionar as melhores estratégias a serem adotadas no contexto em que forem aplicadas (GOLDSCHMIDT; PASSOS, 2005).

Sendo assim, o processo de DCBD possui como um de seus intuitos dirimir os problemas mencionados e potencializar a descoberta de informações, o que é corroborado de acordo com a seguinte definição: KDD é um processo não trivial, composto por diversas etapas, tipicamente interativo e iterativo, de identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de um conjunto de dados (FAYYAD *et al.*, 1996).

O processo de DCBD baseia-se, portanto, na exploração de dados armazenados a fim de permitir a obtenção de informações outrora ocultas e desconhecidas em constatações compreensíveis e aplicáveis e que assistam, portanto, à tomada de decisões estratégicas (GALVÃO; MARIN, 2009).

Cabe salientar que as definições de descoberta de conhecimento em bases de dados e mineração de dados (DM), que será abordada adiante, não são equivalentes, pois o KDD corresponde a todo o processo de extração de conhecimento em bases de dados enquanto *data mining* compreende-se por ser uma de suas etapas e culmina na utilização de técnicas de inteligência artificial atreladas a áreas como estatística, banco de dados, reconhecimento de

padrões e interface humano-máquina (CÔRTEZ; PORCARO; LIFSCHITZ, 2002; CARVALHO, 2005).

Historicamente, o termo descoberta de conhecimento em bases de dados foi cunhado no primeiro *workshop* sobre KDD em 1989 para se referir ao amplo processo de encontrar conhecimento em dados e para enfatizar a aplicação de alto nível de determinados métodos de mineração de dados. Este termo é mais empregado por pesquisadores das áreas de inteligência artificial e de aprendizado de máquina, enquanto a nomenclatura *data mining* é geralmente usada por estatísticos e analistas de dados (FAYYAD *et al.*, 1996).

### 2.3.1 Etapas do Processo de Descoberta de Conhecimento

O processo de descoberta de conhecimento em bases de dados é subdividido em cinco etapas: seleção, pré-processamento, transformação, mineração de dados e interpretação/avaliação (FAYYAD *et al.*, 1996). Cada uma delas será detalhada adiante e são ilustradas na figura 4:

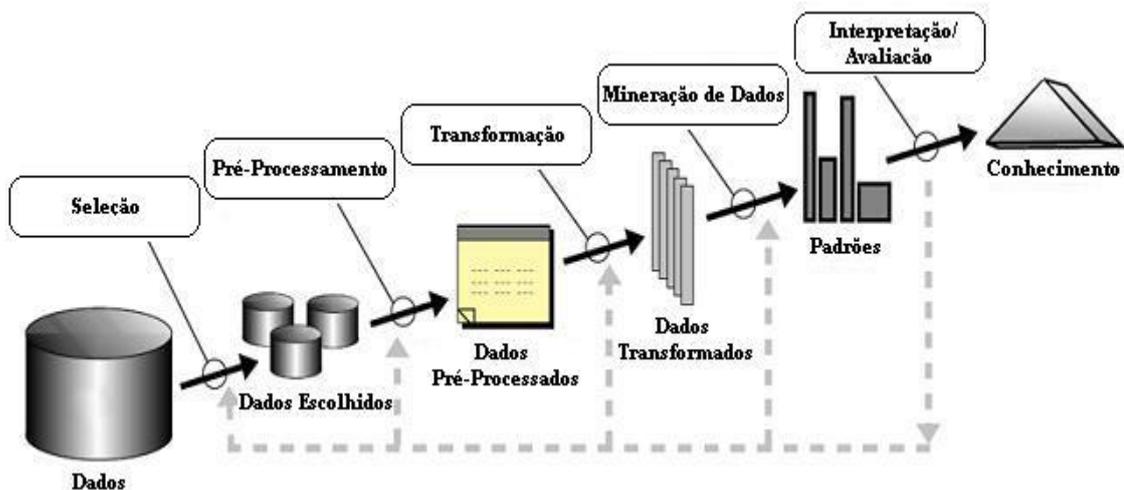


Figura 4 - Etapas do KDD.

Fonte: Adaptado de Fayyad *et al.* (1996).

#### 2.3.1.1 Etapa de Seleção dos Dados

Estágio onde são predefinidos o desenvolvimento e compreensão do domínio da aplicação, o conhecimento prévio e os propósitos a serem alcançados de acordo com a base escolhida. Concentra-se ainda na seleção de um conjunto de dados ou de um subconjunto de variáveis ou amostras de dados onde se pretende realizar a descoberta a ser alcançada (DUNHAN, 2006).

### **2.3.1.2 Etapa de Pré-Processamento**

Nessa etapa são empregadas operações básicas para a remoção de ruídos, *outliers* e correção de valores inconsistentes, decidindo dessa forma sobre as estratégias a serem adotadas para lidar com campos de dados faltantes, representação de informações relativas a sequências temporais, etc (FAYYAD *et al.*, 1996).

### **2.3.1.3 Etapa de Transformação dos Dados**

Nesse passo são utilizados alguns recursos para consolidar a representação dos dados dependendo do objetivo da tarefa a ser escolhida na etapa de mineração e métodos para redução da dimensionalidade e do número de variáveis (CÔRTEZ; PORCARO; LIFSCHITZ, 2002; HAN; KAMBER; PEI, 2011).

### **2.3.1.4 Etapa de Mineração dos Dados**

É considerada o núcleo do processo de descoberta de conhecimento. Nela, é definida a tarefa, o(s) método(s) e o(s) algoritmo(s) que serão empregados sobre o modelo obtido para extração de padrões ou comportamentos de interesse que delimitam suas características numa

forma de representação compreensível (CÔRTEZ; PORCARO; LIFSCHITZ, 2002).

### **2.3.1.5 Etapa de Interpretação/Avaliação**

Fundamenta-se na elucidação dos padrões minerados e validação do conhecimento adquirido, buscando identificar e avaliar padrões e inferências que propiciem suporte a decisões às partes interessadas (CÔRTEZ; PORCARO; LIFSCHITZ, 2002; HAN; KAMBER; PEI, 2011).

### **2.3.2 Aspectos relacionados à Mineração de Dados**

A mineração de dados constitui-se numa área de pesquisa multidisciplinar que inclui especialmente tecnologias de banco de dados, estatística, reconhecimento de padrões, computação de alto desempenho e visualização de dados. Mesmo existindo muita documentação sobre o tema, há inúmeras abordagens e não ocorre uma padronização universalmente aceita devido aos diversos públicos-alvo que as originam e se destinam (CARDOSO; MACHADO, 2008).

Ainda, *Data mining* encerra-se numa metodologia que busca uma descrição lógica ou matemática, por vezes complexa, de padrões e regularidades num conjunto de dados. A mesma pode ser definida, por vezes, como a descoberta de padrões, mudanças, anomalias e estruturas estatísticas em dados. Com isso, afirma-se que a DM abrange uma gama de técnicas empregadas para examinar conjuntos de dados visando trazer à tona complexas relações entre eles, e essas técnicas não necessitam, obrigatoriamente, que os arranjos entre os dados sejam conhecidos preliminarmente, pois o processo viabiliza o surgimento de novas associações (CARDOSO; MACHADO, 2008).

Dessarte, o objetivo da mineração de dados é de favorecer a descoberta do conhecimento implícito e não utilizado de informações armazenadas, e resumidamente, sua aplicação propicia algumas possibilidades, tais como (ELMASRI; NAVATHE, 2010; CARDOSO; MACHADO, 2008):

- Classificação - divisão dos dados em classes ou categorias distintas determinadas com base em combinações de parâmetros;
- Identificação - quando há a viabilidade de se evidenciar, a partir de certos padrões, a ocorrência de um item, evento ou atividade;
- Previsão - quando há a necessidade de verificar como atributos irão se comportar no futuro.

Uma aplicação prática da mineração de dados pode ser destacada, por exemplo, num exemplo clássico empregado na rede de supermercados WalMart, onde o estudo das associações de produtos que constavam num mesmo cupom fiscal identificou a tendência relacional da venda de fraldas infantis e cervejas. O perfil delineado dos consumidores destes produtos particularizava-se por serem homens casados, na faixa dos 25 a 30 anos, que os compravam no período vespertino e no deslocamento trabalho → residência. A partir desta constatação, buscou-se redefinir o layout de forma que houvesse uma aproximação das seções de bebidas e de produtos infantis e, de acordo com as inferências proporcionadas pela mineração de dados, verificou-se que as vendas aumentaram cerca de 30% após o redesenho das prateleiras que armazenavam os referidos produtos (BORTOLI, 2012).

Todavia, antes da efetiva execução da mineração de dados, é essencial determinar a(s) tarefa(s), o(s) método(s) e algoritmo(s) a serem aplicados, e seus conceitos estão dispostos adiante.

### **2.3.2.1 Tarefas de Mineração de Dados**

O conceito de tarefa consiste fundamentalmente na especificação do que se objetiva lograr dos dados selecionados, a fim de adquirir uma resposta para o problema que se possui (REZENDE, 2003; GARCIA, FROZZA, 2013).

Os principais objetivos da mineração de dados propendem-se a destacar tarefas de predição e descrição. A predição faz uso de algumas variáveis para presumir valores desconhecidos ou futuros de outras variáveis de interesse, enquanto a descrição concentra-se na descoberta de padrões descritivos e que sejam humanamente interpretáveis. Entretanto, o limite entre essas atividades é muito tênue, pois alguns modelos preditivos podem ser descritivos e vice-versa. As tarefas mais comuns e utilizadas para a predição são as de

classificação e regressão, enquanto que para a descrição apontam-se as de clusterização, associação e sumarização (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; CARVALHO, 2005; GARCIA, FROZZA, 2013).

### 2.3.2.1.1 Tarefa de Classificação

A classificação consiste numa função de aprendizado que delimita um conjunto de dados ou registros em diversas classes pré-definidas. Assim sendo, seu foco concentra-se na descoberta de uma função que mapeie um conjunto de registros num único grupo de variáveis previamente estabelecidas (classes), de maneira que seja possível prever a que classe, inclusive, novos registros se enquadrariam caso sejam a ela submetidos (GOLDSHMIDT; PASSOS, 2005; GALVÃO; MARIN, 2009).

Essa tarefa tem sido a mais estudada e utilizada ao longo dos anos, e o intuito principal da definição de um classificador constitui-se na descoberta da relação entre atributos previsores (conjunto de registros) e os atributos classes. Exemplificando o exposto, a figura 5 apresenta um classificador hipotético que possui como objetivo apontar a relação existente entre os atributos previsores  $A_1$  e  $A_2$  e os valores das classes (+ e -). Seu funcionamento é instituído por recursivos particionamentos do espaço de dados que a cada estágio verificam se novas subáreas devem ser criadas a fim de se constituir um novo modelo de separação das classes (CARVALHO, 2005).

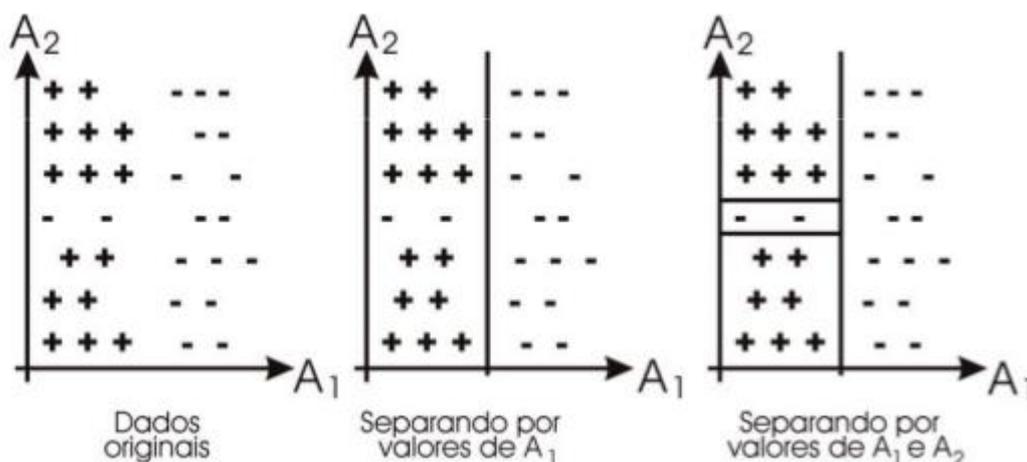


Figura 5 - Modelo de classificador.

Fonte: Carvalho (2005).

### 2.3.2.1.2 Tarefa de Regressão

A regressão baseia-se numa função de aprendizado onde a variável a ser predita corresponde a um atributo numérico existente num banco de dados com valores reais, o que permite estimar o valor de uma variável observando-se os valores das demais. Sua implementação utiliza, geralmente, métodos estatísticos e redes neurais e suas aplicações são inúmeras, como a dedução da probabilidade de recuperação de um paciente de acordo com os resultados de uma bateria de exames, previsão da quantidade de biomassa numa floresta proporcionada por medições realizadas por sensoriamento remoto, etc (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; GALVÃO; MARIN, 2009; CAMILO; SILVA, 2009).

### 2.3.2.1.3 Tarefa de Clusterização

A clusterização é uma tarefa descritiva utilizada para segregar os registros de uma base de dados em agrupamentos (subconjuntos), permitindo identificar um conjunto finito de categorias para descrever os dados. Um *cluster* denota-se por ser um conjunto de registros agrupados de acordo com sua similaridade, e os registros são associados de tal maneira que as proximidades *intraclusters* (mesmo *cluster*) sejam maximizadas e as *interclusters* (*clusters* diferentes) sejam minimizadas, conforme observado na figura 6. Essa tarefa se diferencia basicamente da de classificação pois apresenta características de aprendizado não-supervisionado (não necessita que os registros sejam categorizados preliminarmente) e uma vez que os *clusters* são definidos, os registros são atribuídos a seus agrupamentos correspondentes. Uma aplicação prática pode ser descrita nos casos em que se deseja encontrar subpopulações homogêneas de consumidores para fins de marketing (CARVALHO, 2005; CAMILO; SILVA, 2009).

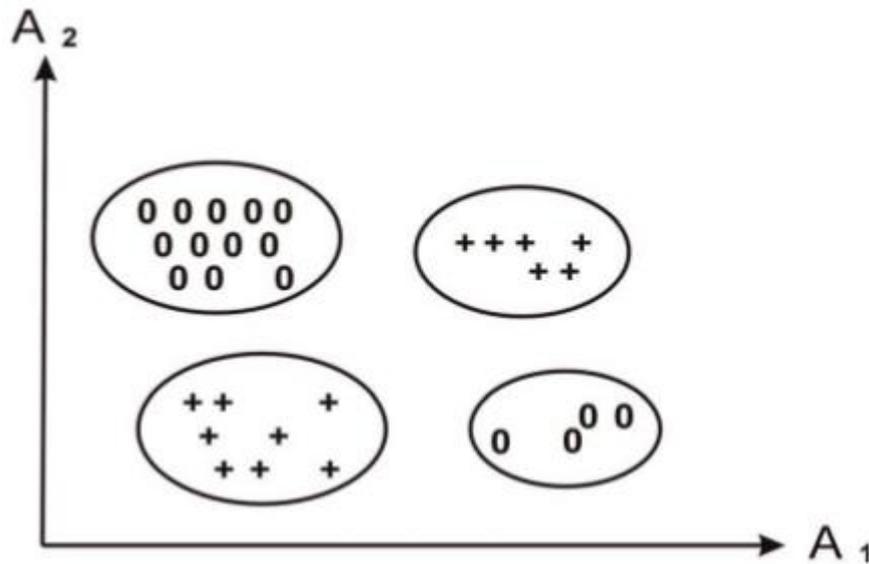


Figura 6 - Exemplo de *Clustering*.

Fonte: Carvalho (2005).

#### 2.3.2.1.4 Tarefa de Associação

Esta tarefa possui o intuito de determinar e descrever correlações entre atributos que são constatadas empreendendo-se múltiplos passos iterativos sobre o conjunto de dados utilizado (CARVALHO, 2005).

