



## PROCESSOS DE AQUISIÇÃO MAIS INTELIGENTES PELA ANÁLISE E AVALIAÇÃO DOS PADRÕES DE CONSUMO

Marcelo Fabiano Nascimento de Oliveira

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientador: Alexandre de Assis Bento Lima

Rio de Janeiro  
Dezembro de 2017

PROCESSOS DE AQUISIÇÃO MAIS INTELIGENTES PELA ANÁLISE E  
AVALIAÇÃO DOS PADRÕES DE CONSUMO

Marcelo Fabiano Nascimento de Oliveira

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO  
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)  
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM  
CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:

---

Prof. Alexandre de Assis Bento Lima, D.Sc.

---

Prof. Geraldo Zimbrão da Silva, D. Sc.

---

Prof. Rodrigo Salvador Monteiro, D.Sc

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2017

Oliveira, Marcelo Fabiano Nascimento de

Processos de Aquisição mais inteligentes pela análise e avaliação dos Padrões de consumo/ Marcelo Fabiano Nascimento de Oliveira. – Rio de Janeiro: UFRJ/COPPE, 2017.

XIV, 87 p.: il.; 29,7 cm.

Orientador: Alexandre de Assis Bento Lima

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia de Sistemas e Computação, 2017.

Referências Bibliográficas: p. 81-87.

1. Descoberta de Conhecimento em Bases de Dados. 2. Mineração de Dados. 3. Mineração de Padrões Sequenciais. 4. Logística I. Lima, Alexandre de Assis Bento. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

Dedico este trabalho a meus pais amados, Maria da Graça Nascimento e Marcellino de Oliveira Filho, a minha querida esposa, Célia Regina Lima Sá Batista Oliveira, que muito me ajudou nas horas infindáveis de pesquisa em que precisei de tranquilidade para pensar e a meu querido filho, Arthur Lima de Oliveira.

## **Agradecimentos**

Agradeço a todos os cientistas e pesquisadores que desbravaram a área de estudo de Descoberta de Conhecimento, sem o qual este trabalho não poderia ser desenvolvido em sua totalidade. Agradeço a meu orientador, sempre presente, prestativo e inteligente nas orientações para problemas desafiadores durante o curso do projeto. Agradeço à Marinha do Brasil pela oportunidade ofertada de cursar a cadeira de Mestrado, sem a qual nada seria possível.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## PROCESSOS DE AQUISIÇÃO MAIS INTELIGENTES PELA ANÁLISE E AVALIAÇÃO DOS PADRÕES DE CONSUMO

Marcelo Fabiano Nascimento de Oliveira

Dezembro/2017

Orientador: Alexandre de Assis Bento Lima

Programa: Engenharia de Sistemas e Computação

Para dar início ou manter diversas de suas atividades, a Administração Pública precisa de inúmeros tipos de insumos, como alimentos e material de escritório, comprados do setor privado através de processos conhecidos como Compras Governamentais, as quais podem tomar diversas formas legais permitidos pela Lei 8666/93, também conhecida como lei de licitações. Além dos recursos financeiros não serem infinitos, forçando a Administração a estabelecer um critério de priorização, o tempo pode ser um inimigo poderoso, podendo, inclusive, inviabilizar a normalização de um estoque de uma emergência de um hospital. Assim, qualquer alternativa que pretenda amenizar a influência do tempo numa licitação, é bem vinda. Esta dissertação pretende apresentar um caminho diferente, propondo a busca de padrões que possam representar o comportamento de consumo de itens de fardamento, de uma organização em particular, conduzindo a elaboração da previsão de demanda baseada em padrões e, assim, possibilitar a redução do número de licitações necessárias.

Este trabalho envolve a área de Descoberta de Conhecimento (KDD, em inglês) para melhorar o gerenciamento de suprimentos logísticos, através da introdução da Mineração de Padrões Sequenciais, investigando os padrões na linha do tempo.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

SMARTER PROCUREMENT PROCESSES FOR THE ANALYSIS AND EVALUATION  
OF CONSUMER PATTERN

Marcelo Fabiano Nascimento de Oliveira

December/2017

Advisor: Alexandre de Assis Bento Lima

Department: Systems Engineering and Computing

In order to initiate or maintain several of its activities, the Public Administration needs numerous types of inputs, such as food and office supplies, purchased from the private sector through processes known as Government Procurement, which may take various legal forms permitted by Law 8666 / 93, also known as the bidding law. In addition to the fact that financial resources are not infinite, forcing Management to establish a prioritization criterion, time can be a powerful enemy and may even make it impossible to normalize a stock of a hospital emergency. Thus, any alternative that intends to soften the influence of time in a bid, is welcome. This dissertation intends to present a different path, proposing the search for patterns that can represent the consumption behavior of uniforms, of a particular organization, leading to the elaboration of the demand forecast based on patterns and, thus, to reduce the number of the necessary bids.

This work involves the Knowledge Discovery (KDD) area to improve the management of logistics supplies through the introduction of Sequential Standards Mining, investigating patterns in the timeline.

## Sumário

Lista de Figuras.....	xi
Lista de Tabelas .....	xiii
1- Introdução.....	1
1.1 Formulação da Situação Problema .....	2
1.2 Proposta da pesquisa .....	2
1.3 Objetivos de pesquisa .....	3
1.4 Questões de Pesquisa .....	3
1.5 Delimitações da Pesquisa.....	3
1.6 Solução Adotada.....	3
1.7 Resultados Obtidos .....	4
1.8 Organização do Trabalho .....	5
2 – Conceitos Básicos .....	6
2.1 Gerenciamento da Cadeia de Suprimentos.....	6
2.2 Logística.....	7
2.3 Processo de Compras no Setor Público: Licitação .....	8
2.4 Descoberta de Conhecimento .....	8
2.5 Métodos aplicados ao problema de Mineração .....	10
2.5.1 Método Supervisionado .....	10
2.5.2 Método Não Supervisionado .....	11
2.6 Mineração de Dados .....	11
2.6.1 Mineração de Padrões Sequenciais (MPS) .....	12
2.6.2 Abordagens para Mineração Sequencial de Dados .....	13
2.7 Ferramentas de Mineração de Padrões Sequenciais .....	14
2.7.1 <i>Generalized Sequential Pattern (GSP)</i> .....	14
2.7.2 <i>Sequential Pattern Discovery Using Equivalent Classes (SPADE)</i> .....	16
2.7.3 <i>Sequential Pattern Mining (SPAM)</i> .....	17
2.7.4 CMSPADE e CMSPAM.....	17
2.7.5 <i>Last Position Induction Sequential Pattern (LAPIN)</i> .....	18
2.7.6 PREFIXSPAN .....	18
2.8 Taxonomia .....	19
Taxonomia Canônica de Unidades Sistemáticas.....	21
Taxonomia de Domínio .....	21
Taxonomia de Processos ou Tarefas.....	21



2.9	Teoria da Classificação Facetada (TCF) .....	22
2.10	Biblioteca para apoio à pesquisa .....	24
3	– Metodologia .....	26
3.1	Visão Geral .....	26
	Passo 1: Tratamento e Formatação dos dados originais .....	26
	Passo 2: Mineração de padrões seqüenciais .....	28
	Passo 3: Interpretação dos resultados .....	28
	Passo 4: Formulação da Conclusão .....	29
3.1.1	Padrões Ordinários e ExtraOrdinários .....	29
3.2	Fluxo Principal da Solução .....	30
3.2.1	Geração de Entrada Padronizada .....	31
3.2.2	Escolha da abordagem: Ano a Ano ou período de 11 anos .....	32
3.2.3	Aplicação da Mineração de Padrões Sequenciais.....	33
3.2.4	Cálculo de Métricas e geração de categorias .....	33
3.2.5	Seleção de critérios e aplicação de Filtros .....	34
3.2.6	Análise de Cenários .....	35
3.2.7	Geração da Previsão de Demanda .....	36
4	– Processos de Aquisição mais inteligentes pela análise dos padrões de consumo .....	37
4.1	Tratamento do Conjunto de dados .....	37
4.2	Escolha da Abordagem: Ano a Ano ou Por 11 anos .....	38
4.3	Busca de Padrões Sequenciais .....	39
4.4	Tamanho Médio de Janela (TMJ) .....	40
4.5	Cálculo de Recorrência e Abrangência .....	41
4.6	Construção da Taxonomia: Aplicação da Teoria da Classificação Facetada.....	42
4.7	Seleção do algoritmo para a geração dos resultados.....	46
	4.7.1 Abordagem Ano a Ano.....	46
	4.7.2 Abordagem 11 anos .....	47
4.8	Análise de padrões de consumo.....	49
	4.8.1 Abordagem Ano a Ano.....	49
	4.8.2 Abordagem de 11 anos.....	60
4.9	Previsão de demanda gerada .....	63
	4.9.1 Abordagem Ano a Ano, cenário de curto Prazo e padrões ordinários .....	63
	4.9.2 Abordagem Ano a Ano, cenário de curto prazo e padrões extraordinários....	64
	4.9.3 Abordagem Ano a Ano, cenário de médio prazo e padrões extraordinários ..	64

4.9.4	Abordagem Ano a Ano, cenário de longo prazo e padrões extra-ordinários .	64
4.9.5	Abordagem de 11 anos, cenário de curto prazo e padrões ordinários .....	64
4.9.6	Abordagem de 11 anos, cenário de curto prazo e padrões extraordinários....	65
4.9.7	Abordagem de 11 anos, cenário de médio prazo e padrões extraordinários ..	65
4.9.8	Abordagem de 11 anos, cenário de longo prazo e padrões extraordinários ...	65
4.10	Mundo real: reflexo sobre as Licitações.....	66
4.10.1	Abordagem Ano a Ano.....	67
4.10.2	Abordagem de período de 11 anos .....	72
4.11	Resultados e Discussões.....	77
5	- Conclusões e Trabalhos Futuros .....	78
	Referências Bibliográficas .....	81

## Lista de Figuras

Figura 1 – Etapas de Descoberta de Conhecimento .....	9
Figura 2 – Processo de Descoberta de Conhecimento (KDD) .....	9
Figura 3 – Os 4 passos desenvolvidos na solução .....	27
Figura 4 – Representação da Planilha eletrônica recebida .....	27
Figura 5 – Fluxo principal da solução.....	30
Figura 7 – Representação da seleção do algoritmo.....	35
Figura 8 – Distribuição das sequências no período de 2006 a 2016.....	38
Figura 9 – Representação das Abordagens Ano a Ano e de 11 anos.....	39
Figura 10 – Representação da Escolha da ferramenta de Mineração .....	39
Figura 11 – Categoria Energia.....	44
Figura 12 - Taxonomia, Teoria da Classificação Facetada, Categoria Personalidade.....	45
Figura 13 – Recorrência dos padrões da tabela 22 .....	50
Figura 14 – Abrangência dos padrões da tabela 22 .....	50
Figura 15 – Recorrência dos padrões da tabela 23 .....	51
Figura 16 – Abrangência dos padrões da tabela 23 .....	51
Figura 17 – Recorrência dos padrões da tabela 24 .....	52
Figura 18 – Abrangência dos padrões da tabela 24 .....	52
Figura 19 – Recorrência dos padrões da tabela 25 .....	53
Figura 20 – Abrangência dos padrões da tabela 25 .....	53
Figura 21 – Recorrência dos padrões da tabela 26 .....	54
Figura 22 – Abrangência dos padrões da tabela 26 .....	54
Figura 23 – Recorrência dos padrões da tabela 27 .....	55
Figura 24 – Abrangência dos padrões da tabela 27 .....	55
Figura 25 – Recorrência dos padrões da tabela 28 .....	56
Figura 26 – Abrangência dos padrões da tabela 28 .....	56
Figura 27 – Recorrência dos padrões da tabela 29 .....	57
Figura 28 – Abrangência dos padrões da tabela 29 .....	57
Figura 29 – Recorrência dos padrões da tabela 30 .....	58
Figura 30 – Abrangência dos padrões da tabela 30 .....	58
Figura 31 – Recorrência dos padrões da tabela 32 .....	59
Figura 32 – Abrangência dos padrões da tabela 32 .....	59
Figura 33 – Distribuição dos itens do cenário de curto prazo .....	66
Figura 34 – Distribuição dos itens para o cenário de curto prazo .....	66
Figura 35 – Distribuição dos itens para o cenário de curto prazo, padrões ordinaries .....	66
Figura 36 – Distribuição dos itens para o cenário de curto prazo, padrões extraordinários .....	66
Figura 37 – Representação do processo de redução do número de licitações.....	67
Figura 38 – Curto Prazo ano de 2014, padrões extraordinários .....	68
Figura 39 – Curto Prazo ano de 2015, padrões extraordinários .....	68
Figura 40 – Curto Prazo ano de 2016, padrões extraordinários .....	68
Figura 41 – Curto Prazo ano de 2014, padrões ordinários .....	68
Figura 42 – Curto Prazo ano de 2015, padrões ordinários .....	68
Figura 43 – Curto Prazo ano de 2016, padrões ordinários .....	69
Figura 44 – Médio Prazo ano de 2014, padrões extraordinários .....	69