Uma exemplificação da aplicação da tarefa de associação pode ser observada por regras que descrevem relações como {produto X, produto Y  $\rightarrow$  produto Z} ou {antecedente  $\rightarrow$  consequente}. Tal regra permite inferir que se um cliente compra o produto X e o produto Y, o mesmo tem a tendência de também comprar o produto Z.

Uma distinção da tarefa de associação em relação à de classificação concerne-se à questão sintática, pois as regras de associação podem ter mais de um atributo no seu consequente, enquanto as regras de classificação possuem apenas um. Nesse sentido, afirma-se que a tarefa de classificação é assimétrica em relação a seus itens pois seus atributos previsores podem unicamente ocorrer no antecedente e o atributo classe apenas no consequente. Em contrapartida, a tarefa de associação é dita simétrica, uma vez que seus atributos podem ocorrer tanto no antecedente ou no consequente da regra (CARVALHO, 2005).

### **2.3.2.1.5 Tarefa de Sumarização**

A sumarização envolve técnicas empregadas para descrever ou apontar características comuns para um subconjunto de dados. A aplicação da tarefa de sumarização é geralmente aplicada na análise exploratória de dados com geração automatizada de relatórios e no pré-processamento dos dados, quando métodos estatísticos são utilizados para determinar valores incoerentes. As técnicas de sumarização mais elaboradas são denominadas de visualização, que podem ser diagramas baseados em proporção e dispersão, histogramas, etc (SFERRA; CORRÊA, 2003; REZENDE, 2003; GARCIA; FROZZA, 2013).

### **2.3.2.2 Métodos de Mineração de Dados**

Os métodos utilizados no processo de descoberta de conhecimento são tecnologias que proporcionam a obtenção de resultados após a definição dos objetivos a serem alcançados no processo de mineração. Dessa forma, deve-se verificar a melhor técnica a fim de se alcançar os melhores resultados com o máximo de precisão (CÔRTEZ; PORCARO; LIFSCHITZ, 2002).

Geralmente, os métodos de mineração de dados são delimitados às categorias de aprendizado supervisionado (preditivo) e não supervisionado (descritivo), que se distinguem elementarmente pelo fato dos métodos de aprendizado supervisionado possuírem no seu conjunto de dados uma variável (ou atributo) classe que consiste a base em que os demais registros existentes serão classificados. Já os métodos de aprendizado não supervisionado não carecem de nenhuma pré-categorização, ou seja, emprega instâncias sem um atributo classe (CAMILO; SILVA, 2009; DAMASCENO, 2010). A seguir, são apresentados alguns dos principais e mais empregados métodos de mineração, sendo eles: Árvores de Decisão, Redes Neurais Artificiais, Regras de Associação, Algoritmos Genéticos e Lógica Fuzzy.

As Árvores de Decisão, representadas graficamente por nós, galhos (ramificações) e folhas, são também denominadas árvores de classificação ou de regressão, e seus nós indicam um teste realizado sobre o valor de um atributo, os galhos representam os valores possíveis do

teste efetuado no nó superior e as folhas denotam a categoria (classe) a qual um registro está atribuído. Assim sendo, quando se deseja classificar um novo registro a partir de uma Árvore de Decisão montada, basta seguir o fluxo criado observando os testes dos nós que descendem a partir do nó raiz até chegar a uma das folhas. A intenção da aplicação de uma Árvore de Decisão é de construir um modelo de predição categórico ou de encontrar a estrutura preditiva de um problema, sendo que os dois propósitos não são excludentes, e podem inclusive ocorrerem num mesmo estudo (CAMILO; SILVA, 2009). Para uma melhor ilustração, a figura 7 apresenta um modelo de representação de uma Árvore de Decisão:

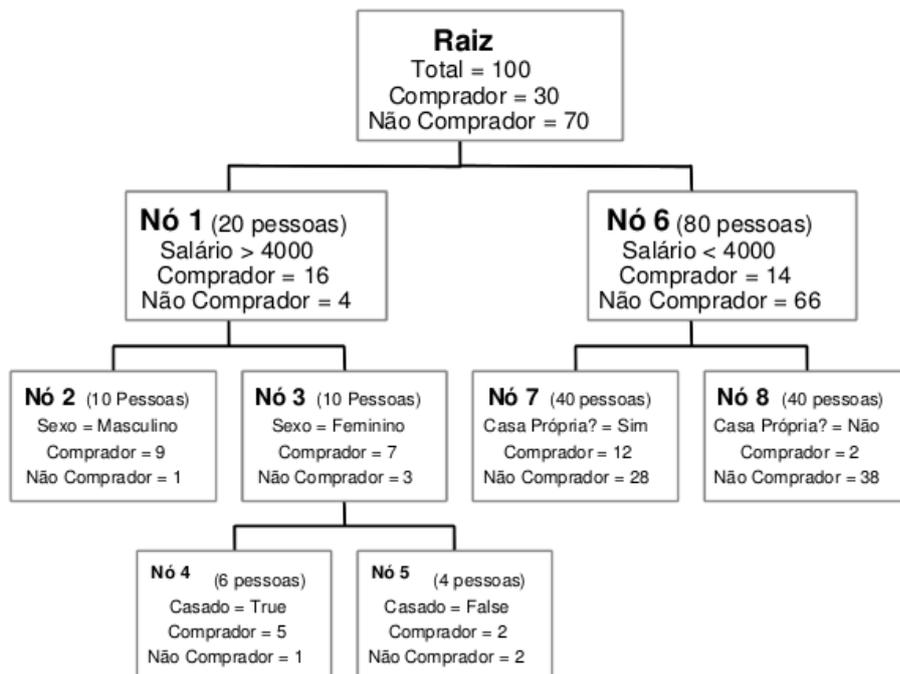


Figura 7 - Representação de uma Árvore de Decisão.

Fonte: CÔRTEZ; PORCARO; LIFSCHITZ (2002).

Uma rede neural artificial pode ser definida como uma técnica computacional inspirada no funcionamento do cérebro humano que apresenta um conjunto de unidades simples de processamento (neurônios) de entrada e saída interligadas por um grande número de conexões (sinapses) a camadas intermediárias, com cada uma das ligações possuindo um peso associado que durante o processo de aprendizado é ajustado para armazenar o conhecimento adquirido permitindo assim classificar corretamente um objeto (CAMILO; SILVA, 2009).

Sua vantagem sobre métodos convencionais corresponde ao fato de não necessitarem de conhecimento detalhado dos processos a serem modelados, podendo ainda identificar

padrões para os quais não foram treinadas. Um exemplo de algoritmo que implementa este conceito é o *backpropagation*, muito utilizado nos anos 80 e que realiza aprendizado através da detecção de erros (TARAPANOFF; JÚNIOR; CORMIER, 2000; KOVÁCS, 2002; GALVÃO; MARIN, 2009). A figura 8, mostrada a seguir, apresenta a esquematização de uma rede neural artificial:

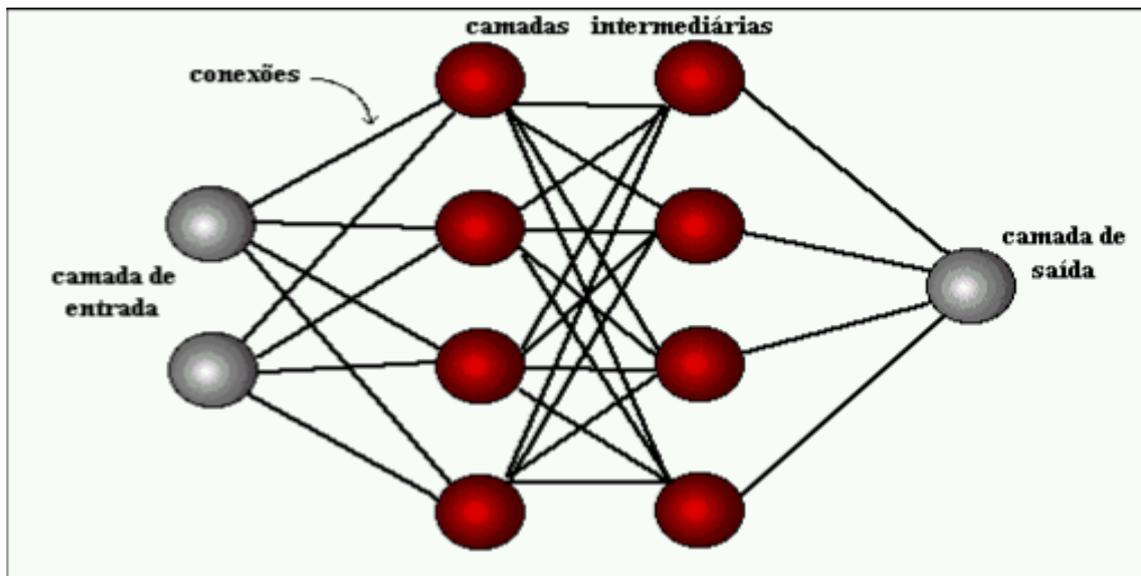


Figura 8 - Rede Neural Artificial.

Fonte: CÔRTEZ; PORCARO; LIFSCHITZ (2002).

O método de Regras de Associação consiste num dos mais conhecidos na mineração de dados e funcionalmente apresenta um padrão de relacionamentos entre itens que ocorrem frequentemente numa base de dados. Sua representação pode ser dada por  $I = \{i_1, i_2, \dots, i_m\}$ , sendo  $I$  um conjunto de itens (*itemsets*), e suas regras associativas são expressas na forma  $A \rightarrow B$ , onde  $A$  é o antecedente,  $B$  o conseqüente, e  $A \subset I$ ,  $B \subset I$  e  $A \cap B = \emptyset$  (AGRAWAL; SRIKANT, 1994; GILLMEISTER; CAZELLA, 2007).

A seguir, e visando a demonstrar regras de associação a partir de um caso real, a figura 9 ilustra algumas regras geradas pelo algoritmo Apriori no *software* WEKA©:

```

Associator output

Apriori
=====

Minimum support: 0.1 (255 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 3
Size of set of large itemsets L(2): 3
Size of set of large itemsets L(3): 1

Best rules found:

1. curso=Segurança do Trabalho estudou_em_escola_publica=NAO 2024 ==> trabalha=NAO 1982   conf:(0.98)
2. estudou_em_escola_publica=NAO 2324 ==> trabalha=NAO 2274   conf:(0.98)
3. curso=Segurança do Trabalho trabalha=NAO 2076 ==> estudou_em_escola_publica=NAO 1982   conf:(0.95)
4. trabalha=NAO 2385 ==> estudou_em_escola_publica=NAO 2274   conf:(0.95)
5. curso=Segurança do Trabalho 2227 ==> trabalha=NAO 2076   conf:(0.93)
6. curso=Segurança do Trabalho 2227 ==> estudou_em_escola_publica=NAO 2024   conf:(0.91)
7. curso=Segurança do Trabalho 2227 ==> estudou_em_escola_publica=NAO trabalha=NAO 1982   conf:(0.89)
8. estudou_em_escola_publica=NAO trabalha=NAO 2274 ==> curso=Segurança do Trabalho 1982   conf:(0.87)
9. estudou_em_escola_publica=NAO 2324 ==> curso=Segurança do Trabalho 2024   conf:(0.87)
10. trabalha=NAO 2385 ==> curso=Segurança do Trabalho 2076   conf:(0.87)
11. estudou_em_escola_publica=NAO 2324 ==> curso=Segurança do Trabalho trabalha=NAO 1982   conf:(0.85)
12. trabalha=NAO 2385 ==> curso=Segurança do Trabalho estudou_em_escola_publica=NAO 1982   conf:(0.83)

```

Figura 9 - Regras de associação obtidas pelo algoritmo Apriori.

Fonte: O autor.

A toda regra de associação está atrelado um grau de confiança, que corresponde a porcentagem das transações do *itemset*  $A \cup B$  dentre todas as transações em que  $A$  ocorre, o que serve, portanto, para validar uma determinada regra (AMO, 2012). A equação 1, apresentada abaixo, explicita a modelagem matemática  $\text{conf}(A \rightarrow B)$ :

$$\text{conf}(A \rightarrow B) = \frac{\text{número de transações que suportam } A \cup B}{\text{número de transações que suportam } A}$$

Equação 1 - Cálculo do grau de confiança.

Fonte: Adaptado de Amo (2012).

Da mesma forma, toda regra de associação está também atrelada a um grau de suporte, que corresponde ao número de transações que contem um *itemset* dividido pelo número total de transações geradas, o que favorece a indicação da significância estatística da regra (ROMÃO *et al.*, 1999). A representação de seu modelo matemático é exibida na equação 2:

$$\text{sup}(A \rightarrow B) = \frac{\text{número de transações que suportam } A \cup B}{\text{número de transações geradas}}$$

Equação 2 - Cálculo do grau de suporte.

Fonte: O autor.

Os Algoritmos Genéticos fundamentam-se na utilização de técnicas de otimização algorítmica para resolução de problemas e são baseados em princípios como a genética e seleção natural. Seu funcionamento baseia-se na definição de uma população no estágio inicial de forma aleatória e, seguindo o princípio de sobrevivência do mais apto, uma nova população é criada com base na primeira geração, com seus indivíduos sendo submetidos recursivamente a processos de cruzamento e mutação. Tal processo se estende até que novos indivíduos mais fortes sejam gerados ou ao se alcançar algum critério de parada, convergindo assim numa solução ideal. Devido a sua capacidade de obter soluções em paralelo, são considerados excelentes quanto utilizados no processo de mineração de dados, entretanto, apresentam como desvantagem a superprodução de soluções individuais e a alta demanda de processamento computacional (CÔRTEZ; PORCARO; LIFSCHITZ, 2002; CAMILO; SILVA, 2009; GALVÃO; MARIN, 2009; HAN; KAMBER; PEI, 2011).

A Lógica Nebulosa ou Lógica *Fuzzy* encerra-se numa teoria matemática que propicia uma modelagem aproximada de raciocínio, simulando a capacidade humana de tomar decisões racionais perante situações de imprecisão e/ou incertezas. Dessarte, a Lógica Nebulosa baseia-se num mecanismo de manipulação de informações imprecisas que proporciona inferir uma solução inexata frente a uma questão baseada num conhecimento impreciso ou não confiável. Algumas de suas utilizações são empregadas em sistemas de controle e de suporte à decisão (GALVÃO; MARIN, 2009; HAN; KAMBER; PEI, 2011).

### **2.3.2.3 O algoritmo Apriori**

O algoritmo Apriori é empregado na mineração de itens frequentes e seu funcionamento básico concentra-se na realização de múltiplos passos sobre a base de dados onde inicialmente são determinados o índice de suporte de cada item individualmente para determinar quais deles são *itemsets* frequentes, e nos passos seguintes os *itemsets* candidatos (*itemsets* frequentes encontrados) são gerados e seus índices de suporte mínimo são comparados aos especificados como parâmetro pelo usuário (AGRAWAL; SRIKANT, 1994; GILLMEISTER; CAZELLA, 2007). Na figura 10, exibida a seguir, seu pseudocódigo é apresentado:

```

Procedure Apriori (in D: transactional Database; minsup: minimum support)
1   $L_1 = \{\text{large 1-itemsets}\}$ 
2   $k = 2$ 
3  while  $L_{k-1} \neq \emptyset$  do
4  begin
5       $C_k = \text{apriori\_gen}(L_{k-1})$ 
6      for all transactions  $t$  in  $D$  do
7      begin
8           $C^t = \text{subset}(C_k, t)$ 
9          for all candidates  $c \in C^t$  do
10              $c.\text{count} = c.\text{count} + 1$ 
11         end
12     // 2nd pruning step
13      $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$ 
14      $k = k + 1$ 
15 end

```

Figura 10 - Algoritmo Apriori.

Fonte: Al-Hamamy *et al.* (200-).

O primeiro passo deste algoritmo consiste na obtenção do número de ocorrências dos itens para determinar os *itemsets* frequentes. No passo seguinte, chamado de  $k$ , ocorrem duas fases: na primeira, os *itemsets* frequentes ( $L_{k-1}$ ) definidos no passo anterior são utilizados para conceber os *itemsets* candidatos ( $C_k$ ), usando-se para isso a função *apriori-gen*. Posteriormente, é efetuado um novo escaneamento da base de dados para averiguar o suporte de cada candidato em  $C_k$  (AGRAWAL; SRIKANT, 1994; ROMÃO *et al.*, 1999). A seguir, a figura 11 expõe o pseudocódigo da função *apriori-gen*:

```

apriori_gen ( $L_{k-1}$ );
15  $C_k = \emptyset$ 
16 for all itemsets  $X \in L_{k-1}$  and  $Y \in L_{k-1}$  do
17   if  $X_1 = Y_1 \wedge \dots \wedge X_{k-2} = Y_{k-2} \wedge X_{k-1} < Y_{k-1}$  then begin
18      $C = X_1 X_2 \dots X_{k-1} Y_{k-1}$ 
19     add  $C$  to  $C_k$ 
20   end
21 // 1st pruning step
22 delete candidate itemsets in  $C_k$  whose any subset is not in  $L_{k-1}$ 

```

Figura 11 - Função *apriori-gen* do algoritmo Apriori.

Fonte: Al-Hamamy *et al.* (200-).

A função *apriori-gen* possui duas finalidades, sendo que a primeira baseia-se na formação dos *itemsets* frequentes para então gerar os *itemsets* candidatos  $C_k$ . Sendo assim, os *itemsets* candidatos estarão ordenados lexicograficamente, eliminando-se com isso aqueles

que porventura dispõem de itens iguais. A segunda finalidade encerra-se na poda dos *itemsets* candidatos, partindo-se da premissa de que cada subconjunto de um conjunto de itens frequentes também deve ser frequente, ou seja, caso algum *itemset* candidato gerado contenha um subconjunto não frequente, o mesmo será podado (GILLMEISTER; CAZELLA, 2007).

#### 2.3.2.4 O software de Mineração de Dados WEKA©

A *suite* de mineração WEKA© (*Waikato Enviroment for Knowledge Analysis*) é uma coleção de algoritmos de aprendizado de máquina e de ferramentas de pré e pós-processamento de dados. Foi desenvolvida para propiciar o fácil uso de seus métodos em novos conjuntos de dados, e possui amplo suporte a todo o processo de mineração, desde a preparação à visualização dos dados de entrada e resultados do processo de aprendizagem. Todas suas funcionalidades podem ser acessadas por meio de uma interface comum, de forma a permitir que seus utilizadores identifiquem e comparem os métodos mais adequados para o problema em estudo (WITTEN; FRANK, 2005).

O WEKA© foi desenvolvido na Universidade de Waikato, que se encontra na Nova Zelândia, e sua denominação corresponde também a uma ave endêmica que possui a particularidade de não voar. O sistema é escrito na linguagem de programação Java™, o que o torna utilizável em vários sistemas operacionais (multiplataforma) e é distribuído sob os termos da licença GNU *General Public License* (WITTEN; FRANK, 2005; DAMASCENO, 2010).

Esta aplicação possui quatro formas de utilização, e a principal é a interface gráfica chamada de Explorer, que provê acesso a todas as funcionalidades utilizando, para isso, menus selecionáveis e formulários que são preenchidos durante seu uso. Uma segunda interface, denominada KnowledgeFlow, apresenta um *layout* onde um usuário pode selecionar alguns componentes a partir de uma barra de ferramentas e conectá-los a fim de gerar um fluxograma para representar a análise e processamento dos dados (WITTEN; FRANK, 2005; BOUCKAERT *et. al*, 2012).

A terceira interface gráfica disponível, intitulada Experimenter, permite a um usuário criar, executar, modificar e analisar experimentos de uma maneira mais conveniente do que processando-os individualmente. Exemplificando: um usuário pode criar um experimento

onde se executam vários esquemas para uma série de *datasets* e, através do Experimenter, é possível analisar os resultados obtidos para determinar qual dos esquemas é, estatisticamente, o melhor. A quarta e última forma de utilização, a Simple CLI, consiste numa interface de linha de comando onde é possível executar as classes `weka.jar` diretamente (WITTEN; FRANK, 2005; BOUCKAERT *et. al*, 2012).

A seguir, são apresentados alguns algoritmos existentes na ferramenta e suas definições básicas (WITTEN; FRANK, 2005):

- **Algoritmos de Classificação**
  - NaiveBayes - este algoritmo implementa o classificador probabilístico Naïve Bayes;
  - MultilayerPerceptron - rede neural que utiliza *backpropagation* para treinamento;
  - ZeroR – algoritmo usado para prever uma classe majoritária (se nominal) ou o valor médio (se numérico);
  - J48 - é uma implementação em código Java™ do algoritmo C4.5 existente na ferramenta WEKA© utilizado para gerar árvores de decisão;
- **Algoritmos de Agrupamento**
  - Cobweb - implementa os algoritmos de agrupamento Cobweb e Classit;
  - MakeDensityBasedClusterer - *meta-clusterer* que envolve um algoritmo de agrupamento para fazê-lo retornar distribuição e densidade;
  - SimpleKMeans - *cluster* que utiliza o método k-means;
- **Algoritmos de Associação**
  - PredictiveApriori - permite encontrar regras de associação ordenadas por acurácia preditiva;
  - Tertius - geralmente utilizado para encontrar regras de associação ou de classificação com múltiplas condições no consequente de acordo com uma medida de confirmação.

Cabe ressaltar que o WEKA© utiliza um padrão de arquivo específico desenvolvido pelo departamento de ciência da computação da Universidade de Waikato para a representação de conjuntos de dados denominado *arff* (*attribute-relation file format*), que consiste num arquivo de texto ASCII que descreve uma lista de instâncias de um conjunto de atributos. Um arquivo *arff* é segmentado em duas partes, onde na primeira se encontram o

nome do conjunto de dados e a lista dos atributos bem como o tipo que estes últimos representam, e a segunda compreende os valores que os registros dos atributos possuem. A figura 12, exibida abaixo, exemplifica a estrutura de um arquivo arff (WITTEN; FRANK, 2005; DAMASCENO, 2010):

```
% 1. Title: Iris Plants Database|
%
% 2. Sources:
%   (a) Creator: R.A. Fisher
%   (b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)
%   (c) Date: July, 1988
%
@RELATION iris

@ATTRIBUTE sepallength NUMERIC
@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE petallength NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE class       {Iris-setosa,Iris-versicolor,Iris-virginica}

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
```

Figura 12 - Representação de um arquivo arff.

Fonte: BOUCKAERT *et al.* (2012).

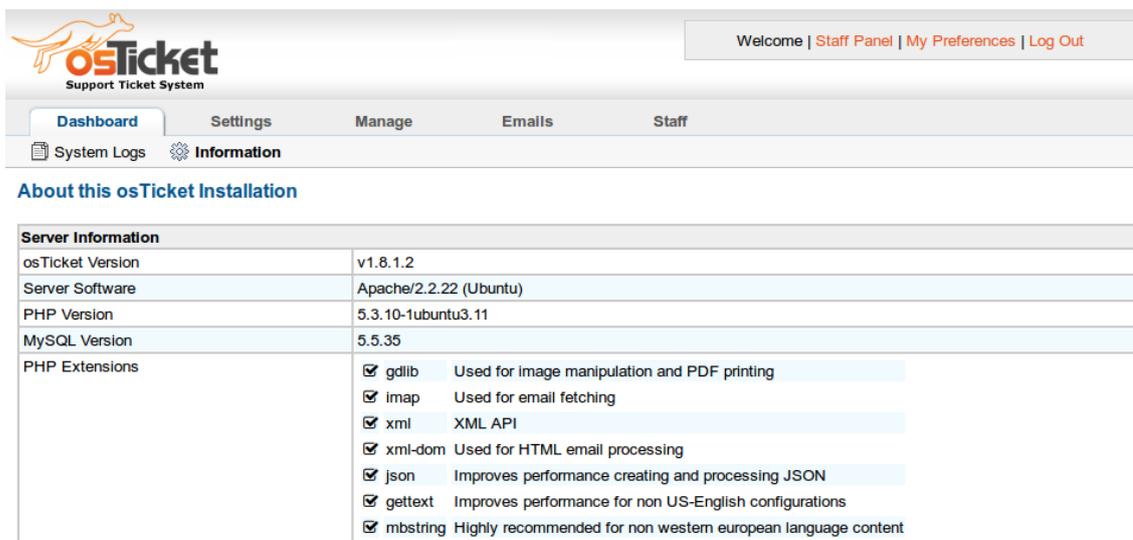
### 3. METODOLOGIA

Neste capítulo, convencionou-se adotar como sinônimos os termos solicitação de serviço, requisição de serviço, ordens de serviço, tíquetes e chamados e, para uma identificação mais inteligível, os nomes dos atributos foram colocados entre aspas duplas ("").

#### 3.1. O SISTEMA DE SOLICITAÇÃO DE SERVIÇOS DE TI

O sistema utilizado pela diretoria de TI da reitoria da instituição federal de ensino para recebimento e tratamento de solicitações de serviços de TI é o osTicket© na versão 1.6 ST, que se caracteriza por ser um sistema *web* de código aberto e é apontado, pelos seus desenvolvedores, como o sistema de suporte de tíquetes mais popular do mundo (osTicket, 2014).

Este sistema utiliza *softwares* e pacotes que são requeridos preliminarmente à sua instalação e essenciais para seu devido funcionamento, e são o servidor *web* Apache™, o interpretador e algumas extensões associadas à linguagem de programação PHP© e o banco de dados MySQL™, como pode ser visto na figura 13.



The screenshot shows the osTicket administration interface. At the top, there is a navigation bar with the osTicket logo and a user menu containing 'Welcome', 'Staff Panel', 'My Preferences', and 'Log Out'. Below the navigation bar, there are tabs for 'Dashboard', 'Settings', 'Manage', 'Emails', and 'Staff'. The 'Information' tab is selected, showing the 'About this osTicket Installation' page. The page contains a table with the following data:

Server Information	
osTicket Version	v1.8.1.2
Server Software	Apache/2.2.22 (Ubuntu)
PHP Version	5.3.10-1ubuntu3.11
MySQL Version	5.5.35
PHP Extensions	<ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> gdlib Used for image manipulation and PDF printing</li> <li><input checked="" type="checkbox"/> imap Used for email fetching</li> <li><input checked="" type="checkbox"/> xmi XML API</li> <li><input checked="" type="checkbox"/> xmi-dom Used for HTML email processing</li> <li><input checked="" type="checkbox"/> json Improves performance creating and processing JSON</li> <li><input checked="" type="checkbox"/> gettext Improves performance for non US-English configurations</li> <li><input checked="" type="checkbox"/> mbstring Highly recommended for non western european language content</li> </ul>

Figura 13 - Tela de informações do sistema osTicket© na versão 1.8.1.2.

Fonte: O autor.

Em se tratando de seu funcionamento, a abertura de solicitações de serviços de TI é realizada por três formas distintas:

- Através do envio de um e-mail, com a descrição da solicitação desejada em seu corpo, para o endereço de correio eletrônico de suporte da diretoria de TI;
- Através de uma página web, conforme consta na figura 14, exibida a seguir;
- Pelo painel de administração do sistema. Entretanto, essa operação é restringida apenas aos responsáveis pelo seu gerenciamento.

The image shows a web interface for a support ticket system. At the top left is the logo of Instituto Federal de Educação, Ciência e Tecnologia Fluminense, consisting of a grid of green squares and a red circle. To the right of the logo is the text 'INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA FLUMINENSE'. In the top right corner, it says 'SUPPORT TICKET SYSTEM'. Below the header is a navigation bar with three links: 'Pagina Inicial' (with a home icon), 'Novo Ticket' (with a document icon), and 'Andamento Ticket' (with a document icon). Below the navigation bar is a message: 'Por favor preencha os dados do formulário abaixo para abrir nova solicitação.' The form contains the following fields: 'Nome Completo:' with a text input field; 'Endereço de Email:' with a text input field; 'Telefone:' with a text input field and 'Ramal' with a smaller text input field; 'Departamento:' with a dropdown menu showing 'Selecione'; 'Assunto:' with a text input field; and 'Mensagem:' with a large text area. At the bottom of the form are three buttons: 'Enviar', 'Limpar', and 'Cancelar'.

Figura 14 - Página de solicitação de serviços via web.

Fonte: O autor.

Assim que um tíquete é criado por qualquer uma das formas apresentadas, é gerado automaticamente um número que corresponde à solicitação efetuada, o que permite o acompanhamento posterior da requisição pelo solicitante bem como a comunicação com o atendente responsável caso seja necessária uma interação entre ambas as partes.

### 3.2. A SELEÇÃO DOS DADOS

Neste sistema, toda e qualquer informação relativa a um tíquete é armazenada em tabelas, tanto as que são informadas pelo usuário solicitante quanto as inseridas pelo atendente durante a resolução da solicitação. A figura 15, apresentada abaixo, expõe as tabelas associadas ao sistema:

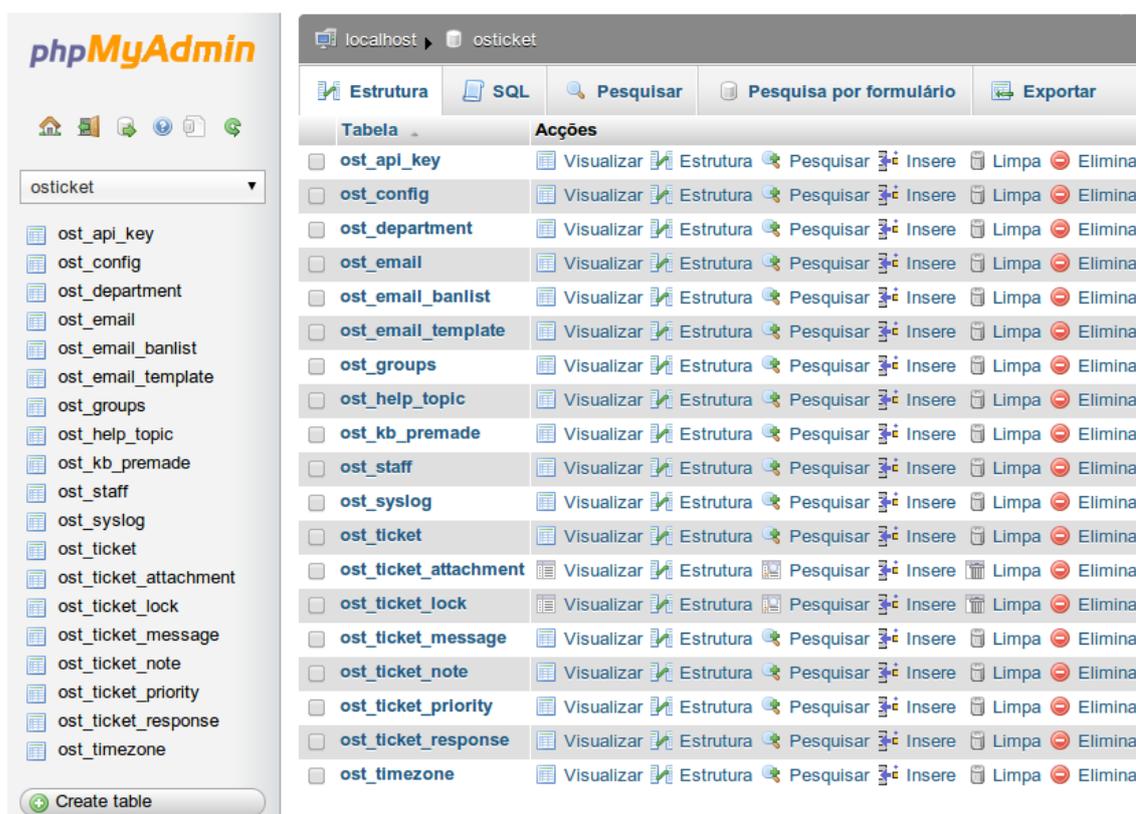


Figura 15 - Tabelas do sistema osTicket©.

Fonte: O autor.