Figura 45 – Médio Prazo ano de 2015, padrões extraordinários .....	69
Figura 46 – Médio Prazo ano de 2016, padrões extraordinários .....	69
Figura 47 – Médio Prazo ano de 2014, padrões ordinários .....	70
Figura 48 – Médio Prazo ano de 2015, padrões ordinários .....	70
Figura 49 – Médio Prazo ano de 2016, padrões ordinários .....	70
Figura 50 – Longo Prazo ano de 2014, padrões extraordinários .....	70
Figura 51 – Longo Prazo ano de 2015, padrões extraordinários .....	70
Figura 52 – Longo Prazo ano de 2016, padrões extraordinários .....	71
Figura 53 – Longo Prazo ano de 2014, padrões ordinários .....	71
Figura 54 – Longo Prazo ano de 2015, padrões ordinários .....	71
Figura 55 – Longo Prazo ano de 2016, padrões ordinários .....	71
Figura 56 – Curto Prazo ano de 2014, padrões extraordinários .....	72
Figura 57 – Curto Prazo ano de 2015, padrões extraordinários .....	72
Figura 58 – Curto Prazo ano de 2016, padrões extra-ordinários .....	72
Figura 59 – Curto Prazo ano de 2014, padrões ordinários .....	73
Figura 60 – Curto Prazo ano de 2015, padrões ordinários .....	73
Figura 61 – Curto Prazo ano de 2016, padrões ordinários .....	73
Figura 62 – Médio Prazo ano de 2014, padrões extraordinários .....	73
Figura 63 – Médio Prazo ano de 2015, padrões extraordinários .....	73
Figura 64 – Médio Prazo ano de 2016, padrões extraordinários .....	74
Figura 65 – Médio Prazo ano de 2014, padrões ordinários .....	74
Figura 66 – Médio Prazo ano de 2015, padrões ordinários .....	74
Figura 67 – Médio Prazo ano de 2016, padrões ordinários .....	74
Figura 68 – Longo Prazo ano de 2014, padrões ordinários .....	75
Figura 69 – Longo Prazo ano de 2015, padrões ordinários .....	75
Figura 70 – Longo Prazo ano de 2016, padrões ordinários .....	75
Figura 71 – Longo Prazo ano de 2014, padrões extraordinários .....	76
Figura 72 – Longo Prazo ano de 2015, padrões extraordinários .....	76
Figura 73 – Longo Prazo ano de 2016, padrões extraordinários .....	76

## Lista de Tabelas

Tabela 1 – Geração e Poda de Sequências Candidatas .....	15
Tabela 2 – Exemplo construção Classificação Facetada.....	23
Tabela 3 – Organização dos dados da planilha.....	28
Tabela 4 – Organização das sequências .....	28
Tabela 5 – Representação da Planilha com itens.....	31
Tabela 6 – Representação da Montagem do Itemset .....	32
Tabela 7 – Representação das Sequências .....	32
Tabela 8 – Exemplo de Sequências .....	37
Tabela 9 – Número de sequências entre 2006 e 2016.....	38
Tabela 10 – Lista de Padrões, SID's e datas .....	40
Tabela 11 – Lista de SID e períodos.....	41
Tabela 12 – Categoria Energia .....	43
Tabela 13 – Categoria Personalidade/Entidade, com um especificador .....	43
Tabela 14 – Categoria Personalidade/Entidade, com dois especificadores .....	43
Tabela 15 – Categoria Personalidade/Entidade, com três especificadores .....	44
Tabela 16 – Abordagem Ano a Ano, cenário de curto prazo .....	46
Tabela 17 – Abordagem Ano a Ano, cenário de médio prazo .....	47
Tabela 18 – Abordagem Ano a Ano, cenário de longo prazo .....	47
Tabela 19 – Abordagem de 11 anos, cenário de curto prazo .....	48
Tabela 20 – Abordagem de 11 anos, cenário de médio prazo.....	48
Tabela 21 – Abordagem de 11 anos, cenário de longo prazo.....	48
Tabela 22 – Resultados para o padrão { Sapato Preto, Blusa Azul ou Branca } .....	49
Tabela 23 – Resultados para o padrão { Calção Azul P e Blusa Azul ou Branca } .....	50
Tabela 24 – Resultados para o padrão { Sapato e Calça } .....	51
Tabela 25 - Resultados para o padrão { Gandola e Sapato } .....	52
Tabela 26 – Resultados para o padrão { Sapato Preto, Calção Azul e Calça } .....	53
Tabela 27 – Resultados para o padrão { Calça Azul e Bota }.....	54
Tabela 28 – Resultados para o padrão { Sapato Preto e Bota } .....	55
Tabela 29 – Resultados para o padrão { Calça Azul e Dólmã }.....	56
Tabela 30 – Resultados para o padrão { Calça Azul e Bota }.....	58
Tabela 31 – Resultados para o padrão { Calça Azul, Sapato Preto e Blusa Azul } .....	58
Tabela 32 – Resultados para o padrão { Sapato, Calção, Bota e Gorro }.....	59
Tabela 33 – Resultados para o padrão { Calça e Camisa } .....	60
Tabela 34 – Resultados para o padrão { Camisa e Calção } .....	60
Tabela 35 – Resultados para o padrão { Camisa e Sapato } .....	61
Tabela 36 – Resultados para o padrão { Calção Azul e Calção Azul } .....	61
Tabela 37 – Resultados para o padrão { Sapato, Calção e Calça } .....	61
Tabela 38 – Resultados para o padrão { Calça Camuflada e Calça } .....	61
Tabela 39 – Resultados para o padrão { Calção e Bota } .....	62
Tabela 40 – Resultados para o padrão { Bota e Distintivo } .....	62
Tabela 41 – Resultados para o padrão { Bota e Gorro }.....	62
Tabela 42 – Resultados para o padrão { Calça e Gorro } .....	63
Tabela 43 – Previsão de demanda para padrões ordinários .....	63

Tabela 44 – Previsão de demanda para padrões extraordinários .....	64
Tabela 45 – Previsão de demanda para padrões extraordinários .....	64
Tabela 46 – Previsão de demanda para padrões extraordinários .....	64
Tabela 47 – Previsão de demanda para padrões ordinários .....	64
Tabela 48 – Previsão de demanda para padrões extraordinários .....	65
Tabela 49 – Previsão de demanda para padrões extraordinários .....	65

# 1- Introdução

Para cumprir suas funções prioritárias, muitas vezes, o Setor Público necessita adquirir inúmeros itens, através das Compras Governamentais, fornecidas pelo mercado privado. Por isso, há a necessidade das autoridades públicas definirem com clareza o que desejam adquirir, ou seja, se conseguem realizar suas previsões de demanda, sem excessos empregando recursos além do necessário ou em falta, ocasionando, desabastecimento.

FARIA *et al.* (2008), em estudo sobre o tempo de uma licitação, analisa e consegue valorar tempos médios para o transcurso das diversas modalidades de licitação, com dados obtidos do Departamento de Licitações da Prefeitura Municipal de Viçosa. Para ele, um Pregão consome, em média, 15 dias; Já uma concorrência, 60 dias. Enquanto o Sistema de Compras Federal On line, o ComprasNet<sup>1</sup>, menciona 17 dias para o Pregão e 120 dias para concorrência. Intervalos tão incertos e, muitas vezes, longos podem ocasionar consequências catastróficas, se o que estiver em jogo for, por exemplo, suprimentos para atividades essenciais como a hospitalar (medicamentos para o setor de emergência de um hospital) ou, ainda, para o fornecimento de alimentos para uma creche.

A quantidade recursos financeiros disponíveis para executar aquisições é outra preocupação, requerendo a utilização de mecanismos de priorização de aquisições. Para auxiliar nessa escolha, muitas organizações desenvolveram atividades internas que analisam as informações disponíveis e emitem uma sugestão onde o recurso poderia ser aplicado. Hanke *et al.* (1995) chamou a tentativa de prognosticar o futuro, através do exame do passado de previsão (também, com o significado de recomendação). Muitas organizações dispõem de sistemas automatizados para fazer suas previsões de demanda, internamente, que fazem uso de vários modelos desenvolvidos pela academia, sendo os mais comuns o qualitativo e quantitativo. Esses sistemas representam um ganho de precisão e velocidade na análise de dados massivos, fornecendo, ao gestor/administrador, ferramentas e informações mais palpáveis para uma tomada de decisão sobre comprar ou não, de forma mais consciente, eficiente e eficaz.

<sup>1</sup> Sítio governamental disponível em: <https://www.comprasgovernamentais.gov.br/>

Basicamente, os métodos de previsão/recomendação tradicionais analisam séries históricas de itens individuais (normalmente, por meios estatísticos), construindo sugestões de compra para construir um processo de compra normal.

Esta dissertação visa propor um caminho alternativo para adquirir itens, utilizando as técnicas de Mineração de Dados, permitindo a identificação de padrões relevantes ao domínio de conhecimento. A próxima seção apresenta a situação problema a ser tratado neste trabalho.

## **1.1 Formulação da Situação Problema**

A Marinha do Brasil deve prover, entre outros fatores, vestuário apropriado para seus membros, devendo acontecer no momento e na quantidade adequados para atendimento. Há pontos de venda funcionando como lojas comerciais, onde é possível escolher entre vários itens. Uma vez que esta possibilidade existe, a organização precisa envidar esforços para manter seus estoques sempre disponíveis, com vistas a permitir pronto atendimento e evitar uma quebra de estoque.

Atualmente, o processo de controle e indicação de reposição de estoque implantado analise séries históricas num esforço de prever o volume necessário de itens a serem comprados, para então realizar uma nova previsão. A análise é realizada item a item, sem utilizar nenhum mecanismo de identificação de relacionamento entre itens, fazendo com que a aquisição de determinados itens seja realizada de maneira isolada, podendo levar mais tempo para finalizar a compra, custando mais e tendo o risco de levar a uma quebra de estoque. A próxima seção apresenta uma proposta de solução.

## **1.2 Proposta da pesquisa**

A proposta do trabalho é identificar padrões, entre os itens de fardamento ofertados pela Marinha do Brasil, obtidos através da aplicação de um conjunto de algoritmos de Mineração de Dados. O tratamento dos resultados permitirá a identificação de padrões relevantes para a organização, além de permitir observar sua frequência de observação (janela de tempo) e sua distribuição espacial (ocorrência em Território Nacional). Ao final, espera-se a construção de uma previsão de demanda que utiliza como base os padrões identificados, na eterna tomada de decisão sobre o que deve ser ou não ser comprado. Os principais objetivos a serem alcançados com o desenvolvimento da proposta, podem ser observados na próxima seção.



### **1.3 Objetivos de pesquisa**

Os principais objetivos a serem alcançados, nesta dissertação, são:

1. Permitir um melhor agrupamento dos itens de uma licitação, guiado pelo relacionamento estabelecido pela Mineração de Dados. Os padrões encontrados pelo processo de Mineração devem permitir que itens deixem de ser comprados em separado, para serem comprados em conjunto, reduzindo custos de maneira geral;
2. Permitir a ampliação do conhecimento dos padrões ocultos na organização, através da construção de categorias, baseadas em taxonomia;
3. Permitir a seleção dos padrões prioritários, visando à montagem de previsões de demanda com itens de maior saída ou necessidade.