No tocante ao início da seleção dos dados propriamente dita, todas as tabelas e suas respectivas colunas existentes no banco de dados foram avaliadas e três delas foram consideradas as mais promissoras: `ost_ticket`, `ost_ticket_message` e `ost_ticket_response`. Para uma melhor exemplificação do julgamento efetuado, são descritas a seguir as definições de algumas colunas das referidas tabelas que foram estipuladas como as mais relevantes:

- `ost_ticket`: número identificador da solicitação criada, e-mail do solicitante, nome do solicitante, título da solicitação, endereço IP da máquina utilizada para abertura da solicitação, status de atendimento da solicitação, meio utilizado para a abertura da solicitação, data e hora de abertura e data e hora de fechamento;
- `ost_ticket_message`: número identificador da solicitação criada, corpo da mensagem descritiva da solicitação, origem da abertura da solicitação (e-mail, página *web* ou criada pelo painel de administração do sistema) e data e hora de abertura;
- `ost_ticket_response`: número identificador da solicitação criada, nome do responsável técnico pelo atendimento, resposta(s) dada(s) para resolução da

solicitação ou indagações realizadas pelo responsável técnico visando à resolução do atendimento e data e hora em que a(s) resposta(s) foi(foram) efetuadas.

Após esta análise introdutória, foi realizada uma consulta via comandos SQL ao banco de dados do sistema osTicket© para obter os valores numerais ou nominais das colunas das tabelas supracitadas. Esta ação foi processada pelo *software* phpMyAdmin©, que consiste numa ferramenta *web* livre voltada à administração de bancos de dados MySQL™ (phpMyAdmin, 2014) Os referidos comandos SQL são apresentados abaixo e um de seus exemplos exibidos na figura 16:

```
SELECT `ticket_id`, `ticketID`, `email`, `name`, `subject`, `status`, `source`,
`created`, `closed` FROM `ost_ticket` WHERE 1 LIMIT 0, 3000
```

```
SELECT `ticket_id`, `message`, `source`, `created`, `updated` FROM
`ost_ticket_message` WHERE 1 LIMIT 0, 3000
```

```
SELECT `ticket_id`, `staff_name`, `response`, `created` FROM
`ost_ticket_response` WHERE 1 LIMIT 0, 3000
```

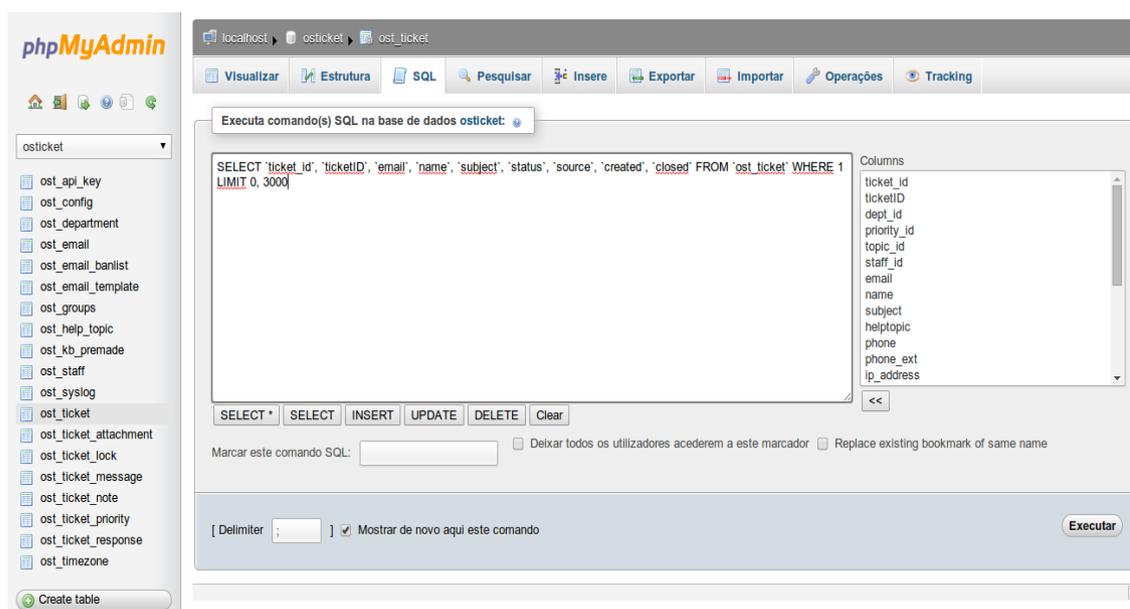


Figura 16 - Consulta SQL realizada no banco de dados do sistema osTicket©.

Fonte: O autor.

Ulteriormente, os dados de cada uma das tabelas foram exportados pela mesma interface de gerenciamento de banco de dados diretamente no formato csv (*comma-separated values*), como pode ser observado na figura 17. Essa operação gerou três arquivos respectivos à cada uma delas, e permitiu obter uma amostra com 1610 solicitações que compreendem os registros abertos desde a implantação do sistema, que ocorreu em 05/06/2012, a 24/02/2014.

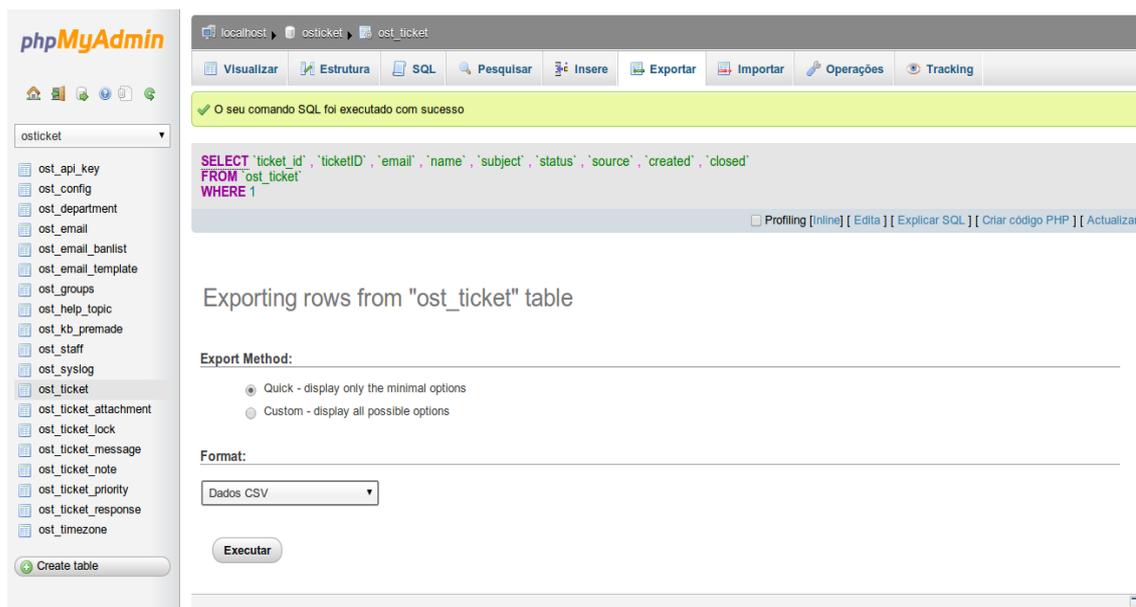


Figura 17 - Tela que exibe a exportação das colunas da tabela ost\_ticket.

Fonte: O autor.

Numa segunda avaliação mais aprofundada, foram utilizados os arquivos concernentes às tabelas ost\_ticket e ost\_ticket\_message para uma concreta determinação das colunas que seriam analisadas a fim de contribuir, posteriormente, à obtenção dos primeiros atributos. Assim sendo, as colunas explicitadas abaixo foram caracterizadas como as mais significativas e agrupadas num único arquivo, sendo elas:

- Número identificador da solicitação;
- E-mail do solicitante;
- Nome do solicitante;
- Título da solicitação;
- Meio utilizado para a abertura das solicitações;
- Data e hora de abertura;
- Data e hora de fechamento;
- Corpo da mensagem descritiva da solicitação.

Quanto às colunas provenientes da tabela ost\_ticket\_response, somente a que abrangia o texto da resolução técnica e/ou questionamentos realizados pelos atendentes de TI para tratamento das solicitações fora mantida. Entretanto, essa coluna foi utilizada exclusivamente para fins de consulta, não sendo, dessa forma, agrupada às demais.

### 3.3. ETAPAS DE PRÉ-PROCESSAMENTO E TRANSFORMAÇÃO DOS DADOS

No início da etapa de pré-processamento, decidiu-se que a coluna relativa ao meio utilizado para a abertura das solicitações, proveniente da tabela *ost\_ticket*, originaria a criação do primeiro atributo escolhido, nomeado como “*meio\_de\_abertura*”. Não foi necessário, para ele, nenhum tipo de normalização, e seus valores nominais são e-mail, *web* e *other* (aberto pelo painel de administração do sistema), e sua distribuição percentual é exibida no gráfico 1. Assim sendo, tais valores foram removidos para um novo arquivo csv, que passou a ter o propósito de armazenar os valores deste e dos futuros atributos assim que cada um deles fosse determinado e normalizado. Cabe ressaltar que foi realizada uma cópia da coluna que continha o número identificador da solicitação, juntamente à coluna do atributo “*meio\_de\_abertura*”, para que se efetuasse uma conferência da numeração dos tíquetes após a união dos demais atributos no novo arquivo csv gerado, evitando assim possíveis equívocos neste processo.

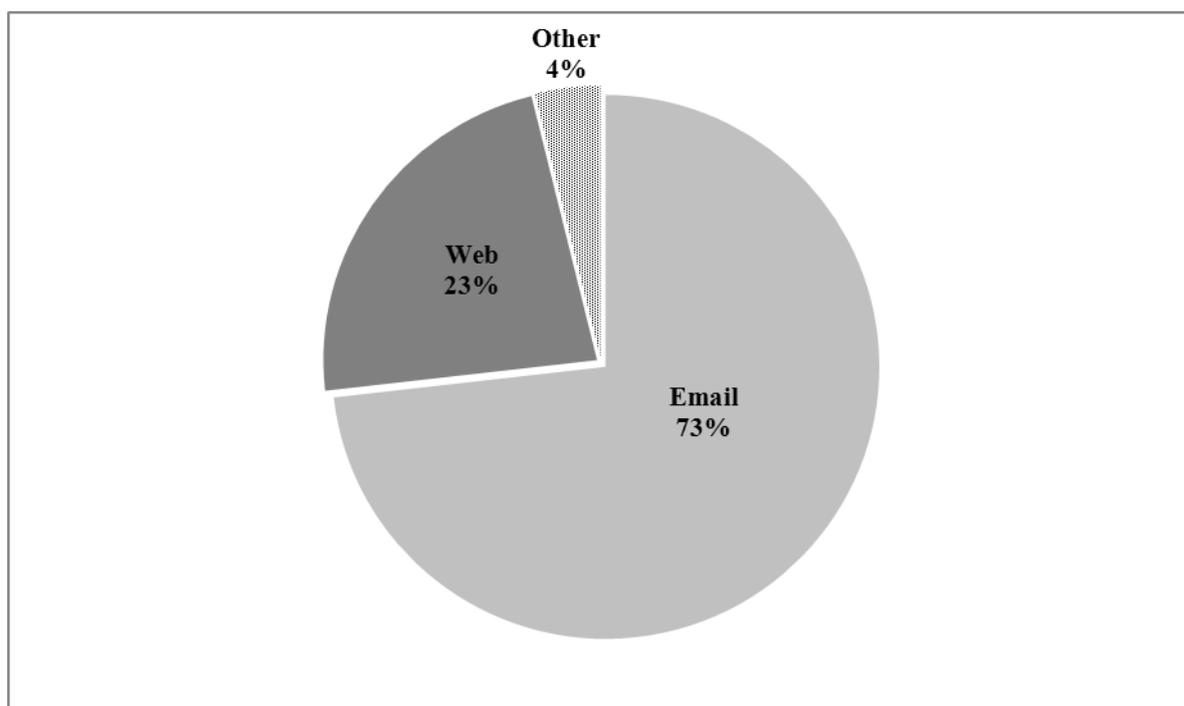


Gráfico 1- Distribuição percentual do atributo “*meio\_de\_abertura*”.

Fonte: O autor.

Diante da diversidade de e-mails e nomes existentes, vislumbrou-se a possibilidade de utilizar os dados dessas colunas para determinar o campus de lotação dos solicitantes. Com isso, as solicitações passariam a se vincular a um dos nove *campi* da instituição ou à reitoria, e

não a uma pessoa especificamente. Tal processo fora viabilizado com o uso do módulo de recursos humanos do sistema unificado de administração pública utilizado pela instituição de ensino, uma vez que propicia encontrar facilmente, e por qualquer servidor, dados relativos aos seus funcionários pesquisando-os por seus nomes. Esse procedimento culminou com a concepção do segundo atributo, intitulado de “campus\_de\_origem”, e novamente os dados do mesmo foram transferidos ao novo arquivo csv, bem como a coluna que continha o número identificador da solicitação para fins de conferência. O gráfico 2, exibido a seguir, expressa a distribuição do atributo “campus\_de\_origem”:

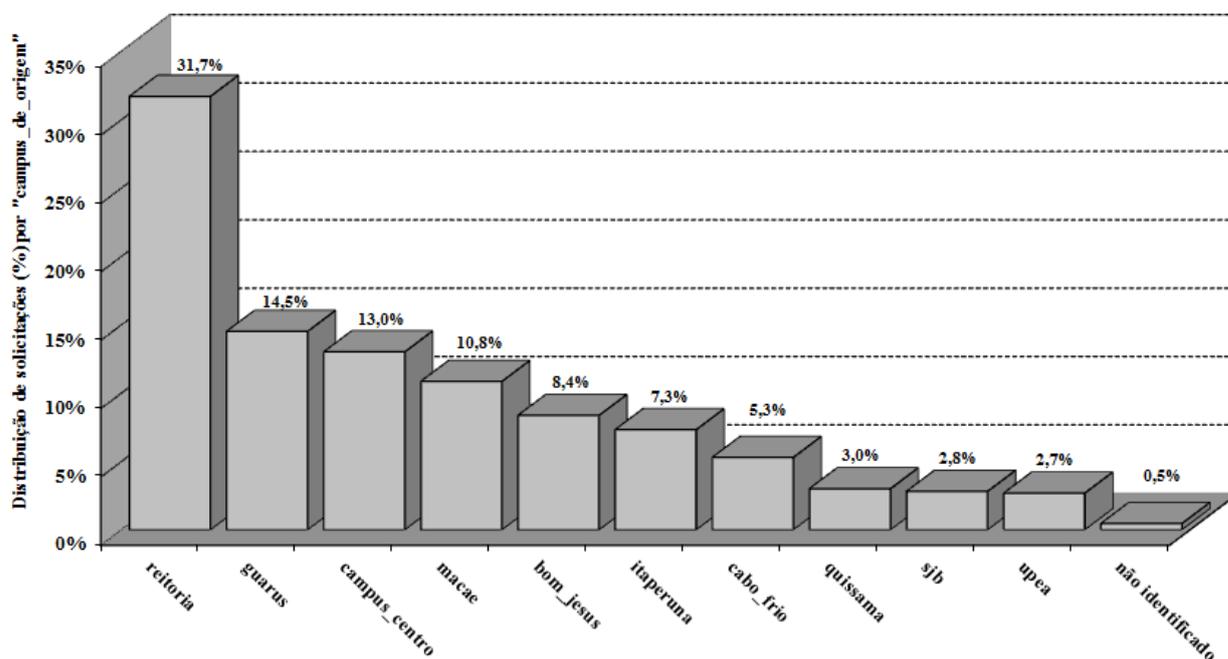


Gráfico 2 – Distribuição percentual do atributo “campus\_de\_origem”.

Fonte: O autor.

Seguindo a ideia de padronização abordada acima, os dados das colunas data e hora de abertura e data e hora de fechamento foram convertidos em intervalos de tempo para aferir o período médio de resolução dos tíquetes. Esse procedimento foi desenvolvido através do emprego das funções existentes no *software* de planilhas Excel© DATADIF e PROCV, e o atributo resultante fora chamado de “tempo\_de\_resolucao”. Na determinação dos valores desse atributo, primeiramente foi aplicada a função DATADIF aos campos data e hora de abertura e data e hora de fechamento para gerar uma coluna chamada de quantidade de dias. Uma vez estabelecida essa quantidade em todos os chamados, aplicou-se a função PROCV para associar os valores da coluna quantidade de dias a uma outra coluna com intervalos numéricos para que a operação resultante exibisse os valores do atributo

“tempo\_de\_resolução” conforme mostrados abaixo:

- 1\_dia;
- 2\_e\_3\_dias;
- 4\_e\_5\_dias;
- 6\_e\_7\_dias;
- 8\_e\_15\_dias;
- 16\_e\_23\_dias;
- 24\_e\_31\_dias;
- >=\_32\_dias.

O gráfico 3, exibido a seguir, explicita a quantidade de solicitações e seu tempo de resolução associado:

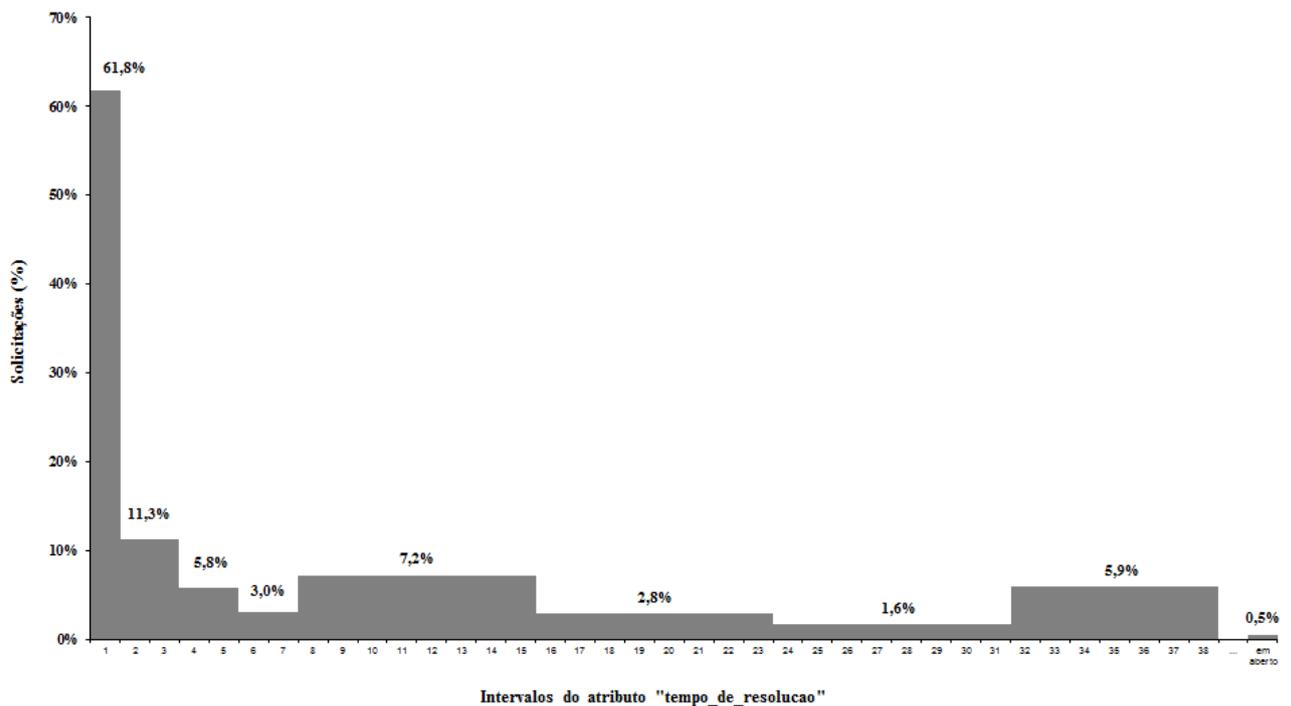


Gráfico 3 - Distribuição percentual das solicitações pelo seu tempo de resolução.

Fonte: O autor.

### 3.3.1 Processo de criação do Modelo de Categorias de Tratamento de Serviços de TI

Neste ponto, cada tíquete passou a ser avaliado pelo corpo da mensagem descritiva da solicitação, e o intuito dessa técnica era identificar e classificar tudo o que fora requisitado

pelos usuários para conceber um modelo de categorias de tratamento de solicitações de serviços de TI. Além disso, ressalta-se que a criação de tal modelo de categorias também permitiu definir, na segunda avaliação completa da amostra que será abordada adiante, os valores nominais correspondentes de três atributos intitulados de “categoria”, “subcategoria” e “objeto\_central”.

Para uma melhor ilustração do acima disposto, supõe-se um caso onde um usuário criou uma ordem de serviço informando no corpo da mensagem descritiva que desejava ter acesso ao sistema de controle acadêmico. Como o objetivo era de criar um modelo de categorias de tratamento de serviços de TI, esta situação hipotética, por corresponder a uma solicitação de acesso a um sistema institucional, originaria a definição da classificação dos atributos “categoria”, “subcategoria” e “objeto\_central” respectivamente em: “sistemas” → “acesso” → “nome do sistema de controle acadêmico”.

Assim sendo, as solicitações iniciais foram originando as primeiras categorias, subcategorias e objetos centrais das requisições feitas. E após atingir cerca de um terço da amostra, os chamados examinados desde o princípio foram reavaliados para corroborar as categorias do modelo proposto até então ou para efetuar modificações e/ou inserções de novas classificações. Destaca-se que esse procedimento fora repetido assim que se alcançaram, também, dois terços da amostra.

Salienta-se que a necessidade de geração de categorias de serviços é decorrente das inúmeras atribuições das coordenações vinculadas à diretoria de TI, e pretende-se com isso fazer com que estas categorias reflitam genericamente tais atribuições, oportunizando depreender as características dos serviços ofertados e/ou atendidos. Já em relação aos atributos “subcategoria” e “objeto\_central”, estes se apresentam como as especificidades dos serviços solicitados. Outra medida que se tornou imprescindível foi o desmembramento de chamados que continham mais de uma requisição no corpo da mensagem descritiva de uma mesma solicitação e que, explicitamente, eram de classificações de categorias dessemelhantes. Isto significa que caso uma ordem de serviço contivesse três solicitações que se associavam claramente a três categorias distintas, um novo registro era criado para cada nova categoria identificada, contudo possuindo o mesmo número identificador da requisição original. Com isso, foi possível manter a vinculação da classificação de cada solicitação da amostra a uma categoria específica.

Ao fim da análise acima exposta, foi factível estabelecer os valores nominais para os atributos “categoria” e “subcategoria” conforme exibidos a seguir:

- “aplicativos” - categoria que engloba todas as requisições de serviços

relativas a aplicativos utilizados tanto para uso dos setores administrativos da instituição quanto em laboratórios voltados ao ensino. Os valores do atributo “subcategoria”, para esta categoria, correspondem a “instalacao” e “licenca”;

- “e-mail” - categoria que agrega todos os tíquetes relacionados a correio eletrônico e os valores associados ao seu atributo “subcategoria” são “acesso”, “criacao” e “suporte”;
- “hardware” - todos os chamados dessa classe, como os concernentes a estações de trabalho, *tablets* e *scanners*, foram assim categorizados e os valores de suas subcategorias encerram-se em “solicitacao” e “suporte”;
- “impressao” - os chamados referentes a impressoras e seus insumos (*toners*, por exemplo) foram assim definidos e suas subcategorias são “solicitacao” e “suporte”;
- “infraestrutura” - as solicitações que englobam redes de computadores e serviços computacionais atrelados, telefonia fixa, móvel e internet, qualificaram-se pertencentes a esse grupo, e suas subcategorias são “indisponibilidade”, “instabilidade”, “solicitacao” e “suporte”;
- “sistemas” - toda e qualquer ordem de serviço alusiva a sistemas de uso institucional receberam essa classificação e suas subcategorias dividem-se em “acesso”, “instalacao” e “suporte”;
- “SO” - grupo de requisições de serviço referentes a sistemas operacionais e foi a única categoria que propiciou a concepção de somente uma subcategoria: “solicitacao”.

Portanto, após a finalização da primeira análise completa da amostra, o modelo criado foi apresentado ao diretor de TI da reitoria para que o mesmo expusesse suas colocações, excluindo com isso qualquer discrepância que destoasse da realidade. Efetuadas as mudanças sugeridas, o modelo de categorias na sua versão final é exposto na tabela 2 (ressalta-se que as ocorrências dos valores do atributo “objeto\_central” já correspondem à verificação final):

Tabela 2 - Modelo de categorias de tratamento de solicitações de serviços de TI.

categoria	subcategoria	objeto_central	quantidade
aplicativos	instalacao	adobe	1
		arduino	2
		asc_timetables	1
		office	1
		tesc	1

	licenca	adobe asc_timetables autocad corel photoshop	1 1 2 1 1
<b>Total desta categoria</b>			<b>12</b>
e-mail	acesso	informacoes	39
		lista	16
		mudanca_de_senha	151
criacao	lista	13	
	usuario	152	
suporte	informacoes	95	
	quota	61	
	redirecionamento	24	
<b>Total desta categoria</b>			<b>551</b>
hardware	solicitacao	desktop	1
		informacoes	3
		notebook	3
		perifericos	12
	suporte	desktop	3
perifericos		3	
servidor		3	
tablet		19	
<b>Total desta categoria</b>			<b>47</b>
impressao	solicitacao	aquisicao	5
		instalacao	6
		insumos	16
	suporte	informacoes	7
insumos		1	
<b>Total desta categoria</b>			<b>35</b>
infraestrutura	indisponibilidade	acesso_a_internet	5
		acesso_a_sistema	5
		link	1
		portais	12
		sistema	3
		telefonia	10
	instabilidade	equipamento	7
		link	4
		servicos_de_rede	2
	solicitacao	acesso_a_rede	2
acesso_rdweb		78	
cabeamento		11	
desbloqueio_de_site		7	
equipamento		4	
servicos_de_rede		31	
telefonia	2		

		videoconferencia	7
	suporte	informacoes	9
		servicos_de_rede	75
		telefonica	45
		videoconferencia	11
<b>Total desta categoria</b>			<b>331</b>
sistemas	acesso	atualizacao_cadastral	4
		iff_rotinas	15
		lime_survey	1
		luma	64
		ponto_biometrico	6
		portais	7
		portal_capes	9
		portal_institucional	1
		qualidata	6
		SAC	1
		sigas	2
		sisep	1
		sistema_de_inscricao	1
		sistema_de_projetos	1
		suap	37
		sistemas	instalacao
lime_survey	1		
luma	5		
ponto_biometrico	2		
qualidata	1		
sisep	1		
universidade_petrobras	1		
sistemas	suporte	atualizacao_cadastral	5
		banco_de_mobilidade	1
		cap_compras	1
		gitlab	1
		iff_rotinas	2
		informa	23
		informacoes	1
		luma	7
		moodle	8
		ponto_biometrico	35
		portais	15
		portais_de_eventos	7
		portal_institucional	117
		qualidata	12
		SAC	1
		sigas	3
		sisep	15
sistema_de_gestao	9		

		sistema_de_inscricao	14
		sistema_de_projetos	1
		suap	87
		universidade_petrobras	5
<b>Total desta categoria</b>			<b>539</b>
SO	solicitacao	instalacao	7
		licenca	7
		informacoes	1
<b>Total desta categoria</b>			<b>15</b>
<b>Total geral</b>			<b>1530</b>

### 3.3.2 Processo de classificação das solicitações de serviços segundo o modelo criado

Posteriormente à criação do modelo de categorias de tratamento de solicitações de serviços de TI, todas as requisições foram reexaminadas com a finalidade de efetivamente classificá-las segundo o modelo elaborado. Porém, a classificação da categoria da solicitação era consumada mediante a averiguação do conteúdo da resolução prestada pelo responsável do atendimento, extraído da tabela `ost_ticket_response`, associando-o à categoria mais próxima que estivesse definida no modelo de categorias de tratamento. Ou seja: a primeira avaliação de todos os tíquetes foi realizada para criar o modelo de categorias de tratamento de solicitações de serviços de TI de acordo com o corpo da mensagem descritiva informada pelo usuário, e a segunda avaliação completa destinou-se a classificar as requisições observando a resolução efetuada pelo atendente técnico do tíquete relacionando-a a uma das categorias do modelo proposto na primeira avaliação.

Tal procedimento foi adotado pois nem sempre a classificação da categoria do que um usuário solicita corresponde exatamente à classificação da categoria da resolução do atendente. E como a ideia é de ter uma compreensão mais apurada do que é ofertado e/ou atendido, essa técnica se apresentou como a mais coerente.

Contudo, acentua-se que a classificação das solicitações conforme acima apresentado ocorreria apenas após a observação de três premissas, expressas abaixo segundo sua ordem de relevância:

1. Qualquer solicitação que correspondesse a uma indisponibilidade, seja de acesso à internet ou a um sistema, assim seria classificada. O motivo para tal é de promover subsídios caso haja uma futura utilização dos processos de Gerenciamento

de Incidente e Gerenciamento de Problema da ITIL®;

2. Nos casos em que o responsável pelo atendimento não expusesse com clareza o que fizera na resolução, o que impedia uma correta classificação, a categoria da solicitação se enquadraria segundo o que o usuário solicitara;

3. As requisições que o atendente não realizava nenhuma ação resolutive eram categorizadas segundo o que era inserido no corpo descritivo da solicitação, como nas situações onde o usuário solicitante não criava um tíquete por seu e-mail institucional, o que não possibilitava a garantia de sua autenticidade, e era orientado a refazer sua solicitação por esta via.

Devido à possibilidade de se fornecerem subsídios para a implantação dos processos da ITIL® supracitados, vislumbrou-se a viabilidade de classificar as ordens de serviço segundo alguns conceitos existentes no estágio de ciclo de vida do serviço *Service Operation*. Esta prática desencadeou a criação do último atributo, denominado “tipo\_de\_chamado”, e seus valores nominais correspondem a “incidente”, “demanda\_interna”, “evento” ou “requisicao\_de\_servico”. Assim sendo, as requisições eram classificadas em relação a este último atributo prioritariamente em:

- “incidente” - qualquer solicitação que se constatasse no corpo da mensagem descritiva ou de acordo com a resposta do atendente a uma falha ou redução da qualidade existente de um serviço/sistema;
- “demanda\_interna” - qualquer chamado que fosse aberto por um membro da diretoria de TI direcionado a outro da mesma equipe sobre serviços/sistemas somente por eles utilizados/administrados. Todavia, caso uma solicitação fosse criada para reportar uma falha de um serviço de monitoramento de uma rede de computadores por exemplo, esta seria classificada como incidente;
- “evento” - qualquer requisição aberta mediante uma ocorrência que pode ocasionar um incidente caso não seja devidamente tratada, como o limite de quota excedido de uma conta de e-mail;
- “requisicao\_de\_servico” - toda ordem de serviço que não se enquadra nos casos anteriores.