### **1.4 Questões de Pesquisa**

Durante o desenvolvimento da metodologia algumas questões pertinentes foram suscitadas, devendo ser respondidas no desenrolar da pesquisa:

1. O conjunto de dados obtido da Marinha do Brasil contém dados em quantidade suficiente para delinear o comportamento de consumo e, assim, representar a realidade para sugerir padrões a serem adquiridos?

### **1.5 Delimitações da Pesquisa**

O trabalho prevê a introdução da Mineração de Dados para construir previsões de demanda, tendo uma forte preocupação com a variável “tempo”, identificando padrões seqüenciais. Não há a utilização de outras teorias como a introdução da Mineração de Hierarquias (LU *et al.*, 1998) ou nem Mineração de Séries Temporais (ALENCAR, 2007).

### **1.6 Solução Adotada**

Buscou-se um conjunto de dados que pudesse apoiar a pesquisa de padrões, fato que foi conseguido através de sistema de distribuição de fardamentos (uniformes) da organização. Os dados foram recebidos em forma de planilhas eletrônicas, que foram carregadas, seus dados foram corrigidos, limpos de ruídos e formatados, gerando uma entrada padronizada (de acordo com a literatura sobre busca de padrões seqüenciais, sobretudo na demonstração do problema original) (AGRAWAL-SKIRANT, 1995).

Duas abordagens foram definidas para investigar o intervalo médio de tempo entre as aparições de um padrão: Ano a Ano e Por Período de 11 anos. Após a realização do processo de Mineração, emerge a necessidade de interpretação dos resultados, em múltiplos critérios importantes, dentre eles, a seleção de padrões pela distribuição espacial e quais são mais observados ao longo dos anos. Em seguida, a análise de Cenários pretende formular três realidades, baseada na informação gerada até então, favorecendo uma possível tomada de decisões: **Cenário de Padrões Curto Prazo, Cenário de Médio Prazo e Cenário de Longo Prazo**. Ao final, pretende-se indicar ou sugerir o conjunto de padrões mais favoráveis a aquisição. Os resultados são apresentados na próxima seção.

## 1.7 Resultados Obtidos

Os principais resultados obtidos foram o desenvolvimento de categorização dos padrões, tendo por base uma taxonomia apoiada na Teoria da Classificação Facetada. Foram identificadas as facetas Personalidade/Entidade e Energia. Este expediente permitiu a classificação dos padrões em extraordinários (quando supera as expectativas do usuário) e ordinários (quando atinge as expectativas do usuário).

A mineração de dados resultou em 74 padrões recuperados que, após a aplicação de critérios de seleção, resultaram na construção de previsão de demanda contendo de 26 padrões mais representativos. Numa simulação da montagem de processos de compra, baseada em padrões, houve redução do número de licitações nos três anos selecionados de 2014, 2015 e 2016. Assim, na abordagem ano a ano, para o ano de 2014 há uma redução (simulando uma compra baseada nos padrões recuperados) de 37.78%, envolvendo todos os cenários; para o ano de 2015, uma redução de 38.34%; e para o ano de 2016, de 36% no número de licitações realizadas no ano. Para a abordagem de 11 anos, para um período de observação maior, tem-se 37.12% de redução para o ano de 2014; 36.11% para o ano de 2015; e 51.11% para o ano de 2016.

Todos os resultados sugerem a formação bem sucedida de um modo alternativo de montagem de licitações, utilizando menores quantidades de recursos, como dinheiro e tempo, melhorando a qualidade do processo e sua efetividade.

## **1.8 Organização do Trabalho**

Este trabalho está organizado da seguinte forma: o capítulo 2 apresenta todo o referencial teórico (ou conceitos básicos) necessário para embasar o trabalho, apresentando as ferramentas e técnicas utilizadas. O capítulo 3 apresenta a metodologia desenvolvida e aplicada ao trabalho; O capítulo 4 apresenta todas as etapas da solução, os resultados e as discussões; O capítulo 5 apresenta as conclusões e as de proposições de melhoria futuras para o trabalho.

## 2 – Conceitos Básicos

Este capítulo introduz toda a teoria que serve de base para o desenvolvimento da pesquisa, apresentando os principais conceitos básicos utilizados nesta dissertação.

### 2.1 Gerenciamento da Cadeia de Suprimentos

O Gerenciamento da Cadeia de Suprimentos é uma área de conhecimento bastante antiga e consolidada no cenário empresarial que lida com as mais diversas atividades de transformação e manipulação de materiais. Igualmente, é grande o número de autores que discorrem sobre o assunto, imprimindo seus pontos de vista originais sobre ele. Uma dessas referências é BERTAGLIA (2000), para o qual o Gerenciamento de Cadeia de Suprimentos corresponde ao conjunto de processos requeridos para obter materiais, agregar-lhes valor, de acordo com a concepção dos clientes e consumidores, além de disponibilizar os produtos para o lugar (onde) e para a data (quando) que os clientes consumidores os desejam.

As empresas varejistas são o principal exemplo do desenvolvimento da Cadeia de Suprimentos gerida para produzir resultados. Um dos principais benefícios, dessa abordagem, é a redução dos custos operacionais. BERTAGLIA (2000) menciona, ainda, que a estrutura da cadeia de suprimentos requer o domínio de várias competências, dentre elas, do profundo conhecimento de processos e de suas variáveis, de padrões de mercado e suas demandas, de modelos de distribuição, de níveis de serviço, das distâncias, das modalidades de transporte, da distribuição geográfica e etc. Isto demonstra que a atividade detém uma grande complexidade, uma vez que envolve elementos sobre os quais não se tem controle, como as constantes mudanças no tráfego rodoviário, por exemplo, devido às condições do tempo, acidentes ou intenso fluxo de veículos, podendo adicionar tempo significativo.

A Cadeia de Suprimentos, cada vez mais, é vital para as organizações que fazem gerência de materiais, tanto dos setores públicos como privados, no trato diário com manipulação de mercadorias até seu destino final. Além da redução dos custos mencionada, vários outros objetivos podem figurar como necessidades da organização, por exemplo, para fazer frente a uma investida de um concorrente, aproveitar uma oportunidade ocasionada por uma variação de cenário ou atender a uma mudança de hábito (às vezes repentino) de seus consumidores. Como observado, o Gerenciamento de Cadeia de Suprimento desempenha uma função importante em organizações

com necessidade de fornecer produtos e/ou serviços a quem necessita, na hora e na quantidade certa, objetivos esses perfeitamente aderentes para organizações, tanto públicas ou privadas, em que o tempo é uma variável escassa e que necessita ser trabalhada da melhor forma possível. A situação, em lide, representa bem a Organização Pública Federal estudada neste trabalho quando analisada sob a perspectiva de fornecedora de itens de fardamento para seus membros, através de seus diversos postos de venda direta. A próxima seção menciona a principal componente da Cadeia de Suprimentos: a Logística.

## **2.2 Logística**

A origem da atividade Logística faz menção direta ao meio militar, em tempos remotos de guerra, cujo objetivo era providenciar o abastecimento de tropas com armas, roupas e alimentos necessários para uma campanha de guerra. Somente em tempos mais recentes, a Logística ganhou os contornos conhecidos, passando a designar gestão de recurso. Em tempos ainda mais recentes, a Logística aparece sob várias denominações diferentes: Logística Reversa, Logística Integrada ou Logística Empresarial. No entanto, na essência, sua função não se altera.

Segundo o COUNCIL OF LOGISTICS MANAGEMENT<sup>2</sup> a Logística pode ser entendida como uma parte dos processos de Gerenciamento da Cadeia de Suprimentos que planeja, desenvolve e controla o efetivo fluxo e estocagem de bens, serviços e informações correlatos desde o ponto de origem até o ponto de consumo, com o objetivo de atender as necessidades dos clientes. BALLOU (1995) apresenta a Logística como sendo formada por atividades relacionadas aos fluxos tanto da entrada de materiais, como da saída de produtos, mencionando que essa visão integrada, no longo prazo, tende a fornecer, de forma conjunta, mais eficiência e eficácia ao processo. Para BOWERSOX-CLOSS (2001), a atividade responsável pelo fluxo de materiais, sendo processados e remetidos aos consumidores, através dos canais de distribuição, é a Logística. Segundo os autores, um aspecto importante da Logística é o esforço para fornecer ou tornar disponível os itens necessários em locais de distribuição, onde possam ser negociados. A disponibilidade, porém, tem custos, sendo um deles ligado a regularidade de compras de insumos ou itens finais, recaindo em um processo formalizado de Compras Governamentais. A seção seguinte analisa o principal instrumento para a realização de Compras Governamentais: Licitação.

<sup>2</sup> COUNCIL OF LOGISTICS MANAGEMENT, Órgão internacional que normatiza Conhecimento sobre a área de Cadeia de Suprimentos, <<https://cscmp.org/>>, acessado em 24 de maio de 2017.

### **2.3 Processo de Compras no Setor Público: Licitação**

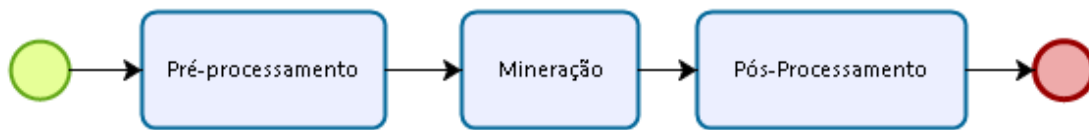
Para MEIRELLES (2001), Licitação é um procedimento administrativo de compras, com a seleção da proposta mais lucrativa, buscando propiciar igualdade de oportunidades aos que desejam realizar acordos com órgãos públicos, com padrões estabelecidos pela Administração, atuando com eficiência e moralidade. Dessa forma, a Licitação surge como uma ferramenta que atinge vários objetivos, como a isonomia no trato com terceiros privados, a formalização do processo de compras e a competição dos terceiros para a obtenção da proposta mais vantajosa.

A licitação contempla várias etapas, como elaboração de edital, Julgamento de habilitação e Julgamento da Proposta, e possui diversas modalidades, dentre elas, a Concorrência, o Convite, a Tomada de Preços, o Pregão Presencial e o Pregão Eletrônico. A seção seguinte apresenta a área de Descoberta de Conhecimento em Bases de Dados, que é mais ampla e contém a Mineração de Dados como uma de suas etapas.

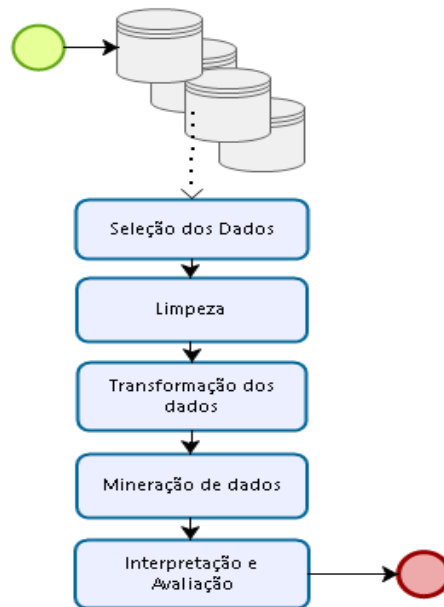
### **2.4 Descoberta de Conhecimento**

A área da Descoberta de Conhecimento em Bases de Dados (em inglês, *Knowledge Discovery in Databases - KDD*) nasceu prometendo fazer uso inteligente e eficiente dos dados acumulados pelas organizações por anos a fio, transformando-os em informação e conhecimento útil, seguindo a linha aberta por ZELENY (1987), responsável por cunhar a expressão conhecida por DIKWE, um acrônimo, em inglês, que corresponde aos conceitos de Dados, Informação, Conhecimento, Sabedoria e Iluminação. GEOBEL-GRUENWALD (1999) acrescenta que o *KDD* envolve um processo interativo e iterativo, incluindo a execução de inúmeros passos para basicamente tornar dados brutos em conhecimento útil.