Abaixo, é apresentado o gráfico 4 que expressa a relação percentual do atributo “tipo\_de\_chamado” com todas as ordens de serviço efetuadas:

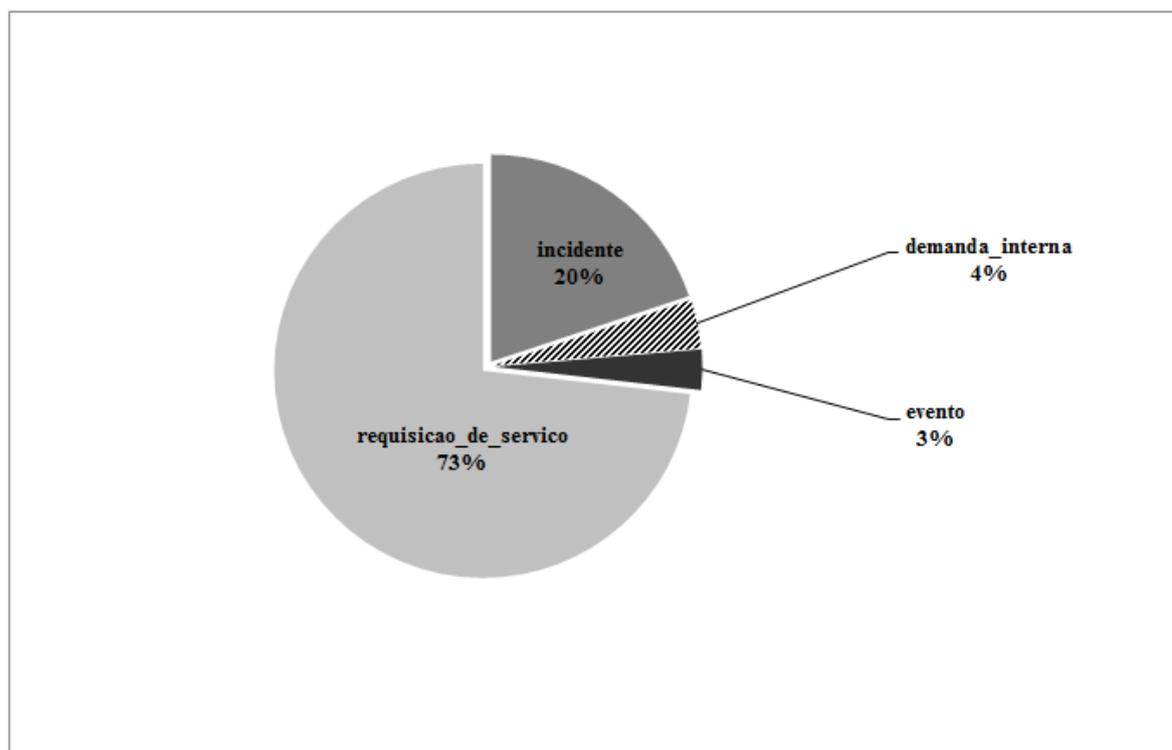


Gráfico 4 - Distribuição percentual do atributo “tipo\_de\_chamado”.

Fonte: O autor.

Por último, foram praticadas algumas correções na base, como a inserção do caractere “?” nos campos do atributo “campus\_de\_origem” que não puderam ser determinados, e nos campos do atributo “tempo\_de\_resolucao” que estavam vazios, pois alguns chamados ainda não haviam sido resolvidos e não possuíam, dessa forma, uma data e hora de fechamento. Ainda, houve a exclusão de alguns tíquetes da base devido a diversos fatores, tais como solicitações duplicadas, pedidos não pertinentes à área de TI, pedidos de cancelamento de solicitações anteriores, tíquetes abertos mediante envio de SPAM, etc. Subtraindo-se os tíquetes desses casos mencionados, a amostra passou a ser de 1530 ordens de serviço

Finalizadas as etapas de pré-processamento e transformação dos dados, os sete atributos obtidos foram revistos e validados para a continuidade da pesquisa, estando organizados na base na seguinte ordem: “categoria”, “subcategoria”, “objeto\_central”, “tipo\_de\_chamado”, “campus\_de\_origem”, “meio\_de\_abertura” e “tempo\_de\_resolucao”.

### **3.4. ETAPA DE MINERAÇÃO DE DADOS**

O *software* adotado para a realização da mineração de dados é o WEKA© e foram utilizados a tarefa de associação, o método de regras de associação e o algoritmo Apriori para a mineração dos itens frequentes determinados.

Cabe salientar que vários arranjos entre os atributos foram testados, e somente aqueles que apresentaram informações mais relevantes foram considerados.

## 4. RESULTADOS

Primeiramente, são elencadas as regras que apresentam as correlações mais expressivas entre os atributos "categoria" e "tipo\_de\_chamado". Além disso, são exibidos os números de ocorrências dos atributos em cada transação e o respectivo índice de confiança e suporte mínimo. Ressalta-se que a geração das regras de associação foi realizada em 3 etapas, e seus resultados são expressos na tabela 3:

Tabela 3 - Regras de associação dos atributos “categoria” → “tipo\_de\_chamado”.

antecedente			consequente			
	categoria		tipo_de_chamado		conf.	sup.
1	e-mail	551	requisicao_de_servico	459	0.83	0.1
2	sistemas	539	requisicao_de_servico	420	0.78	0.1
3	infraestrutura	331	requisicao_de_servico	162	0.49	0.1
4	infraestrutura	331	incidente	125	0.38	0.05
5	sistemas	539	incidente	115	0.21	0.05
6	impressao	35	requisicao_de_servico	33	0.94	0.01
7	hardware	47	requisicao_de_servico	27	0.57	0.01
8	hardware	47	incidente	18	0.38	0.01
9	infraestrutura	331	demanda_interna	44	0.13	0.01
	tipo_de_chamado		categoria		conf.	sup.
10	requisicao_de_servico	1120	e-mail	459	0.41	0.1
11	requisicao_de_servico	1120	sistemas	420	0.38	0.1
12	requisicao_de_servico	1120	infraestrutura	162	0.14	0.1
13	incidente	304	infraestrutura	125	0.41	0.05
14	incidente	304	sistemas	115	0.38	0.05
15	evento	48	e-mail	48	1	0.01
16	demanda_interna	58	infraestrutura	44	0.76	0.01
17	incidente	304	e-mail	43	0.14	0.01

A primeira mineração realizada, com o índice de confiança configurado para ser maior ou igual a 10% e suporte mínimo em 10%, apresentou somente 6 regras. Em se tratando das

mesmas, observa-se que para as categorias e-mail, sistemas e infraestrutura o tipo de chamado correlacionado mais frequente é “requisicao\_de\_servico”, com uma taxa de 83%, 78% e 49% de associação. Já observando a relação “tipo\_de\_chamado” → “categoria”, é possível corroborar as constatações supracitadas, uma vez que tais categorias compreendem o maior quantitativo de chamados do tipo “requisicao\_de\_servico”, totalizando juntas cerca de 93%.

Já para a segunda mineração, configurada com um índice de confiança maior ou igual a 10% e suporte mínimo em 5%, mais 4 regras foram geradas, e verifica-se que 38% e 21% das ordens de serviço vinculadas às categorias infraestrutura e sistemas tratam-se de incidentes, e do total de 304 incidentes identificados, 41% são respectivos à categoria infraestrutura e 38% à sistemas.

Ainda relativo à tabela 3 foi realizada uma terceira mineração, com taxas de confiança maior ou igual a 10% e suporte de 1%, para prospecção de transações com menos registros na base de dados, e com isso constatou-se algumas particularidades, tais como: 94% das solicitações sobre impressão são “requisicao\_de\_servico”, 38% das solicitações sobre *hardware* são “incidente” e 44 dos 331 chamados de infraestrutura são “demanda\_interna”.

Posteriormente, realizou-se a mineração dos atributos “categoria” x “campus\_de\_origem”, e a tabela 4, expressa a seguir, exhibe as regras de associação geradas:

Tabela 4 - Regras de associação dos atributos “categoria” → “campus\_de\_origem”.

antecedente		consequente				
	campus_de_origem	categoria	conf.	sup.		
1	reitoria	485	sistemas	251	0.52	0.1
2	guarus	222	e-mail	126	0.57	0.05
3	campos_centro	199	e-mail	107	0.54	0.05
4	reitoria	485	infraestrutura	136	0.28	0.05
5	quissama	46	e-mail	29	0.63	0.01
6	upea	41	infraestrutura	21	0.51	0.01
7	cabofrio	81	e-mail	38	0.47	0.01
8	bom_jesus	128	e-mail	57	0.45	0.01
9	macae	166	e-mail	63	0.38	0.01
10	itaperuna	112	e-mail	42	0.38	0.01
11	campos_centro	199	sistemas	69	0.35	0.01
12	itaperuna	112	sistemas	34	0.3	0.01
13	macae	166	sistemas	47	0.28	0.01
14	bom_jesus	128	sistemas	35	0.27	0.01
15	cabofrio	81	sistemas	22	0.27	0.01
16	guarus	222	sistemas	55	0.25	0.01
17	cabofrio	81	infraestrutura	20	0.25	0.01
18	itaperuna	112	infraestrutura	26	0.23	0.01
19	bom_jesus	128	infraestrutura	25	0.2	0.01
20	macae	166	infraestrutura	31	0.19	0.01
21	guarus	222	infraestrutura	34	0.15	0.01
22	reitoria	485	e-mail	70	0.14	0.01
	categoria	campus_de_origem	conf.	sup.		

23	sistemas	539	reitoria	251	0.47	0.1
24	infraestrutura	331	reitoria	136	0.41	0.05
25	e-mail	551	guarus	126	0.23	0.05
26	e-mail	551	campos_centro	107	0.19	0.05
27	hardware	47	macae	16	0.34	0.01

Seguindo o mesmo conceito já apresentado, efetuou-se uma primeira mineração com taxa de confiança maior ou igual a 10% e suporte mínimo também com 10%, e somente duas regras foram obtidas. Com isso, certifica-se que 52% das ordens de serviço oriundas da reitoria eram relativas a sistemas, o que comprova que a mesma é a maior solicitadora de suporte para chamados classificados nesta categoria, conforme aprova a regra de número 23.

Na segunda mineração, com o índice de confiança mantido em 10% e suporte mínimo decrescido para 5%, 6 regras foram criadas, e as mesmas indicam que a maior parte das requisições provenientes dos *campi* Guarus e Campos-Centro são referentes a e-mail, que juntos, somam 42% do total de solicitações desta natureza, conforme apontado nas regras 25 e 26. Outro destaque é que dos 485 chamados derivados da reitoria, 136 pertenciam à categoria infraestrutura, ou seja, 41% (136 chamados de 331) do total desta categoria.

Na terceira mineração, com o índice de confiança mantido e suporte mínimo decrescido para 1%, 19 regras foram produzidas e as mesmas mostram que 3 grupos de categorias são os mais solicitados:

- e-mail: os *campi* Quissamã, Cabo Frio, Bom Jesus, Macaé e Itaperuna, além dos já apontados pela segunda mineração, são os maiores solicitadores de chamados concernentes à essa categoria;
- sistemas: os *campi* Campos-Centro, Itaperuna, Macaé, Bom Jesus, Cabo Frio e Guarus, assim como a reitoria, figuram como seus maiores utilizadores;
- infraestrutura: terceiro grupo de categoria mais utilizado, e de todos os *campi*, somente Quissamã e Campos-Centro não estão listados como seus maiores requisitantes.

Abaixo, é apresentada a associação dos atributos “categoria” x “meio de abertura”, e suas regras associativas são expostas na tabela 5:

Tabela 5 - Regras de associação dos atributos “categoria” → “meio\_de\_abertura”.

antecedente			consequente			
	categoria		meio_de_abertura	conf.	sup.	
1	sistemas	539	Email	436	0.81	0.1
2	infraestrutura	331	Email	254	0.77	0.1
3	e-mail	551	Email	357	0.65	0.1
4	e-mail	551	Web	187	0.34	0.1

5	sistemas	539	Web	92	0.17	0.05
6	impressao	35	Email	27	0.77	0.01
7	hardware	47	Email	29	0.62	0.01
8	hardware	47	Web	15	0.32	0.01
9	infraestrutura	331	Web	45	0.14	0.01
	<b>meio_de_abertura</b>		<b>categoria</b>		<b>conf.</b>	<b>sup.</b>
10	Web	351	e-mail	187	0.53	0.1
11	Email	1120	sistemas	436	0.39	0.1
12	Email	1120	e-mail	357	0.32	0.1
13	Email	1120	infraestrutura	254	0.23	0.1
14	Web	351	sistemas	92	0.26	0.05
15	Other	59	infraestrutura	32	0.54	0.01
16	Web	351	infraestrutura	45	0.13	0.01

Assim como ocorreu nos casos anteriores, para a tabela 5 foram efetuadas 3 minerações. A primeira possuía 10% ou mais de confiança e de suporte mínimo, e com essa configuração 8 regras foram criadas. A partir delas, analisa-se que “Email” é a via mais empregada para a abertura de chamados relativos a sistemas, infraestrutura e e-mail, que unidas perfazem 94% de 1120 ordens de serviço. Outro ponto a se considerar é que 34% dos chamados que correspondiam à categoria e-mail foram produzidos pela página *web*, que por si só atingia 53% de todas as solicitações abertas via *web*.

A segunda mineração, mantida com taxa maior ou igual a 10% de confiança e suporte mínimo reduzido a 5%, originou apenas duas regras, que indicavam que 17% dos chamados caracterizados como sistemas foram abertos via *web*, e esta categoria correspondia a 26% do total de 351 solicitações.

Na terceira mineração, a taxa de confiança fora mantida e o suporte mínimo limitado a 1%, e com isso 6 regras foram geradas. Tais regras indicam que 77% e 62% das solicitações referentes a impressão e *hardware*, respectivamente, foram abertas por e-mail, e que 32 e 14 por cento dos chamados de *hardware* e infraestrutura foram criadas pela *web*. Já para o meio de abertura *Other*, utilizado somente pela equipe de TI da reitoria, verifica-se que de todos os 59 chamados concebidos por este meio, 32 eram referentes à categoria infraestrutura.

Outra importante verificação foi a observação da correlação do atributo “categoria” e o tempo de resolução para cada um de seus valores. A tabela 6, exibida abaixo, mostra as regras associativas desse contexto:

Tabela 6 - Regras de associação dos atributos “categoria” → “tempo\_de\_resolucao”.

antecedente		consequente				
	categoria		tempo_de_resolucao	conf.	sup.	
1	e-mail	551	1_dia	416	0.75	0.1
2	sistemas	539	1_dia	335	0.62	0.1
3	infraestrutura	331	1_dia	149	0.45	0.05

4	hardware	47	1_dia	23	0.49	0.01
5	infraestrutura	331	>=_32_dias	45	0.14	0.01
6	infraestrutura	331	2_e_3_dias	44	0.13	0.01
7	e-mail	551	2_e_3_dias	60	0.11	0.01
8	infraestrutura	331	8_e_15_dias	36	0.11	0.01
	<b>tempo_de_resolucao</b>		<b>categoria</b>		<b>conf.</b>	<b>sup.</b>
9	1_dia	946	e-mail	416	0.44	0.1
10	1_dia	946	sistemas	335	0.35	0.1
11	1_dia	946	infraestrutura	149	0.16	0.05
12	6_e_7_dias	46	sistemas	25	0.54	0.01
13	>=_32_dias	90	infraestrutura	45	0.5	0.01
14	4_e_5_dias	89	sistemas	37	0.42	0.01
15	16_e_23_dias	43	infraestrutura	17	0.4	0.01
16	8_e_15_dias	110	sistemas	41	0.37	0.01
17	2_e_3_dias	173	e-mail	60	0.35	0.01
18	8_e_15_dias	110	infraestrutura	36	0.33	0.01
19	2_e_3_dias	173	sistemas	56	0.32	0.01
20	4_e_5_dias	89	e-mail	24	0.27	0.01
21	2_e_3_dias	173	infraestrutura	44	0.25	0.01
22	8_e_15_dias	110	e-mail	25	0.23	0.01
23	4_e_5_dias	89	infraestrutura	20	0.22	0.01
24	>=_32_dias	90	sistemas	20	0.22	0.01

Novamente, as regras de associação foram determinadas em 3 estágios. Primeiramente, a definição da confiança baseou-se num índice maior ou igual a 10%, o mesmo para o suporte mínimo. Sendo assim, foram concebidas 4 regras, que mostraram que 75% e 62% das requisições relativas às categorias e-mail e sistemas foram atendidas em um dia. Considerando que essas categorias constituem as mais numerosas, ressalta-se que as duas correspondem a 79% do total de chamados resolvidos neste período.

Num segundo momento, manteve-se a confiança em 10% e o suporte mínimo fora limitado a 5%, e assim sendo apenas 2 regras foram criadas, que apontam que 45% das solicitações que envolviam a categoria infraestrutura foram solucionadas num único dia e que ela corresponde a 16% do total de chamados atendidos com este período.

Por último, o suporte mínimo foi reduzido a 1% e a confiança mantida em 10%, e dessa forma, 16 regras foram originadas. Algumas ponderações relevantes sobre tais regras destacam que: a categoria *hardware* possui quase metade de suas solicitações atendidas num único dia; a categoria infraestrutura possui, de suas 331 ordens de serviço vinculadas, 45 atendidas em mais de 32 dias, 44 entre 2 e 3 dias e 36 entre 8 e 15 dias; e a categoria e-mail possui o intervalo de 2 e 3 dias, em termos quantitativos, como seu segundo tempo médio de resolução.