Para gerar conhecimento, através de *KDD*, é preciso um processo metódico e bem definido, como o representado na Figura 1, que possui as etapas de Pré-processamento (incluindo os movimentos de preparação dos dados a serem fornecidos como entrada para os algoritmos de Mineração), Mineração de Dados (onde ocorre o processamento propriamente dito) e Pós-processamento (etapa de depuração e/ou síntese dos padrões descobertos) (CALIL, 2008).



**Figura 1 – Etapas de Descoberta de Conhecimento**



**Figura 2 – Processo de Descoberta de Conhecimento (KDD)**

FAYYAD *et al.* (1996b) fornece uma visão bastante interessante do processo de Descoberta de Conhecimento (representado na Figura 2) composto por diversas etapas, ampliando o entendimento demonstrado por CALIL (2008): **Seleção de dados**, prevendo a coleta e seleção de dados; **Limpeza**, prevendo a análise dos dados, retirando possíveis efeitos adversos como ruídos e valores ausentes; **Transformação**, diz respeito à agregação ou criação de novos dados, a partir dos existentes; **Mineração**, aplicação dos algoritmos de Mineração de Dados de interesse para a busca de padrões ou relações em determinado conjunto de dados; **Interpretação e Avaliação**, que verifica a qualidade do conhecimento gerado, validando a coerência e relevância frente ao problema original.

PIATETSKY-SHAPIRO (1991) é reconhecido como o primeiro a cunhar a expressão Descoberta de Conhecimento em Bases de Dados (*KDD*), em workshop no ano de 1989,

ênfatizando que o conhecimento é o produto final de uma descoberta dirigida por dados, sendo bastante popularizado nas áreas de Aprendizado de Máquina e Inteligência Artificial.

Segundo FAYYAD *et al.* (1996b), a Descoberta de Conhecimento foca na busca por aspectos de padrões compreensíveis, os quais podem ser interpretados como conhecimentos úteis, oriundos de grandes bases de dados. O processo utilizado para identificar **padrões** é não trivial, entendendo dados como um conjunto de **fatos**, **padrões** como uma **expressão** em uma linguagem qualquer, descrevendo subconjuntos de dados ou modelos aplicados ao subconjunto; por fim, o **processo** é entendido como uma **sucessão de etapas**; o termo não trivial remete à ideia da necessidade de realização de alguma busca de dados ou inferência de resultados. FAYYAD *et al.* (1996b) menciona ainda que a componente matemática (estatística) fornece os meios necessários (linguagem e instrumental) para quantificar a incerteza resultante da tentativa de inferir padrões gerais oriundos de uma amostra retirada de uma população. De acordo com a literatura, a Descoberta de Conhecimento pode atender aos objetivos que, ao que tudo indica, são complementares: o teste de uma hipótese aventada (**Verificação**) e a busca por padrões (**Descoberta**); A Descoberta de padrões, ainda, pode ser subdividida em um processo de **Predição** (quando o interesse é prever um comportamento futuro) e a **Descrição** (onde o interesse é buscar padrões e apresentá-los de forma adequada a seres humanos).

Em seguida serão apresentados os dois métodos que tornam efetiva a busca de conhecimento através da Mineração de Dados: o Método Supervisionado e o Método Não-Supervisionado.

## 2.5 Métodos aplicados ao problema de Mineração

De acordo com AHLEMEYER-STUBBE *et al.* (2014), a Mineração de Dados faz uso de uma variedade de métodos que analisam e avaliam os dados para descobrir conhecimento, na forma de padrões, os quais podem surgir de forma não conhecida, inesperada, interessante e relevante, além de relacionamentos entre os dados que podem ser usados para tornar previsões válidas e acuradas. Mencionam-se dois métodos para atingir os objetivos mencionados, anteriormente: o método Supervisionado e o método Não Supervisionado, a serem descritos na próxima seção.

### 2.5.1 Método Supervisionado

Nesta modalidade, a análise dos dados é usada para estimar uma dependência desconhecida de dados de entrada e saídas conhecidos, sendo que os dados de entradas podem estar sob a forma quantitativa (artigos de compra, por exemplo) e os dados de saída, conhecidos como alvos na



Mineração de dados, são os objetivos a alcançar. Ou seja, as amostras de entradas são processadas pelo sistema de aprendizado, comparando sua saída com a saída da amostra, ocorrendo ajustes no sistema de aprendizado caso haja alguma discrepância (AHLEMEYER-STUBBE *et al.*, 2014).

Geralmente, são utilizados para técnicas preditivas, uma vez que tentam prever qual é a classe de uma instância desconhecida, baseadas em exemplos utilizados em seu treinamento. O exemplo mais comum é visto na forma de classificadores.

### **2.5.2 Método Não Supervisionado**

Nesta modalidade não são necessários ajustes finos, pois, a descoberta de padrões é feita, diretamente, pelos algoritmos de Mineração de Dados, sem variável alvo. Há apenas a inclusão de valores de entrada, sem necessidades de validações. Dessa forma, o objetivo do método Não Supervisionado é descobrir estruturas Naturais nos dados de entrada (AHLEMEYER-STUBBE *et al.*, 2014).

Em outras palavras, este método utiliza instâncias sem a determinação do atributo classe. Ou seja, é um tipo de aprendizado utilizado, geralmente, para análise exploratória dos dados, utilizando técnicas de Agrupamento ou Regras de Associação. As técnicas de associação objetivam buscar regras que contêm associações entre atributos presentes no conjunto de dados. Já o agrupamento, objetiva relacionar instâncias com características em comuns.

Uma etapa importante do processo de Descoberta de Conhecimento em Bases de Dados é a Mineração de Dados, com o emprego de diversas técnicas, dentre elas, a busca de Regras de Associação e o Agrupamento, mencionados no parágrafo anterior. Na próxima seção serão apresentados os pontos teóricos mais importantes sobre a Mineração de Dados.

## **2.6 Mineração de Dados**

De acordo com a literatura, a Mineração de Dados (*Data Mining*) é um ramo da Computação, com início nos anos 80, período em que foram demonstradas as primeiras preocupações das organizações com a imensa massa de dados gerada e, continuamente, armazenada, demandando novos espaços em disco a cada dia, porém, sem utilidade aparente como diferencial competitivo para essas mesmas organizações (DEVEZA, 2011).

A literatura, historicamente, atribui à Mineração de Dados a convergência de três importantes áreas de conhecimento, Inteligência Artificial, Banco de Dados e a Estatística. DE AMO (2004) define Mineração de Dados como um processo de extração ou mineração de

conhecimento de grandes volumes de dados. Para ZAKI (2000), a tarefa de Mineração de Dados é descobrir um conjunto de atributos, compartilhados, ao longo do tempo, entre um grande número de objetos em um dado banco de dados.

Atualmente, o problema da geração e armazenamento de informação se tornou muito mais complexo, incisivo e necessário, na medida em que muitos sistemas foram construídos, e milhões de pessoas têm acesso, por exemplo, a redes sociais que geram informação (na forma de voz, de texto, de fotos e de vídeos), tornando a quantidade de dados ainda maior, graças aos avanços propalados e alcançados pela área de Tecnologia da Informação e pelo desenvolvimento de tecnologias de armazenamento de grande capacidade, de forma heterogênea e distribuída.

A Mineração de Dados pode ser entendida, essencialmente, como a busca por compreensão de um fenômeno, cuja expressão visível e palpável se constata na forma de padrões conseguidos com o emprego de diversas disciplinas complementares, na busca por resultados úteis e/ou vitais para a organização. A próxima seção apresenta um ramo da Mineração de Dados, que leva em consideração a dimensão tempo: Mineração de Padrões Sequenciais.

### 2.6.1 Mineração de Padrões Sequenciais (MPS)

Atribui-se a AGRAWAL-SRIKANT (1995), originalmente, a formulação do problema de Mineração de Padrões Sequenciais, considerada uma extensão da análise de Regras de Associação, porém, levando em conta não apenas as relações entre atributos, mas também, a ordem temporal da ocorrência dos eventos. Como ponto de partida, a técnica pressupõe a existência de um banco de dados  $D$ , contendo registros de transações de compras identificadas por algum atributo, possuindo uma lista de itens e uma data correspondendo ao momento da execução da ação.

Alguns conceitos são importantes, dentre eles, o de sequência, que é uma lista ordenada de um conjunto de itens, sendo representada por  $s = (s_1, s_2, s_3, s_4, \dots, s_n)$ ;  $s_i$  está contido em  $I$  (que é uma lista de itens de interesse) e  $i$  pertence a  $[1, \dots, m]$ , onde  $m$  é o tamanho da sequência medido pelo número de itemsets presentes e cuja representação é  $|s|$ . Um itemset é escrito na forma  $I = (i_1, i_2, i_3, \dots, i_n)$ , onde  $i_j$  é um item. Assim, uma sequência com tamanho  $l$  é chamada de uma  $l$ -sequência (HAN *et al.*, 2001). Uma sequência  $S_a = (a_1, a_2, a_3, \dots, a_n)$  está contida em uma outra sequência  $S_b = (b_1, b_2, b_3, \dots, b_n)$  se existem inteiros  $1 \leq i_1 < i_2 < i_3 < \dots < i_n \leq m$ , de tal forma que  $a_1 \subseteq b_{i_1}, a_2 \subseteq b_{i_2}, \dots, a_n \subseteq b_{i_n}$ . Dessa forma, se uma sequência  $S_a$  está contida em uma sequência  $S_b$ , então considera-se a primeira como uma subsequência de  $S_b$  ou, de outra forma,  $S_b$  é uma supersequência

de  $S_a$ . Como exemplo, a sequência  $\{(4,5),(7,8)\}$  está contida na sequência  $\{(3), (2,4,5,6,7), (5,7,8), (0,3)\}$ , uma vez que  $(4,5) \subseteq (2,4,5,6,7)$ ,  $(7,8) \subseteq (5,7,8)$ .

Para AGRAWAL-SRIKANT (1995) nenhuma sequência de dados possui mais do que uma transação com o mesmo tempo de duração, sendo estes tempos utilizados como critérios de identificação do mesmo. Da mesma forma, as quantidades de itens não são consideradas em uma transação. O suporte de uma sequência  $S$  é definido como a razão entre a quantidade de consumidores que tenham  $S$  e o total de consumidores.

ADAMO (2001), em estudo sobre tamanho do espaço de busca, afirma ser problemático o objetivo da Mineração de Padrões Sequenciais, pois, o espaço pode ser calculado pela expressão  $(2^A)^P$ , onde  $A$  é o conjunto de todos os atributos considerados no problema e  $P$  é o comprimento da maior sequência de eventos encontrada.

Os objetivos da realização de uma Mineração de Padrões Sequenciais são especificados de maneira completa por WANG et. al. (2005), na forma de padrões freqüentes, regulares, periódicos, aperiódicos, estatisticamente significativos e aproximados.

Em seguida, são explicitadas as duas abordagens mais conhecidas para um processo de Mineração Sequencial de Dados: Abordagem de Geração e Teste de Candidatos e Abordagem por Crescimento de Padrões.

### **2.6.2 Abordagens para Mineração Sequencial de Dados**

A literatura sobre Mineração de Padrões Sequenciais apresenta duas abordagens principais para a obtenção de resultados: “Geração e Testes de Candidatos” e “Crescimento de Padrões”. Na **Abordagem de Geração e Teste de Candidatos**, o processo tem início em um conjunto grande de sequências, conhecido como sequências candidatas, cujo suporte é definido através de inúmeras execuções sobre a semente. Ao final, sequências cujos critérios estabelecidos tenham sido atendidos serão reservadas e apresentadas como resultado (HAN *et al.*, 2001).

A principal ferramenta representativa da Geração e Teste de Candidatos é o algoritmo Apriori e seus algoritmos correlatos, como o GSP (*Generalized Sequential Patterns*), cuja característica mais importante é o padrão linear de um número de sequências de dados, além da existência de propriedades escaláveis, que respeitam a ordem da sequência, a quantidade de transações e os itens que constam nelas (HAN *et al.*, 2001). Outros exemplos, dessa abordagem, podem ser observados no algoritmo SPADE (ZAKI, 2001), que faz uso de propriedades

combinatórias para decompor o problema original em problemas menores (mais fáceis de serem resolvidos de maneira independente), armazenados na memória principal, através de técnicas eficientes de procura e operações de junções.

A **Abordagem de Crescimento de Padrões** tem o objetivo de reduzir o número de candidatos gerados, ao contrário do que acontece com a abordagem de Geração e Teste, onde há geração de grande quantidade de candidatos gerados. Uma vez escolhido um limite mínimo de suporte muito baixo, há a possibilidade de continuidade de custos não triviais como: a geração de um grande número de sequências candidatas em banco de dados; a execução de múltiplas varreduras nos bancos de dados na mineração; e, a geração de grande número de candidatos quando o objetivo é minerar padrões sequenciais longos, independente das técnicas utilizadas. Um de seus representantes é o FP-Growth (HAN *et al.*, 1999), que utiliza a estrutura de dados FP-tree, armazenando informação quantitativa sobre padrões frequentes de forma comprimida.

O Crescimento de Padrões Frequentes está apoiado em três pilares principais, sendo o primeiro, a mudança da fase de geração de candidatos que é substituída pela análise de contagem de frequência de conjuntos de dados relevantes, preservando agrupamentos essenciais de informação; segundo, o conjunto de dados e o conjunto de padrões são particionados e são examinados em cada passo; terceiro, uma otimização dos dados vindos do banco, reduzindo seu tamanho e o tornando factível de caber na memória principal. O principal exemplo dessa abordagem é a ferramenta PrefixPan.

A próxima seção apresenta as sete principais ferramentas para a Mineração de Padrões Sequenciais, descrevendo toda a teoria envolvida nelas.

## **2.7 Ferramentas de Mineração de Padrões Sequenciais**

Esta seção visa a apresentar, formalmente, as principais ferramentas utilizadas, nesta dissertação, destacando seu processo de busca de padrões sequenciais, começando com o introdutor da área, o GSP.

### **2.7.1 Generalized Sequential Pattern (GSP)**

Proposto por AGRAWAL-SRIKANT (1995), o algoritmo GSP busca minerar os chamados padrões sequenciais, baseando-se na abordagem de Geração e Teste de Candidatos e sendo dividida em duas etapas: geração de candidatos e cálculo do suporte. Por sua vez, a etapa de geração de candidatos divide-se em junção e poda. O conceito de K-sequência é o número de vezes, em

diferentes elementos da sequência, em que o padrão é observado.  $L_k$  representa o conjunto de padrões sequências frequentes.

O GSP funciona, basicamente, por iterações, onde a primeira varredura do banco de dados busca itens frequentes que, nesta etapa, são chamados de 1-sequências; numa segunda rodada, são formadas as 2-sequências, as quais podem conter um itemset composto de dois itens frequentes ou composto de dois itemsets, onde cada item, sofre a supervisão do suporte mínimo para selecionar sequências. O processo continua para as k-sequências ditas candidatas, sempre a partir das (k-1)-sequências anteriores, até não restarem mais sequências frequentes ou sequências candidatas geradas, completando o processo de geração.

A Fase de Geração de Candidatos é composta de duas etapas: Junção e Poda. A etapa de Junção de um par de (k-1)-sequências é realizada pela união de uma sequência  $S_1$  com uma sequência  $S_2$ , retirando-se o primeiro item de  $S_1$  e o último item de  $S_2$ , obtendo sequências idênticas. O resultado é a sequência candidata gerada com o item retirado de  $S_1$  adicionado ao último elemento de  $S_2$ , formando um novo candidato. Por exemplo, dadas duas 3-sequências  $\langle(2,4),(5)\rangle$ ,  $\langle(4),(5),(7)\rangle$ ,  $\langle(3,9),(10)\rangle$ , a junção da primeira com a segunda sequência fornece duas 2-sequências iguais, pela eliminação de 2, na primeira, e de 7 na segunda, resultando na 4-sequência candidata  $\langle(2,4),(5),(7)\rangle$ .

A Fase da Poda elimina sequências tomando por base o suporte mínimo e a contagem numa K-sequência. A Tabela 1 apresenta outro exemplo das operações de Junção e Poda.

**Tabela 1 – Geração e Poda de Sequências Candidatas**

3-Sequências Frequentes	Candidatos 4-sequências	
	Depois da Junção	Depois da Poda
$\langle(1,2) (3)\rangle$	$\langle(1,2) (3,4)\rangle$	$\langle(1,2) (3,4)\rangle$
$\langle(1,2) (4)\rangle$	$\langle(1,2) (3) (5)\rangle$	
$\langle(1) (3,4)\rangle$		
$\langle(1,3) (5)\rangle$		
$\langle(2) (3,4)\rangle$		
$\langle(2) (3) (5)\rangle$		

É possível observar que a sequência  $\langle(1,2) (3)\rangle$  e a  $\langle(2) (3,4)\rangle$  formaram a sequência  $\langle(1,2) (3,4)\rangle$ , assim como a junção de  $\langle(1,2) (3)\rangle$  com  $\langle(2) (3) (5)\rangle$ , resultaram na formação de

$\langle(1,2) (3) (5)\rangle$ . Esta operação foi possível, uma vez que, o primeiro elemento deveria ser casado com outro que assumisse a forma  $\langle(2) (3,x)\rangle$  permitindo a geração da nova sequência. Caso diferente ocorreu com as outras sequências, onde esta paridade não pode ser encontrada. Ao final das etapas de geração e poda, o candidato 4-sequências  $\langle(1,2) (3,4)\rangle$  é encontrado, resolvendo o problema de busca de padrões.

A etapa do cálculo de suporte é executada em cada k-sequência candidata, para o qual é verificado se seu valor está abaixo do valor estipulado como mínimo, sendo eliminada em seguida e armazenadas aquelas com suporte maior.

Há uma otimização com o uso de estruturas de árvore e tabelas *hash*. A primeira, armazena um conjunto de sequências candidatas, além de permitir uma poda mais eficiente de elementos a cada iteração (DEVEZA, 2011), e a segunda, armazena a lista de sequências candidatas que é apontada por um nó da árvore.

### **2.7.2 Sequential Pattern Discovery Using Equivalent Classes (SPADE)**

O algoritmo GSP introduziu diversas inovações em idéias e possibilidades para a pesquisa de padrões sequenciais, além de alguns desafios, principalmente, de otimização e desempenho. Uma ferramenta desenvolvida na linha de melhoria é o SPADE, criado por ZAKI (2001), que deve utilizar-se de propriedades combinatórias para decompor o problema original em subproblemas menores, que podem se resolvidos, independentemente na memória principal, usando técnicas eficientes de rotulagem e operações de junção.

A estrutura (*Lattice*) (DAVEY, 2002) permite a decomposição do espaço de busca original em partes menores e o tratamento independente na memória principal; Aliada à proposta de decomposição do problema original, o algoritmo introduz outro conceito chave que é a adoção do formato de banco de dados vertical, enquanto outros adotam o formato horizontal, como o GSP, proporcionando melhor desempenho da fase de Geração e Testes de Candidatos e da fase de Poda (AGRAWAL-SRIKANT, 1995), melhorando o tratamento de grandes espaços de busca, da ordem de milhões de registros em bancos de dados.

Por último, propõe duas diferentes estratégias para enumerar sequências frequentes dentro de cada subestrutura (*sub-lattice*): buscas em largura e em profundidade, introduzindo esquemas de busca eficientes, além de reduzir custos computacionais importantes (como os custos de entrada e saída *I/O*).

### 2.7.3 Sequential Pattern Mining (SPAM)

O SPAM utiliza o formato vertical de representação de bases de dados, assim, como o SPADE, sendo desenvolvido por AYRES (2002). Após o processo de varredura na base de dados original, há uma transformação para o formato vertical (levando em consideração o suporte mínimo definido pelo usuário (*minsup threshold*)), além da observação da ordem lexicográfica inerente, no qual um item  $i$  que ocorre antes de um item  $j$ , sendo representados por  $i \leq j$ . Da mesma forma, a ordenação pode ser ampliada para as sequências sempre que uma sequência  $S_a$  for uma subsequência de  $S_b$ , ou seja,  $S_a \leq S_b$ .

Além disso, se  $S_a$  não é uma subsequência de  $S_b$ , não há relação de ordenação entre as sequências. Há dois conceitos importantes para o SPAM: a sequência estendida (conhecida como *s-Step*) e o Itemset estendido (ou *i-Step*); O primeiro representa uma sequência gerada pela adição de uma nova transação consistindo de um item simples ao final de uma determinada sequência de interesse; o segundo gera uma sequência pela adição de um item ao último *itemset* na sequência, desde que este item adicionado seja maior do que qualquer outro presente neste *itemset*. Por exemplo, dada uma sequência  $S_a = ((\{A,F,G\}, \{A,G\}))$ , um *s-Step* possível seria  $(\{A,F,G\}, \{A,G\}, \{H\})$  e um *i-Step* seria  $(\{A,F,G\}, \{A,G,I\})$ .

Uma vez que o algoritmo emprega duas formas complementares para a etapa de geração e teste de candidatos (*s-Step e i-Step*), há um correspondente número de processos de eliminação de candidatos não viáveis, respectivamente, poda de *s-Step* e de *i-Step*, que utilizam a propriedade anti-monótona ou princípio Apriori, que diz: se um conjunto de itens é frequente, então todos os seus subconjuntos, também, devem ser frequentes. A próxima seção investiga uma variação da ferramenta SPADE, proposta por FOURNIER-VIGER *et al.* (2014b).

### 2.7.4 CMSPADE e CMSPAM

De acordo com a proposta de FOURNIER-VIGER *et al.* (2014b), a ideia por trás dos algoritmos SPADE e SPAM continha elementos de melhor desempenho, se comparado ao algoritmo GSP, apesar de, também, pertencer à categoria Geração e Teste de Candidatos, gerando uma grande quantidade de sequências candidatas. Para o autor, os benefícios passam pelo uso de estruturas (*lattice*), permitindo uma decomposição do problema em frações menores e de gerenciamento mais fácil, além de consumir menor tempo de execução. No entanto, o problema da geração de grandes quantidades de candidatos não frequentes ainda persiste.

A introdução do conceito de co-ocorrência Map (CMAP) (FOURNIER-VIGER *et al.*, 2014b) visa, segundo o autor, a resolver parte desse problema com a introdução de uma segunda estrutura de formato vertical, separando a co-ocorrência sobre *itemset* (*i-extension*) e co-ocorrência sobre sequência (*s-extension*). Ainda, segundo o autor, o CMAP é uma estrutura de mapeamento de um item específico para um conjunto de outros itens que o sucedem.

### 2.7.5 *Last Position Induction Sequential Pattern* (LAPIN)

O algoritmo LAPIN, desenvolvido por YANG *et al.* (2005), é outra proposta da abordagem Geração e Teste de Candidatos, com um diferencial de utilização no conceito da última posição de um item  $s$ , que é considerado a chave para julgar se um padrão sequencial de tamanho  $k$  pode ser estendido para o tamanho  $k+1$  e unido a  $s$ , podendo reduzir, consideravelmente, o espaço de busca criado pelo algoritmo, tornando-o mais efetivo. O LAPIN se baseia na ideia de que a descoberta de regras de associação (AGRAWAL *et al.*, 1993) pode exibir padrões de comportamento sequencial, considerando o aspecto temporal presente nos dados  $s$ .

YANG *et al.* (2005) menciona que as regras de associação revelam o aspecto do relacionamento entre elementos de um *itemset* e entre sequências distintas, informando, assim, o aspecto temporal. Além disso, introduz o conceito de prefixo e sufixo associado a um *Itemset* (formando o chamado *Itemset Extension - IE*) ou associado a uma sequência (formando o conceito de *Sequence Extension - SE*).