Ainda analisando a tabela 6, assegura-se que:

- As categorias e-mail, sistemas e infraestrutura correspondem a 44%,

35% e 16% do total de 946 solicitações atendidas em um dia;

- As categorias e-mail, sistemas e infraestrutura correspondem, no total de 173 atendimentos entre 2 e 3 dias, com uma percentagem de 35, 32 e 25 por cento respectivamente;
- As categorias sistemas (37%), infraestrutura (33%) e e-mail (23%) são as categorias mais atendidas entre 8 e 15 dias (110 tíquetes);
- A categoria infraestrutura corresponde por si só a 40% do total de 43 chamados atendidos entre 16 e 23 dias;
- As categorias infraestrutura e sistemas agregam 72% dos tíquetes atendidos acima de 32 dias (90 tíquetes).

Abaixo, são expostas na tabela 7 as regras de associação que correspondem aos atributos “tempo\_de\_resolucao” e “tipo\_de\_chamado”:

Tabela 7 - Regras de associação dos atributos “tempo\_de\_resolucao” → “tipo\_de\_chamado”.

antecedente			consequente			
	tempo_de_resolucao		tipo_de_chamado		conf.	sup.
1	1_dia	946	requisicao_de_servico	747	0.79	0.1
2	2_e_3_dias	173	requisicao_de_servico	115	0.66	0.05
3	1_dia	946	incidente	144	0.15	0.05
4	4_e_5_dias	89	requisicao_de_servico	65	0.73	0.01
5	6_e_7_dias	46	requisicao_de_servico	32	0.7	0.01
6	8_e_15_dias	110	requisicao_de_servico	74	0.67	0.01
7	16_e_23_dias	43	requisicao_de_servico	26	0.6	0.01
8	>=_32_dias	90	requisicao_de_servico	42	0.47	0.01
9	>=_32_dias	90	incidente	37	0.41	0.01
10	16_e_23_dias	43	incidente	15	0.35	0.01
11	8_e_15_dias	110	incidente	29	0.26	0.01
12	2_e_3_dias	173	incidente	43	0.25	0.01
13	4_e_5_dias	89	incidente	19	0.21	0.01
	tipo_de_chamado		tempo_de_resolucao		conf.	sup.
14	requisicao_de_servico	1120	1_dia	747	0.67	0.1
15	incidente	304	1_dia	144	0.47	0.05
16	requisicao_de_servico	1120	2_e_3_dias	115	0.1	0.05
17	evento	48	1_dia	34	0.71	0.01
18	demanda_interna	58	1_dia	21	0.36	0.01
19	incidente	304	2_e_3_dias	43	0.14	0.01
20	incidente	304	>=_32_dias	37	0.12	0.01

Assim como nas tabelas anteriores, as regras de associação foram concebidas em 3 etapas. A primeira baseou-se na definição da confiança num índice maior ou igual a 10%, assim como para o suporte mínimo. Dessarte, foram geradas 2 regras que expuseram que 79% dos tíquetes resolvidos em um dia tratam-se do tipo “requisicao\_de\_servico” e que 67% de todas as ordens de serviço deste tipo foram resolvidos neste intervalo.

Na segunda etapa, manteve-se a confiança em 10% e o suporte mínimo fora reduzido a 5%, e assim sendo 4 regras foram criadas, que apontam que 66% das solicitações atendidas entre 2 e 3 dias consistiam em “requisicao\_de\_servico” e que de todos os chamados atendidos em um dia, 144 deles (15%) são incidentes.

Na terceira etapa, a confiança permaneceu inalterada e o suporte mínimo foi decrescido a 1% e, dessa forma, 14 regras foram originadas, que comprovam algumas relações como as destacadas abaixo:

- 65 (73%) de 89 ordens de serviço atendidas entre 4 e 5 dias são “requisicao\_de\_servico”;
- 70% dos tíquetes resolvidos entre 6 e 7 dias são “requisicao\_de\_servico”;
- 67% dos tíquetes atendidos entre 8 e 15 dias são “requisicao\_de\_servico”;
- 37 (41%) de 90 solicitações atendidas acima de 31 dias são incidentes, e estas geralmente correspondem a problemas relacionados a serviços de rede, telefonia e links instáveis ou inoperantes;
- 35% das solicitações resolvidas entre 16 e 23 dias são incidentes;
- 25% das ordens de serviço solucionadas entre 2 e 3 dias são incidentes.

Visando a obter regras mais especializadas, foi efetuado o arranjo dos atributos “categoria”, “subcategoria”, “objeto\_central” com todos os demais. A tabela 8, a primeira deste tipo, apresenta a correlação dos atributos acima mencionados com o atributo “tipo\_de\_chamado”:

Tabela 8 - Correlação dos atributos “categoria”, “subcategoria”, “objeto\_central” → “tipo\_de\_chamado”.

	antecedente				consequente			
	categoria	subcategoria	objeto_central		tipo_de_chamado	conf.	sup.	
1	impressao	solicitacao	insumos	16	requisicao_de_servico	16	1	0.01
2	sistemas	acesso	iff_rotinas	15	requisicao_de_servico	15	1	0.01
3	sistemas	acesso	luma	64	requisicao_de_servico	63	0.98	0.01
4	e-mail	criacao	usuario	152	requisicao_de_servico	149	0.98	0.01
5	sistemas	acesso	suap	37	requisicao_de_servico	36	0.97	0.01
6	e-mail	acesso	mudanca_de_senha	151	requisicao_de_servico	145	0.96	0.01
7	e-mail	suporte	redirecionamento	24	requisicao_de_servico	23	0.96	0.01
8	e-mail	acesso	informacoes	39	requisicao_de_servico	36	0.92	0.01
9	infraestrutura	solicitacao	acesso_rdweb	78	requisicao_de_servico	71	0.91	0.01
10	sistemas	suporte	suap	87	requisicao_de_servico	74	0.85	0.01
11	hardware	suporte	tablet	19	incidente	16	0.84	0.01
12	sistemas	suporte	portal_institucional	117	requisicao_de_servico	97	0.83	0.01
13	e-mail	suporte	informacoes	95	requisicao_de_servico	74	0.78	0.01

14	e-mail	suporte	quota	61	evento	46	0.75	0.01
15	infraestrutura	suporte	telefonica	45	requisicao_de_servico	27	0.6	0.01
16	sistemas	suporte	ponto_biometrico	35	incidente	21	0.6	0.01
17	infraestrutura	suporte	servicos_de_rede	75	incidente	43	0.57	0.01
18	infraestrutura	solicitacao	servicos_de_rede	31	requisicao_de_servico	16	0.52	0.01
19	infraestrutura	solicitacao	servicos_de_rede	31	demanda_interna	15	0.48	0.01
20	infraestrutura	suporte	telefonica	45	incidente	18	0.4	0.01
21	infraestrutura	suporte	servicos_de_rede	75	demanda_interna	25	0.33	0.01
22	e-mail	suporte	informacoes	95	incidente	20	0.21	0.01
23	sistemas	suporte	portal_institucional	117	incidente	20	0.17	0.01

Os resultados expressos nesta tabela possuem uma taxa de confiança determinada em 10% e o suporte mínimo em 1%, e a constatação inicial a ser destacada remete-se ao fato de que todas as solicitações de insumos de impressão e de acesso ao sistema iff\_rotinas tratam-se de requisições de serviço. Com percentuais altos, de 98 e 97 por cento, as ordens de serviço atreladas às solicitações de acesso (cadastro) ao sistema luma e suap igualmente se caracterizam por serem requisições de serviço.

Em relação à categoria e-mail, os tíquetes referentes à criação de usuário, mudança de senha, redirecionamento e de prestação de informações de acesso, na sua ampla maioria, consistem em chamados definidos como “requisicao\_de\_servico”.

Quanto à categoria infraestrutura, as solicitações de acesso ao rdweb, de suporte à telefonia e de solicitação de serviços de rede apresentam 91%, 60% e 52% de chamados que consistem em “requisicao\_de\_servico”. Entretanto, até pelo caráter da subcategoria ser de assistência, 57% dos tíquetes de suporte a serviços de rede corresponderam a “incidente”, bem como 40% dos tíquetes de suporte à telefonia.

O segundo arranjo mais específico formado corresponde à obtenção das correlações dos atributos “categoria”, “subcategoria”, “objeto\_central” → “campus\_de\_origem”, e seus resultados são expressos na tabela 9:

Tabela 9 - Correlação dos atributos “categoria”, “subcategoria”, “objeto\_central” → “campus\_de\_origem”.

	antecedente			consequente				
	categoria	subcategoria	objeto_central	campus_de_origem	conf.	sup.		
1	sistemas	suporte	portal_institucional	117	reitoria	111	0.95	0.01
2	infraestrutura	suporte	servicos_de_rede	75	reitoria	66	0.88	0.01
3	infraestrutura	solicitacao	servicos_de_rede	31	reitoria	22	0.71	0.01
4	sistemas	acesso	luma	64	reitoria	35	0.55	0.01
5	sistemas	suporte	ponto_biometrico	35	macae	15	0.43	0.01
6	sistemas	acesso	suap	37	campos_centro	15	0.41	0.01
7	sistemas	suporte	suap	87	reitoria	27	0.31	0.01
8	e-mail	criacao	usuario	152	guarus	46	0.3	0.01
9	e-mail	suporte	quota	61	campos_centro	17	0.28	0.01
10	infraestrutura	solicitacao	acesso_rdweb	78	macae	21	0.27	0.01

11	e-mail	suporte	informacoes	95	guarus	22	0.23	0.01
12	e-mail	suporte	informacoes	95	campos_centro	21	0.22	0.01
13	e-mail	criacao	usuario	152	campos_centro	32	0.21	0.01
14	infraestrutura	solicitacao	acesso_rdweb	78	guarus	16	0.21	0.01
15	e-mail	suporte	informacoes	95	reitoria	19	0.2	0.01
16	sistemas	suporte	suap	87	itaperuna	17	0.2	0.01
17	e-mail	acesso	mudanca_de_senha	151	macae	29	0.19	0.01
18	e-mail	acesso	mudanca_de_senha	151	campos_centro	26	0.17	0.01
19	e-mail	suporte	informacoes	95	bom_jesus	16	0.17	0.01
20	e-mail	acesso	mudanca_de_senha	151	guarus	22	0.15	0.01
21	e-mail	acesso	mudanca_de_senha	151	cabo_frio	19	0.13	0.01
22	e-mail	criacao	usuario	152	macae	19	0.13	0.01
23	e-mail	acesso	mudanca_de_senha	151	itaperuna	18	0.12	0.01

Para concepção das regras exibidas na tabela acima, a taxa de confiança foi configurada em 10% e o suporte mínimo em 1%. E a partir das regras reproduzidas, verifica-se que a reitoria foi a maior requisitante dentre todos os chamados de suporte ao portal institucional (95%), de suporte e solicitação de serviços de rede (88% e 71%) e de cadastro ao sistema luma (55%).

Para todas as solicitações de suporte ao sistema suap, novamente a reitoria (31%) e o campus Itaperuna (20%) foram seus maiores utilizadores.

Para a categoria e-mail, as solicitações que se referiam à criação de usuário foram mais numerosas vindas dos *campi* Guarus (30%), Campos-Centro (21%) e Macaé (13%). Para a mesma categoria, mas em relação à subcategoria suporte e objeto central “informacoes”, os *campi* Guarus (23%), Campos-Centro (22%), reitoria (20%) e Bom Jesus (17%) se apresentaram como os mais comuns. Por último, as solicitações de mudança de senha foram criadas com mais frequência pelos *campi* Macaé (19%), Campos-Centro (17%), Guarus (15%), Cabo Frio (13%) e Itaperuna (12%).

A seguir, a tabela 10 exibe o arranjo correspondente à obtenção das correlações dos atributos “categoria”, “subcategoria”, “objeto\_central” → “meio\_de\_abertura”:

Tabela 10 - Correlação dos atributos “categoria”, “subcategoria”, “objeto\_central” → “meio\_de\_abertura”.

	antecedente			consequente				
	categoria	subcategoria	objeto_central	meio_de_abertura	conf.	sup.		
1	sistemas	suporte	portal_institucional	117	Email	117	1	0.01
2	sistemas	acesso	luma	64	Email	61	0.95	0.01
3	sistemas	suporte	informa	23	Email	21	0.91	0.01
4	infraestrutura	solicitacao	acesso_rdweb	78	Email	68	0.87	0.01
5	sistemas	acesso	suap	37	Web	31	0.84	0.01
6	e-mail	criacao	usuario	152	Email	123	0.81	0.01
7	e-mail	suporte	quota	61	Email	49	0.8	0.01
8	infraestrutura	suporte	telefonica	45	Email	36	0.8	0.01

9	sistemas	suporte	suap	87	Email	65	0.75	0.01
10	e-mail	acesso	informacoes	39	Web	28	0.72	0.01
11	sistemas	suporte	ponto_biometrico	35	Email	25	0.71	0.01
12	infraestrutura	solicitacao	servicos_de_rede	31	Email	22	0.71	0.01
13	e-mail	suporte	informacoes	95	Email	66	0.69	0.01
14	infraestrutura	suporte	servicos_de_rede	75	Email	47	0.63	0.01
15	e-mail	acesso	mudanca_de_senha	151	Email	79	0.52	0.01
16	e-mail	acesso	mudanca_de_senha	151	Web	71	0.47	0.01
17	e-mail	suporte	informacoes	95	Web	27	0.28	0.01
18	infraestrutura	suporte	servicos_de_rede	75	Other	19	0.25	0.01
19	sistemas	suporte	suap	87	Web	21	0.24	0.01
20	e-mail	criacao	usuario	152	Web	29	0.19	0.01

As regras demonstradas na tabela acima foram obtidas com os parâmetros da confiança e suporte mínimo configurados em 10% e 1% respectivamente, e a partir das mesmas comprova-se que majoritariamente as seguintes classificações foram abertas via Email: para a categoria sistemas → portal institucional (100% de confiança); acesso ao sistema luma (95%), suporte ao sistema informa (91%), suporte ao sistema suap (75%) e ao sistema de ponto biométrico (71%); para a categoria infraestrutura → solicitação de acesso ao rdweb (87%), suporte a telefonia (80%), solicitação (71%) e suporte (63%) a serviços de rede; para a categoria e-mail → criação de usuário (81%), suporte a quota (80%), informações de suporte (69%) e mudança de senha (52%).

Para o meio de abertura via *web*, as classificações mais comuns para sistemas foram acesso (84%) e suporte (24%) ao suap e, para a categoria e-mail, informações de acesso (72%), mudança de senha (47%), informações de suporte (28%) e criação de usuário (19%).