Por exemplo, para o caso de uma sequência  $A$  composta dos *Itemsets*  $(a,b,c,\dots,n)$  e um item  $\{I\}$ , é possível a concatenação no formato *IE*, ou seja,  $(a,b,c,\dots,n \cup \{I\})$ , exibindo a união entre a sequência original e um item específico, ou, é possível a concatenação da forma *SE*, ou seja,  $(a,b,c,\dots,n, \{I\})$ , exibindo a inclusão do item  $\{I\}$  como último elemento da sequência.

### 2.7.6 PREFIXSPAN

O Prefixspan propõe uma abordagem diferente da Geração e Teste de Candidatos, introduzindo a abordagem de crescimento de padrões frequentes ou FP-Growth (**F**requent **P**attern **G**rowth) (HAN *et al.*, 2000a). O FP-Growth é um método eficiente e escalável para a Mineração de Padrões Frequentes/Sequenciais, sejam eles curtos ou longos, que utilizam a estrutura de FP-Tree (HAN *et al.*, 2000a), cuja base está no crescimento de fragmentos de padrões, que armazenam informação quantitativa sobre padrões frequentes de forma comprimida.



A estrutura em FP-Growth utiliza a metáfora de divisão e conquista, decompondo a tarefa de Mineração e o Banco de dados de acordo com os padrões frequentes/sequenciais obtidos, não gerando candidatos nem testes, nem varrendo a base de dados de forma repetitiva. O PrefixSpan está preocupado em buscar padrões relevantes, estatisticamente. Exemplos de aplicação importantes estão na busca de texto, DNA, Bioinformática, entre outros.

A estrutura em *FP-tree* mantém, entre outros fatores positivos, a completude da informação para um determinado padrão frequente minerado. Segundo HAN *et al.* (2001), o Prefixspan realiza a mineração do conjunto completo de padrões sequenciais, reduzindo os esforços de geração de candidatos subsequentes. A preocupação principal do Prefixspan é a geração excessiva de candidatos bastante comum em algoritmos de geração e testes de candidatos, como o GSP.

Para atingir esse objetivo, HAN *et al.* (2001) menciona, a criação do método freespan HAN *et al.* (2000a), cuja a ideia é usar itens frequentes para, recursivamente, projetar sequências e combiná-las para aumentar seu número, obtendo ao final uma grande base de dados projetada.

YANG *et al.* (2005) afirma que o Prefixspan adota uma estratégia para projetar sequências candidatas em diferentes grupos da chamada base de dados projetada. É importante dizer, que toda sequência presente nesses grupos são diferenciadas pelo prefixo comum dos grupos. Sua idéia principal é que, ao invés de projetar sequências de bancos de dados considerando-se todas as ocorrências possíveis de subsequências frequentes, a projeção é baseada apenas em prefixos frequentes. HAN *et al.* (2000a).

## 2.8 Taxonomia

A utilização de taxonomias, pela ciência, é antiga e remonta ao século dezoito (1735), quando um botânico sueco, de nome Karl Von Linné, criou um sistema para classificar seres vivos, dividindo-os em grupos, de acordo com suas características em comum e com o estabelecimento de uma ordem hierárquica, contidos em seu livro *Systema Naturae*<sup>3</sup>. Seu método ficou conhecido como "Taxonomia de Lineu".

Atualmente, as taxonomias vêm sendo utilizadas para organizar conhecimento em diversas vertentes tanto na internet, como nas corporações, por exemplo, em portais institucionais, em bibliotecas digitais, em web semântica, em ontologias e/ou gestão de informação e de

<sup>3</sup> Publicado em 1735, o autor faz a delimitação das suas ideias para uma classificação hierárquica das espécies. Lineu concebeu o seu "Systema" dividindo a Natureza em três reinos: Animalia, Vegetalia e Mineralia.

conhecimento, sempre com o viés de classificação e organização de conceitos.

Essencialmente, uma taxonomia contém a forte ideia de classificação sistemática e, esta acepção, está contida em várias definições de inúmeros autores, como GRAEF (2001) que apresenta a taxonomia como um sistema para nomear e organizar coisas dentro de grupos que compartilham características similares. Já HANLEY (2005) menciona que taxonomia é uma forma de arranjar tópicos e sub-tópicos relevantes em uma hierarquia ou estrutura de rede. BRUNO-RICHMOND (2003) afirma que taxonomia é a ciência da classificação de acordo com sistemas pré-determinados, resultando em um catálogo usado para fornecer um arcabouço conceitual para discussão, análise ou recuperação de informação. Cada autor contribui com seu ponto de vista, vislumbrando novas aplicações e conhecimentos que se somam. Assim, NOVO (2007) menciona o fato de que a taxonomia é fruto de um trabalho metuculoso, sendo realizada através de vários estudos e investigações, de forma não linear, pois um mesmo modelo não seria capaz de explicar todas as nuances de questões tão complexas e particulares.

Muito da relevância do emprego da taxonomia vem do interesse e disposição natural do ser humano para buscar padrões e criar classificação para os elementos que o cercam. Isso transparece, por exemplo, na tentativa de dar forma conhecida a nuvens, na arrumação de uma prateleira ou na organização de um armário de roupas. A busca de padrões e criação de classificações são lembrados por LANGRIDGE (1997) quando afirma serem estas atividades um fato inerente a todo o ser humano. CAMPOS-GOMES (2008) afirma que a taxonomia é uma classificação sistemática, com classes apresentando uma ordem lógica ligada a princípios, porém, tendo mais restrições em possibilidades de exploração, devido às relações hierárquicas. A taxonomia é de construção mais rápida que tesouros<sup>4</sup>, porém, não menos complexa, apesar de ser mais amigável que o tesouro, possibilitando uma visão geral de todas as ideias. Para TERRA (2005), a taxonomia é um sistema para classificar e facilitar o acesso à informação, com objetivos bem definidos de representar conceitos através de termos, agilizar a comunicação entre especialistas e entre estes e o público, buscar consenso, propor formas de controle da diversidade de significação e fornecer uma visão geral dos processos de conhecimento.

Do ponto de vista empresarial, AGANETTE *et al.* (2010) menciona que as taxonomias estão presentes de forma relevante, apresentando-se como instrumento de organização, recuperação de conhecimento e aprendizagem. Seu objetivo pode variar, haja vista o que se

<sup>4</sup> Tesouro – lista de palavras com significados semelhantes

pretende alcançar. Ainda com a preocupação voltada para o modelo empresarial, CONWAY *et al.* (2002) menciona a classificação de estruturas taxonômicas para aplicação em ambientes corporativos, dividindo-as em taxonomias descritivas, contendo vocabulários controlados advindos de Tesouros, juntamente, com todo o tipo de informação relevante como formas e ortografias; taxonomias para navegação envolvendo agrupamentos de informações, que busca encontrar conhecimento nos comportamentos de usuários e taxonomias para gerenciamento de dados, contendo termos controlados válidos e significativos.

Entre as características essenciais da construção de uma taxonomia estão a existência de uma lista de termos de domínio estruturada; a recuperação e organização de informação; agregação de dados e a evidenciação de modelo conceitual de domínio. É possível elencar as vantagens da utilização de taxonomias, dentre elas, a possibilidade de visualização do conteúdo de um sistema de forma integrada, a minimização do tempo gasto na busca, a padronização da ordem de arquivamento das informações e, um dos mais importantes para este trabalho, a padronização da terminologia utilizada. A literatura apresenta três principais tipos de taxonomias, que são descritas a seguir:

### **Taxonomia Canônica de Unidades Sistemáticas**

Classificação binária do conhecimento, também, chamado de dicotômica. Ela revela relações de família, gênero e espécie. Uma importante característica diz respeito ao fato de não possibilitar agregar novos conhecimentos depois da taxonomia construída.

### **Taxonomia de Domínio**

Sua característica principal é a policotomia, ou seja, a possibilidade de apresentação de mais de duas categorias. É voltada para a representação de um domínio de conhecimento. Seu recorte é determinado pelas características da organização a que serve.

### **Taxonomia de Processos ou Tarefas**

Também apresenta a característica policotômica, sendo composta de conceitos para representação de processos ou tarefas. O recorte é por tarefas/processos, em vários níveis e possui características da organização. A próxima seção apresenta a Teoria da Classificação Facetada como ferramenta para construção de taxonomias.

## 2.9 Teoria da Classificação Facetada (TCF)

A Teoria da Classificação Facetada (ou em Facetas, ou Analítico-Sintética, ou de Dois Pontos) foi introduzida Shiyali Ramamrita Ranganathan, professor de matemática indiano, que viveu entre 1892 e 1972, deixando uma vasta obra voltada para a área de classificação de conteúdo, considerado o precursor do método científico em biblioteconomia. Segundo DAHLBERG (1979), é o pai da Moderna Teoria da Classificação. No início do século 20, Ranganathan publicou sua *Collon Classification*<sup>5</sup> RANGANATHAN *et al.*(1989), no qual aplicou suas idéias de divisão dos assuntos em facetas ou características, agrupando classes por um mesmo princípio de divisão. Sua motivação esteve atrelada à discordância com a rigidez dos métodos disponíveis, até então, a Classificação Decimal de Dewey (CDD) e a Classificação Decimal Universal (CDU). Para Ranganathan, a ideia é produto do pensamento. O conhecimento é representado pela totalidade das ideias armazenadas pela memória humana. E a informação existe quando as ideias são comunicadas (NETO, 2014). RANGANATHAN (1967) conceitua assunto como "um corpo de ideias organizadas ou sistematizadas, cuja extensão e intenção devem ser coerentes com o domínio de interesse e, confortavelmente, ajustadas à competência intelectual e campo especializado de uma pessoa qualquer". A *Collon Classification* é apoiada por dois princípios importantes e complementares, a análise de facetas, na qual os assuntos são divididos em suas partes componentes, e síntese, que postula a combinação e a recombinação destas partes, objetivando a classificação de um documento. MARTÍNEZ *et al.* (2004) afirma que as facetas servem como princípio para organizar e reunir classes, sendo que estas são nomeadas. Por outro lado, as categorias fornecem uma ordem para a disposição dos tópicos numa taxonomia, pois se constituírem em princípio para organização do raciocínio. RANGANATHAN (1967) definiu uma ordem de citação, onde um assunto poderia ser enquadrado em uma de cinco categorias, a saber: Personalidade/Entidade (dos objetos, abstrações, fatos e do conjunto de coisas), Matéria (Material, propriedade), Energia (Processo, operação, fenômenos em si, eventos), Espaço (Dimensão física, local) e Tempo (Dimensão Temporal, período), conhecidos pelo acrônimo PMEST. BARBOSA (1972) menciona que não é necessário que as categorias tenham sempre estas denominações (ou seja, aquelas definidas por Ranganathan), mas sim que a elas correspondam. Também não é preciso que todas elas figurem, ao mesmo tempo, num assunto. Os termos relacionados como Matéria, em

<sup>5</sup> Obra de Shiyali Ramamrita Ranganathan, em que ficou célebre pro matematizar a classificação.

determinado assunto, poderão significar Produtos em outro. A Teoria de Classificação Facetada utiliza unidades classificatórias formadas por conjunto de assuntos básicos e idéias isoladas, propostos por RANGANATHAN (1967), no qual cada assunto básico consistiria em um assunto sem nenhuma idéia isolada como componente, ou seja, seria uma área mais abrangente do conhecimento, como Engenharia. A idéia isolada, por sua vez, consistiria em alguma idéia ou complexo de idéias ajustadas para formar um componente de um assunto, mas, em si mesma, ela não é considerada um assunto, por exemplo, Apartamentos. Dessa forma, Apartamentos seria a idéia isolada, combinada com o assunto básico Engenharia, é possível o assunto composto Construção de Apartamentos. CAMPOS (2001) inclui a noção de especificador, melhorando o formatação da idéia. Para o exemplo da Engenharia, SILVA (2011) apresenta um quadro que organiza e facilita a construção dos assuntos, representado na Tabela 2.

**Tabela 2 – Exemplo construção Classificação Facetada**

<b>Assunto Básico</b>	<b>Ideia Isolada</b>	<b>Assunto (composto)</b>	<b>Especificadores</b>	<b>Assuntos</b>
Engenharia	Apartamentos	Construção de Apartamentos	Luxo	Construção de Apartamentos de Luxo
			Popular	Construção de Apartamentos Populares

Ao final do processo, é possível a observância de duas categorias bem definidas: primeira, Construção de Apartamentos de Luxo e, a segunda, Construção de Apartamentos Populares. CAMPOS *et al.* (2008) oferece um processo de cinco passos para a construção de Taxonomia que, devido a sua importância, são reproduzidos neste trabalho.

**1 - Captura do Conhecimento.** Segundo NETO (2014) Esta etapa realiza o levantamento dos assuntos que deverão ser acomodados, numa estrutura classificatória, servindo como ponto de acesso à informação. Dessa forma, percebe-se que a captura de conhecimento visa buscar e entender o conjunto de informações que deve ser processado para a criação da taxonomia. A fonte pode ser de diversas formas, como em entrevistas com pessoal especializado ou através de documentos disponíveis para consulta. Para este trabalho, um farto material estava disponível para consultas, o Regulamento de Uniformes da Marinha do Brasil (RUMB), que se apresenta na forma de um catálogo contendo todo o material necessário para as atividades da instituição, em matéria de

uniformes.

**2- Análise dos documentos/informações.** Uma vez de posse das informações/documentos, é necessária a realização de varredura para a busca dos chamados descritores conceituais (NETO, 2014).

**3 - Elaboração da estrutura classificatória da taxonomia.** Depois de obter o conjunto de dados (captura do conhecimento) e de uma análise do material em busca dos termos necessários para a montagem da taxonomia, procede-se à elaboração/construção das categorias.

**4 - Escolha do software para a construção da taxonomia.** Para uma definição mais concisa e objetiva da ferramenta para construção da taxonomia, existem algumas opções gratuitas disponíveis no mercado.

**5 - Validação pelo Especialista.** Um dos objetivos de toda a validação é verificar se o que foi construído está aderente à necessidade em lide, além da sua correção e abrangência.

## **2.10 Biblioteca para apoio à pesquisa**

A biblioteca SPMF é um artefato de software de código-livre, escrita em linguagem Java, para Mineração de Dados, especializada em Mineração de Dados, sendo especializada em Mineração de Padrões e, distribuída sob a licença GPL v3. Ela oferece um conjunto de 133 algoritmos de Mineração (até o momento) para várias tarefas diferentes como Mineração de Regras de Associação, Mineração de Padrões Sequenciais, Classificação e Clusterização (Agrupamento).

Atualmente, se encontra na versão 2.30, lançada em 1 de março de 2018. Esta ferramenta é um esforço conjunto e já foi citada, aproximadamente, em 500 trabalhos acadêmicos ao longo dos anos. A biblioteca é fornecida na forma de um arquivo \*.jar (arquivo compactado) ou na forma de classes Java, as quais permitem manipulações no código fonte.

Existem outras plataformas que disponibilizam ambientes de Mineração de Dados, como o Weka (HALL *et al.*, 2009), o Knime<sup>6</sup> e a suíte Pentaho (BOUMAN-VAN DONGEN, 2009); O Weka testado está na versão 3.8.1 e é uma plataforma desenvolvida na Universidade de Waikato, em Hamilton, na Nova Zelândia. Por motivos de desempenho, alguns algoritmos foram retirados da plataforma, como o GSP, por exemplo.

A plataforma Knime oferece um conjunto de ferramentas para classificação, agrupamento e Mineração de Texto, mas não para buscar os padrões sequenciais utilizados neste trabalho.

<sup>6</sup> Knime - <https://www.knime.com/>

A suíte Pentaho utiliza a plataforma Weka, cujas limitações já foram apresentadas. Por essa razão, não foi escolhida como a ferramenta principal para o estudo desenvolvido neste trabalho e, em seu lugar, foi utilizado o Framework SPMF. O próximo capítulo descreve os passos realizados para desenvolver a solução utilizada neste trabalho.

## 3 – Metodologia

Este capítulo trata da metodologia empregada, nesta dissertação, comentando os passos a serem executados e os resultados a serem alcançados, ao longo do trabalho.

### 3.1 Visão Geral

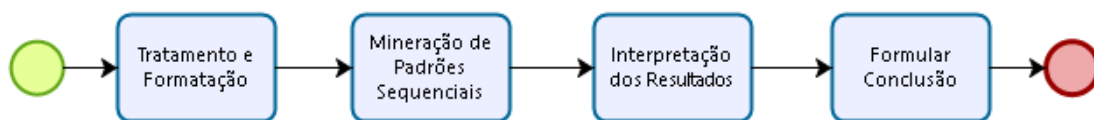
O método científico é algo, fortemente, arraigado no meio acadêmico, permitindo o encadeamento racional dos recursos disponíveis em prol de um objetivo legítimo e, previamente, identificável. (TRUJILLO, 1974), (MARCONI, 2007) e (PRODANOV, 2013). É forte e patente a preocupação com o resultado, e todo o caminho da pesquisa tem o rigor e a formalização, suficientemente, compatíveis com o resultado que se espera alcançar. Diante do exposto e da necessidade de sistematizar, organizar e apresentar um caminho viável com resultados relevantes e palpáveis, este trabalho pretende seguir o esquema proposto por PRODANOV (2013) para a definição da metodologia aplicada ao estudo. A pesquisa deve partir dos aspectos particulares e concretos e generalizá-los, verificando a hipótese inicial, direcionando a pesquisa no sentido do método indutivo (PRODANOV, 2013). Dessa forma, esta dissertação pretende seguir os passos observados na Figura 3, a qual resume o trabalho a ser elaborado, apresentando um fluxo simples de encadeamento das idéias.

#### **Passo 1: Tratamento e Formatação dos dados originais**

Para tornar o projeto viável, a busca de um conjunto de dados representativo foi realizada e conseguida, tendo em vista a plataforma de venda de itens de fardamento, da própria organização, responsável por fazer a venda e distribuição dos produtos, diretamente, a seus membros. Um total de 11 anos foi selecionado, iniciando em 2006 e finalizando em 2016, totalizando, aproximadamente, 4 milhões de registros, recuperados do sistema de vendas e distribuição, entregues na forma de planilhas eletrônicas, conforme representação da Figura 4. Seus dados remetem a idéia de tuplas individuais descrevendo itens de uma mesma compra que, para serem processados, necessitam de transformações para tratamento pelos algoritmos de mineração de padrões sequenciais. A primeira transformação propõe a construção dos conjuntos de itens ou *itemsets* (AGRAWAL-SKIRANT, 1995); ou seja, o agrupamento dos itens baseados no identificador NF ou da Nota Fiscal, que juntos, totalizaram 656 mil compras, realizadas pelos



membros da organização, no período de 11 anos. Agora, ao invés de itens individuais, tem-se grupos de itens, diferenciados pela Nota Fiscal, como por exemplo: Compra A, contendo Sapato e Boné; Compra B, contendo Gorro e Gandola e etc. As compras podem ter sido realizadas pela mesma pessoa, em momentos diferentes, formando, então, o que se convencionou chamar de sequência (AGRAWAL-SKIRANT, 1995); ou seja, de conjuntos de itemsets. Para o exemplo mencionado, uma sequência S poderia conter as compras A e B. Após a realização dos preparativos, o conjunto de itemsets foi reunido em sequências, perfazendo o total de 287 mil exemplares, os quais são, concretamente, as entradas para os algoritmos de mineração de padrões sequenciais, onde a ordem dos itemsets revela (para a ferramenta) uma dimensão temporal bem específica. O conjunto primário de dados foi recebido na forma de planilhas eletrônicas, necessitando sofrer todo o processo de tratamento descrito.



**Figura 3 – Os 4 passos desenvolvidos na solução**

NIP	NOME	IDENTIFICADOR_NF	DATA	NOME_COLOQUIAL	QTDE
1.000	João da Silva	126.584	04/27/2006 00:00:00	SAPATO PRETO 41	1
2.000	Maria Jose	476.930	12/15/2008 00:00:00	GORRO BRANCO 56 (100%POL)	1
2.000	Maria Jose	476.930	12/15/2008 00:00:00	CALCA CAMUFLADA 42	2
3.000	Angelo Santos	476.929	12/15/2008 00:00:00	CALCAO NATACAO G	1
2.000	Maria Jose	476.930	12/15/2008 00:00:00	GANDOLA CAMUFLADA 2	2
2.000	Maria Jose	476.930	12/15/2008 00:00:00	CAPA IMPERMEAVEL 54	1
2.000	Maria Jose	476.930	12/15/2008 00:00:00	GORRO CAMUFLADO	2
3.000	Marcos Alfonso	298.735	10/08/2007 00:00:00	GANDOLA CAMUFLADA 4	1

**Figura 4 – Representação da Planilha eletrônica recebida**

A Figura 4 apresenta uma representação da forma como os dados foram recebidos, incluindo a identificação do indivíduo (nip), o nome, a identificação da nota fiscal, a data, o item e a quantidade comprada. Para montar as sequências, deve-se primeiro organizar os itens por data e nota fiscal, conforme observado na Tabela 3, montando os itemsets. Assim:

**Tabela 3 – Organização dos dados da planilha**

<b>Data</b>	<b>Nota fiscal</b>	<b>Item</b>
27/4/2006	126.584	Sapato Preto 41
15/12/2008	476.930	Gorro Branco 56 (100%POL), Calça Camuflada 42, Gandola Camuflada 2, Capa Impermeável, Gorro Camuflado
15/12/2008	476.929	Calção Natação G
8/10/2007	298.735	Gandola Camuflada 4

O passo seguinte é a organização das sequências, juntando todos os itensets pela identificação do indivíduo, conforme a Tabela 4.

**Tabela 4 – Organização das sequências**

<b>Sequências</b>	<b>Itens</b>
1	{Sapato Preto 41}
2	{Gorro Branco 56 (100%POL), Calça Camuflada 42, Gandola Camuflada 2, Capa Impermeável, Gorro Camuflado}
3	{Calção Natação G}
4	{Gandola Camuflada 4}

Ao final, haverá um arquivo texto contendo 4 sequências de itens variados.

### **Passo 2: Mineração de padrões sequenciais**

Esta etapa é responsável por receber as sequências geradas no passo Passo 1: Tratamento e Formatação dos dados originais e executar cada um dos sete algoritmos de mineração de padrões sequenciais, adotados neste trabalho, aplicando os suportes mínimos e armazenando todos os resultados alcançados.

### **Passo 3: Interpretação dos resultados**

A interpretação de resultados pressupõe o desenvolvimento de três cenários didáticos sugeridos: cenário de curto prazo, avaliando padrões com tamanho de janela entre 0 e 30 dias; cenário de médio prazo, avaliando padrões com tamanho de janela entre 31 e 180 dias; e, Cenário de Longo

Prazo, avaliando padrões com tamanho de janela entre 181 e 365 dias. Tais cenários permitem uma visualização dos dados de maneira mais abrangente, sendo utilizados tanto na abordagem Ano-a-Ano, como na de 11 anos. Ao final, espera-se a identificação dos padrões mais demandados no tempo e no espaço, permitindo ações de planejamento mais adequadas para as demandas da Organização.

#### **Passo 4: Formulação da Conclusão**

Ao final do trabalho, espera-se a geração de condições favoráveis e necessárias à formulação de processos de compras (licitações) por parte da organização. Dentre os benefícios, espera-se o uso mais inteligente de recursos públicos pela identificação dos itens mais demandados, menor quantidade de itens armazenados em estoque e menor permanência dos itens em estoque. A próxima seção apresenta um detalhamento maior da solução adotada.

##### **3.1.1 Padrões Ordinários e ExtraOrdinários**

A Marinha do Brasil possui um catálogo de uniformes, conhecido com Regimento de Uniformes da Marinha do Brasil (RUMB)<sup>7</sup>, apresentando todos os tipos de uniformes possíveis e reconhecidos pela organização, identificados pela circunstâncias de utilização, como o camuflado para operações de campo.

A classificação, em padrões ordinários e extra-ordinários, é uma contribuição desse trabalho e deve ser avaliada, juntamente, com a taxonomia desenvolvida, tendo por base a Teoria da Classificação Facetada. Com ela, é possível estabelecer um parâmetro de diferenciação de padrões, propondo a ampliação do conhecimento da organização sobre a dinâmica de venda de fardamento, com vistas a facilitar sua previsão. Dessa forma, a taxonomia fornece a classificação dos elementos de um padrão que, em conjunto, formam a categorização do padrão. Como exemplo, o padrão formado por CALÇA AZUL 42, CAMISA BRANCA 3 e BOTA 43, de acordo com a taxonomia adotada, tem o primeiro e o segundo classificados como UNIFORME NÃO-OPERATIVO e o ultimo como CALÇADO OPERATIVO. O par CALÇADO OPERATIVO e UNIFORME NÃO-OPERATIVO formam a categoria que identifica o padrão.

<sup>7</sup> RUMB - Disponível em:  
[https://www.marinha.mil.br/com3dn/sites/www.marinha.mil.br.com3dn/files/Rumb%20parte%201\\_0.pdf](https://www.marinha.mil.br/com3dn/sites/www.marinha.mil.br.com3dn/files/Rumb%20parte%201_0.pdf)

Quando o padrão é constituído de partes que pertencem a um mesmo tipo de uniforme, é dito que este padrão é um padrão ordinário. Por exemplo, uma categoria que reúne CALÇADO OPERATIVO e UNIFORME OPERATIVO, para este dissertação, representa um padrão ordinário. Quando, ao contrário reúne classificações diferentes como, por exemplo, CALÇADO NÃO-OPERATIVO e UNIFORME OPERATIVO, é dito que o padrão é representado por uma categoria extraordinária, conseqüentemente, o padrão é considerado extraordinário.

### 3.2 Fluxo Principal da Solução

O fluxo principal apresenta todos os passos da solução em busca de um caminho alternativo para a obtenção da previsão de demanda, pormenorizando todas as fases envolvidas no processo, sendo representado pela Figura 5.

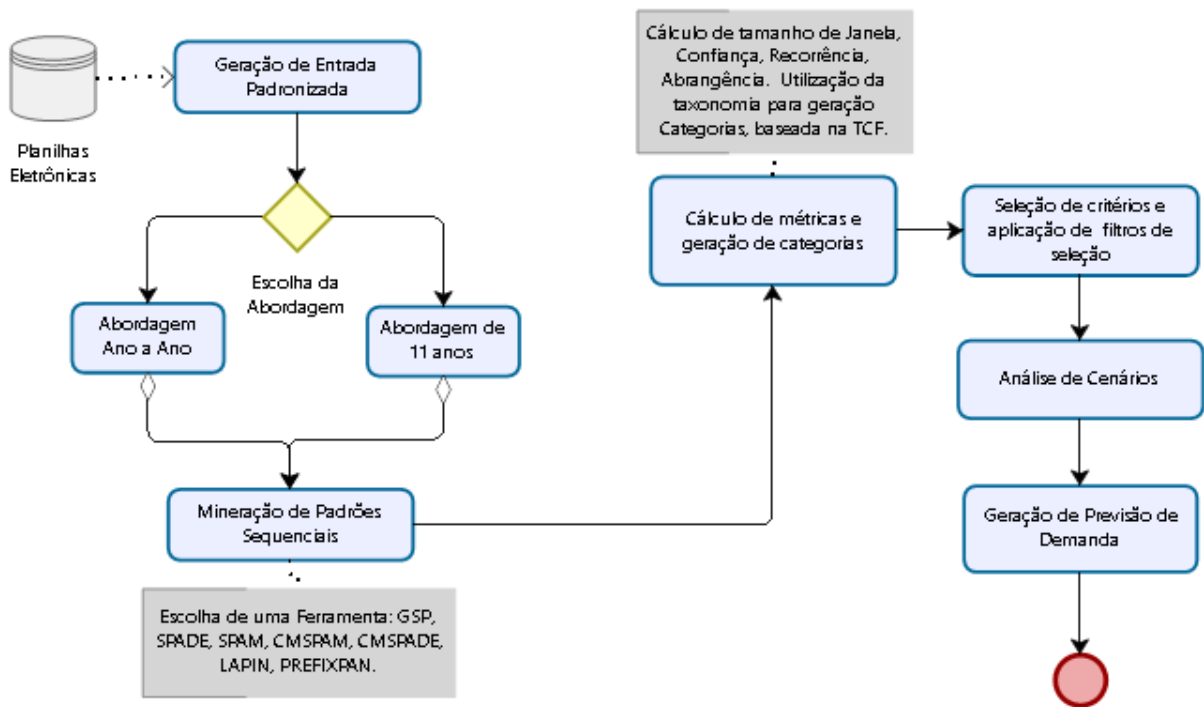


Figura 5 – Fluxo principal da solução

### 3.2.1 Geração de Entrada Padronizada

De acordo com a literatura sobre busca de padrões sequenciais, sobretudo na demonstração do problema original (SKIRANT, 1995), dois conceitos surgem com importância significativa: *Itemsets* e Sequências. Para os contornos desta pesquisa, o *Itemset* é um conjunto de itens e o conjunto de *Itemsets* representa uma Sequência.

Como mencionado na seção **Tratamento e Formatação dos dados originais**, os dados foram entregues na forma de planilhas eletrônicas, requerendo manipulações. Uma delas está em construir *Itemsets* e Sequências requeridos para a execução dos diversos algoritmos. A Tabela 5 é uma representação do extrato da planilha original, contendo duas notas fiscais, dois indivíduos e nove itens.

**Tabela 5 – Representação da Planilha com itens**

<b>Identidade</b>	<b>Nota Fiscal</b>	<b>Item</b>
1.000	126.584	INSIGNIA BRACO BEGE CB-IF
1.000	126.584	CALCA AZUL-MESCLA 38
1.000	126.584	BERMUDA FEM AZUL-MARINHO PRACA M
1.000	127.143	SAPATO PRETO 38
1.000	127.143	INSIGNIA BRACO AZUL-FERR CB-BA
2.000	97.913	CHAPEU BRANCO MN 59
2.000	97.913	INSIGNIA BRACO BRANCA CB-CO
2.000	97.913	INSIGNIA BRACO AZUL-FERR CB-CO
2.000	97.913	CALCAO AZUL GINASTICA OFICIAL XGG
2.000	258.391	CALCA FEM AZUL-MESCLA 44
2.000	258.391	CHAPEU BRANCO MN 60

Os *Itemsets* foram montados agrupando-se os itens, tendo por base a identidade e a nota fiscal. A Tabela 6 exhibe o agrupamento dos itens formando quatro entradas, cujas notas fiscais são 126.584, 127.143, 97.913 e 258.391, contendo um conjunto de itens, cada uma.

**Tabela 6 – Representação da Montagem do Itemset**

<b>Identidade</b>	<b>Nota Fiscal</b>	<b>Item</b>	<b>Compra</b>
1.000	126.584	INSIGNIA BRACO BEGE CB-IF, CALCA AZUL-MESCLA 38, BERMUDA FEM	A
	127.143	SAPATO PRETO 38, INSIGNIA BRACO AZUL-FERR CB-BA	B
2.000	97.913	CHAPEU BRANCO MN 59, INSIGNIA BRACO BRANCA CB-CO, INSIGNIA BRACO AZUL-FERR CB-CO, CALCAO AZUL GINASTICA OFICIAL XGG	C
	258.391	CALCA FEM AZUL-MESCLA 44, CHAPEU BRANCO MN 60	D

A construção das sequências se baseia nos dados gerados pelos Itemsets, impondo uma ordem temporal entre as compras, como exibido na Tabela 7, tendo a identidade do indivíduo como identificador da sequência.

**Tabela 7 – Representação das Sequências**

<b>Identidade</b>	<b>Sequência</b>
1.000	Compra A, Compra B
2.000	Compra C, Compra D

A Tabela 7 exhibe duas sequências adquiridas pelos indivíduos 1.000 e 2.000, ordenadas no tempo, compostas, respectivamente, pela compra A, seguida pela compra B e pela compra C, seguida pela Compra D. Ao final, a estrutura está pronta para a execução por todos os algoritmos utilizados neste trabalho.

### **3.2.2 Escolha da abordagem: Ano a Ano ou período de 11 anos**

As abordagens se propõem a investigar o comportamento da organização, analisando os padrões encontrados pela Mineração de Dados. Na abordagem Ano-a-Ano, a preocupação se concentra na determinação dos padrões cujo intervalo de repetição (janela) seja menor que 365 dias, buscando aqueles que são mais recorrentes, ou seja, aqueles que são observados em maior número de anos dentro do intervalo de 11 anos, e mais abrangentes, ou seja, aqueles que são identificados em mais localidades. dentre as 16 localidades possíveis. Tanto a recorrência como a abrangência são medidas desenvolvidas neste trabalho e podem ser consultadas no capítulo 4. Na abordagem de 11 anos, todo conjunto de dados é colocado em um único arquivo de sequência, sendo submetido a todos os algoritmos de Mineração de Dados utilizados nesta dissertação. O objetivo, neste trabalho, é identificar aqueles padrões que possuam janela maior que 365 dias, com a maior abrangência e confiança possíveis, descartando a recorrência.

### **3.2.3 Aplicação da Mineração de Padrões Sequenciais**

São sete os algoritmos escolhidos para a realização da Mineração de Padrões Sequenciais para esta dissertação: GSP, SPADE, SPAM, CMSPADE, CMSPAM, LAPIN e PREFIXPAN. Todos eles fazem parte da biblioteca SPMF<sup>8</sup> (FOURNIER-VIGER, 2014a), escolhida e utilizada como base deste trabalho, contendo um grande número de algoritmos já desenvolvidos. Cada um deles deve receber a entrada formatada, contendo as sequências, e liberar um arquivo de respostas contendo todos os padrões encontrados, mediante um suporte inicial definido pelo usuário. Em seguida, inicia-se um processo de tratamento dos padrões encontrados, começando com a remoção de dados indesejados, chamados de ruídos.

### **3.2.4 Cálculo de Métricas e geração de categorias**

O cálculo de métricas visa à obtenção de algumas medidas que servem de suporte para a classificação dos padrões de acordo com o interesse do momento. Dessa forma, será possível observar os padrões mais requeridos no tempo e/ou no espaço. Dentre as métricas desenvolvidas para este projeto, estão:

A Recorrência pode ser entendida como a observação de ocorrência de padrões no tempo. Ou seja, havia a necessidade de uma métrica que realizasse uma checagem da solicitação dos padrões ao longo do período de 11 anos, demonstrando que um padrão pudesse ser mais demandado que outro, fornecendo condições de uma comparação direta dos resultados. Em resumo permite observar padrões com maior número de solicitações a partir do histórico de compras.

A Abrangência pode ser entendida como a observação da ocorrência dos padrões distribuídos espacialmente. Esta métrica permite uma avaliação complementar à métrica de recorrência, avaliando quais padrões foram solicitados entre as 16 localidades. Ou seja, permitindo a observação do maior número de localidades que demandaram um padrão específico, espalhadas pelo território nacional.

Dentre as métricas utilizadas neste projeto, estão:

O cálculo de Tamanho Médio de Janela pode ser entendido como o intervalo médio entre idas aos postos de venda de itens de fardamento, ou seja, o intervalo entre duas compras sucessivas. O cálculo de janela foi proposto por AGRAWAL-SKIRANT (1995) e, também, é citado pela biblioteca SPMF, porém, em ambos os casos não foram desenvolvidos.

<sup>8</sup> SPMF: para informações: <http://www.philippe-fournier-viger.com/spmf/>

O cálculo de confiança pode ser entendido como o cálculo do interesse de aquisição de um item, tendo adquirido outro(s) antes. Esta é uma medida probabilística e seu cálculo é realizado mediante a probabilidade condicional, bastante conhecido nas áreas extatas. Ela é bastante utilizada na mineração de dados para a