Abaixo, a tabela 11 apresenta o arranjo que concerne a obtenção das correlações dos atributos “categoria”, “subcategoria”, “objeto\_central” → “tempo\_de\_resolucao”:

Tabela 11 - Correlação dos atributos “categoria”, “subcategoria”, “objeto\_central” → “tempo\_de\_resoucao”.

	antecedente			consequente				
	categoria	subcategoria	objeto_central	tempo_de_resolucao	conf.	sup.		
1	e-mail	acesso	mudanca_de_senha	151	1_dia	128	0.85	0.01
2	e-mail	suporte	redirecionamento	24	1_dia	19	0.79	0.01
3	e-mail	criacao	usuario	152	1_dia	117	0.77	0.01
4	sistemas	acesso	suap	37	1_dia	28	0.76	0.01
5	sistemas	acesso	luma	64	1_dia	48	0.75	0.01
6	sistemas	suporte	informa	23	1_dia	17	0.74	0.01
7	e-mail	suporte	quota	61	1_dia	45	0.74	0.01
8	e-mail	acesso	informacoes	39	1_dia	28	0.72	0.01
9	sistemas	suporte	suap	87	1_dia	56	0.64	0.01
10	sistemas	suporte	portal_institucional	117	1_dia	74	0.63	0.01
11	infraestrutura	solicitacao	acesso_rdweb	78	1_dia	48	0.62	0.01
12	sistemas	suporte	ponto_biometrico	35	1_dia	21	0.6	0.01
13	e-mail	suporte	informacoes	95	1_dia	55	0.58	0.01

14	infraestrutura	suporte	telefonica	45	1_dia	19	0.42	0.01
15	infraestrutura	suporte	servicos_de_rede	75	1_dia	21	0.28	0.01
16	infraestrutura	suporte	servicos_de_rede	75	2_e_3_dias	17	0.23	0.01
17	infraestrutura	suporte	servicos_de_rede	75	>=_32_dias	15	0.2	0.01
18	sistemas	suporte	portal_institucional	117	2_e_3_dias	16	0.14	0.01
19	e-mail	criacao	usuario	152	2_e_3_dias	18	0.12	0.01

As regras exibidas foram geradas com um índice de confiança maior ou igual a 10% e suporte mínimo de 1%, e indicam que as classificações abaixo foram predominantemente solucionadas em um dia:

- Categoria e-mail → mudança de senha (85%), redirecionamento (79%), criação de usuário (77%), suporte a quota (74%), informações de acesso (72%) e informações de suporte (58%);
- Categoria sistemas → acesso ao suap (76%), acesso ao luma (75%), suporte ao informa (74%), suporte ao suap (64%) e suporte ao portal institucional (63%);
- Categoria infraestrutura → solicitação de acesso ao rdweb (62%), suporte a telefonia (42%) e suporte a serviços de rede (28%).

Ainda, verifica-se que as classificações a seguir foram as mais consideráveis quanto aos atendimentos realizados entre 2 e 3 dias: suporte a serviços de rede (23%), suporte ao portal institucional (14%) e criação de usuários de e-mail (12%).

## 5. CONSIDERAÇÕES FINAIS

Indubitavelmente o processo de descoberta de conhecimento aplicado à base de dados do sistema de solicitação de serviços de TI e o conhecimento depreendido a partir do mesmo corroboram ideias até então de natureza meramente intuitivas, e de acordo com os resultados angariados é possível afirmar que a direção de Tecnologia da Informação da reitoria da instituição federal de ensino possui agora conhecimentos sólidos a respeito das características das solicitações de serviços que a mesma oferta e presta suporte. Ressalta-se que a evidenciação deste conhecimento é creditada não somente ao processo de KDD propriamente dito, mas também à criação e aplicação do modelo de categorias de tratamento de solicitações de serviços de TI que fora utilizado na classificação dos tíquetes e ao emprego dos demais atributos determinados.

Ainda, as regras de associação geradas demonstram que para um panorama onde os processos de TI não são gerenciados por intermédio de *frameworks* relacionados à GSTI, existem alguns parâmetros que apresentaram bons resultados, como o tempo médio de resolução das categorias mais numerosas da amostra em sua maioria sendo de apenas um dia seguido pelo período entre 2 e 3 dias. Contudo, também é factível a necessidade de se estabelecer um processo específico de acompanhamento de chamados que possuem um considerável tempo de atendimento, especialmente os associados às áreas de extrema importância como as de infraestrutura e sistemas.

Com isso, as ações e procedimentos de TI podem, a partir de agora, se alicerçar seguramente em dados e fatos exequíveis que podem conduzi-la a uma maior eficiência e eficácia àquilo que a compete e ao contínuo alinhamento aos objetivos estratégicos da instituição de ensino.

Quanto às limitações verificadas, salienta-se que o modelo de categorias e as classificações realizadas se basearam apenas no que fora atendido durante o período que a amostra utilizada se concentra, não contemplando, portanto, serviços ou sistemas que entram em funcionamento posteriormente à sua seleção. Outro ponto a se destacar é a insuficiência que a versão atual do sistema osTicket© possui em propiciar a obtenção de informações essenciais que pudessem subsidiar a uma futura implantação de processos da biblioteca ITIL® por exemplo, o que enfatiza a necessidade de se implementar outra ferramenta de mesmo cunho porém com mais recursos que permitam viabilizar a implantação destacada.

## **5.1. CONTRIBUIÇÕES**

Os subsídios que a presente pesquisa objetiva propiciar correspondem a:

- Concepção de um modelo de categorias de tratamentos de solicitações de serviços de TI baseado nos tíquetes criados através do sistema osTicket©;
- Evidenciação das características dos serviços de TI ofertados e utilizados pelos servidores da instituição de ensino com o intuito de proporcionar uma futura implantação de modelos voltados à gestão de serviços de TI;
- A validação do modelo de categorias proposto por meio de sua aplicação no estudo de caso em questão.

## **5.2. TRABALHOS FUTUROS**

Segundo a conjuntura apresentada nesta pesquisa, expõem-se algumas propostas ulteriores caso a ideia de análise de requisições de serviço prossiga:

- A utilização de um sistema de solicitação de serviços de TI que contemple em sua estrutura funcional algumas informações nativas pertinentes aos chamados, como a classificação em evento, incidente, problema, etc., a fim de minimizar o tempo necessário para a análise dos dados, juntamente com o emprego de ferramentas computacionais que permitam avaliar cada solicitação por intermédio de

aprendizados supervisionado ou não supervisionado;

- Definição de novos atributos que aumentem ainda mais a especificidade do modelo de categorias, como o setor que o solicitante encontra-se lotado;
- Emprego efetivo dos processos de boas práticas dos *frameworks* de gestão de serviços de TI a partir do modelo de categorias, especialmente, os processos da ITIL® de Gerenciamento de Catálogo de Serviço, Gerenciamento de Evento, Gerenciamento de Incidente, Gerenciamento de Requisição, Gerenciamento de Problema e de Mensuração de Serviço.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

AGRAWAL, R.; SRIKANT, R. Fast Algorithms for Mining Association Rules. In: PROCEEDINGS OF THE 20<sup>TH</sup> VLDB CONFERENCE, 1994, Santiago. **Proceedings...** Santiago: 1994. p. 487-499.

AL-HAMAMY, A. H. et al. **Pruning of Apriori-Algorithm's Pruning Steps**. [200-]. Disponível em: < <http://www.iasj.net/iasj?func=fulltext&aId=41181>>. Acesso em: 14 out. 2014.

AMO, S. **Curso de Data Mining**. 2012. Disponível em: <<http://www.deamo.prof.ufu.br/CursoDM2012.html>>. Acesso em: 21 ago. 2014.

ARRAJ, V. **ITIL®: The Basics**. 2013. Disponível em: <[http://www.firesafetyguides.co.uk/gempdf/ITIL\\_The\\_Basics.pdf](http://www.firesafetyguides.co.uk/gempdf/ITIL_The_Basics.pdf)>. Acesso em: 19 jul. 2014.

ASSMANN, C. L.; FROZZA, R.; KIPPER, L. M. Análise Comparativa de Aplicações de Mineração de Dados: uso em diferentes contextos para uma melhor tomada de decisão. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 32., 2012, Bento Gonçalves. **Anais...** Bento Gonçalves: Fundaparque, 2012.

BORTOLI, J. **Data Mining - (Mineração de Dados)**. 2012. Disponível em: <<http://www.joeldebortoli.com/2012/08/data-mining-mineracao-de-dados.html>>. Acesso em: 2 set. 2014.

BOUCKAERT, R. R. et al. **WEKA Manual for Version 3-6-7**. 2012. Disponível em: <[https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB0QFjAA&url=http%3A%2F%2Fftp.jaist.ac.jp%2Fpub%2Fsourceforge%2Fw%2Fproject%2Fwe%2Fweka%2Fdocumentation%2F3.6.x%2FWekaManual-3-6-7.pdf&ei=hD5qVlrlIoOgNvDsg5AG&usg=AFQjCNHkD31g9F1A4xKYJkcdw4cUAwuRyg&sig2=eJxhPPNq2h\\_ZvZQnHqV4pw&bvm=bv.79142246,d.eXY](https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB0QFjAA&url=http%3A%2F%2Fftp.jaist.ac.jp%2Fpub%2Fsourceforge%2Fw%2Fproject%2Fwe%2Fweka%2Fdocumentation%2F3.6.x%2FWekaManual-3-6-7.pdf&ei=hD5qVlrlIoOgNvDsg5AG&usg=AFQjCNHkD31g9F1A4xKYJkcdw4cUAwuRyg&sig2=eJxhPPNq2h_ZvZQnHqV4pw&bvm=bv.79142246,d.eXY)>. Acesso em: 19 ago. 2014.

CAMILO, C. O.; SILVA, J. C. Mineração de Dados: Conceitos, tarefas, métodos e ferramentas. **Universidade Federal de Goiás**, p. 1-29, 2009. Disponível em:

<[http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_001-09.pdf](http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf)>. Acesso em: 09 ago. 2014.

CARDOSO, O. N. P.; MACHADO, R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. **Revista de Administração Pública**, Rio de Janeiro, v. 42, n. 3, p. 495-528, jun. 2008. Disponível em: <<http://www.scielo.br/pdf/rap/v42n3/a04v42n3.pdf>>. Acesso em: 06 ago. 2014.

CARTLIDGE, A. et al. **An Introductory Overview of ITIL® V3**. UK: The UK Chapter of the itSMF. 2007a.

CARTLIDGE, A. et al. **An Introductory Overview of ITIL® 2011**. London: The Stationery Office. 2012b.

CARVALHO, D. R. **Árvore de decisão/ algoritmo genético para tratar o problema de pequenos disjuntos em classificação de dados**. 2005. 173 f. Tese (Doutorado em Ciências em Engenharia Civil) - Coordenação dos Programas de Pós-Graduação em Computação de Alto Desempenho / Sistemas Computacionais, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

CEPIK, M.; CANABARRO, D. R. **Governança de TI: transformando a Administração Pública no Brasil**. Porto Alegre: WS Editor, 2010. 206p.

CORRÊA, U. **Mineração de Dados de Help Desk usando Rattle - o caso Petrobras**. 2007. 107 f. Dissertação (Mestrado Profissionalizante em Administração) - Programa de Pós-Graduação e Pesquisa em Administração e Economia, Faculdade de Economia e Finanças IBMEC, Rio de Janeiro.

CÔRTEZ, S. C.; PORCARO, R. M.; LIFSCHITZ, S. **Mineração de dados - Funcionalidades, Técnicas e Abordagens**. Rio de Janeiro: PUC, 2002.

DAMASCENO, M. Introdução à Mineração de Dados usando o WEKA. In: CONGRESSO NORTE NORDESTE DE PESQUISA E INOVAÇÃO, 5., 2010, Maceió. **Anais...** Maceió, Instituto Federal de Alagoas, 2010.

DUNHAM, M. H. **DATA MINING – Introductory and Advanced Topics**. New Jersey: Pearson Education: 2006. 328p.

ELMASRI, R.; NAVATHE, S. B. **Fundamentals of Database Systems**. 6. ed. Boston: Addison-Wesley, 2010.

ESTEVES, R. A. S. **Implementação do Processo Gestão da Configuração da framework ITIL – um Estudo de Caso**. 2012. 228 f. Dissertação (Mestrado em Sistemas de Informação) - Escola Superior de Tecnologia e de Gestão, Instituto Politécnico de Bragança, Bragança.

FAYYAD, U. M. Data Mining and Knowledge Discovery in Databases: Applications in Astronomy and Planetary Science. In: PROCEEDINGS OF THE THIRTEENTH NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, 1996, Portland. **Proceedings...** Portland: AAAI Press, 1996. p. 1590-1592.

FAYYAD, U. M. et al. **Advances in knowledge discovery and data mining**. California: AAAI Press, 1996. 560p.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI magazine**, v. 17, n. 3, p. 37-54, 1996.

FERNANDES, A. A.; ABREU, V. F. **Implantando a Governança de TI: da Estratégia à Gestão dos Processos e Serviços**. 3. ed. Tijuca: BRASPORT, 2012. 634p.

FREITAS, M. A. S. **Fundamentos do gerenciamento de serviços de TI**. 2. ed. Tijuca: BRASPORT, 2013. 424p.

GALVÃO, N. D.; MARIN, H. F. Técnica de mineração de dados: uma revisão da literatura. **Acta Paulista de Enfermagem**, São Paulo, v. 22, p. 686-690, 2009. Disponível em: <<http://www.scielo.br/pdf/appe/v22n5/14.pdf>>. Acesso em: 12 ago. 2014.

GARCIA, C. A.; FROZZA, R. Sistema de Recomendação de Produtos Utilizando Mineração de Dados. **Revista Tecno-Lógica**, Santa Cruz do Sul, v. 17, n. 1, p. 78-90, jun. 2013. Disponível em: <<https://online.unisc.br/seer/index.php/tecnologica/article/view/3283>>. Acesso em: 12 ago. 2014.

GILLMEISTER, P. R. G.; CAZELLA, S.C. **Uma análise comparativa de algoritmos de regras de associação: minerando dados da indústria automotiva**. 2007. Disponível em: <<http://www.ais.com.br/pos/Uma%20an%20lise%20comparativa%20de%20algoritmos%20de%20regras%20de%20associa%20E7%20E3o%20-%20minerando%20dados%20da%20ind%20FAstria%20automotiva.pdf>>. Acesso em: 12 out. 2014.

GOLDSCHMIDT R; PASSOS E. **Data Mining: Um Guia Prático - Conceitos, Técnicas, Ferramentas, Orientações e Aplicações**. 1. ed. Rio de Janeiro: Editora Campus, 2005.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. 3. ed. Waltham: Morgan Kaufmann, 2011.

KOVÁCS, Z. L. **Redes Neurais Artificiais: Fundamentos e Aplicações**. 4. ed. São Paulo: Editora Livraria da Física, 2006.

LOPES, I. M. **Adopção de Políticas de Segurança de Sistemas de Informação na Administração Pública Local em Portugal**. 2012. 437 f. Tese (Doutorado em Engenharia e Gestão de Sistemas de Informação) - Escola de Engenharia, Universidade do Minho, Braga.

LUNARDI, G. L. **Um Estudo Empírico e Analítico do Impacto da Governança de TI no Desempenho Organizacional**. 2008. 201 f. Tese (Doutorado em Administração) - Escola de Administração, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul.

OGC (Office for Government Commerce). **ITIL® – Service Operation**. [S.I.]: The Stationery Office, 2007.

osTicket. **Support Ticket System**. 2014. Disponível em: <<http://osticket.com/>>. Acesso em: 11 abr. 2014.

PETERSON, R. R. Integration Strategies and Tactics for Information Technology Governance. In: **Strategies for Information Technology Governance**. Hershey: Idea Group Publishing, 2004. Cap. 2, p. 37-80.

phpMyAdmin. **Bringing MySQL to the web**. 2014. Disponível em: <[http://www.phpmyadmin.net/home\\_page/index.php](http://www.phpmyadmin.net/home_page/index.php)>. Acesso em: 12 abr. 2014.

REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. 1. ed. Barueri: Editora Manole Ltda., 2003.

ROMÃO, W. et al. Extração de regras de associação em C&T: O algoritmo Apriori. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 19., 1999, Rio de Janeiro. **Anais...** Rio de Janeiro: UFRJ/PUC, 1999.

SETHIBE, T.; CAMPBELL, J.; MCDONALD, C. IT Governance in Public and Private Sector Organisations: Examining the Differences and Defining Future Research Directions. In: ACIS, 18., 2007, Toowoomba. **Proceedings...** Toowoomba: 2007.

SFERRA, H. H.; CORRÊA, A. M. C. J. Conceitos e Aplicações de Data Mining. **Revista de Ciência & Tecnologia**, Piracicaba, v. 11, n. 22, p. 19-34, jul. 2003. Disponível em: <<http://www.unimep.br/phpg/editora/revistaspdf/rct22art02.pdf>>. Acesso em: 14 ago. 2014.

TARAPANOFF, K.; ARAÚJO, R. H. J.; CORMIER, P. M. J. Sociedade da informação e inteligência em unidades de informação. **Ciência da Informação**, Brasília, v. 29, n. 3, p. 91-100, dez. 2000. Disponível em: <<http://www.scielo.br/pdf/ci/v29n3/a09v29n3.pdf>>. Acesso em: 14 ago. 2014.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniq**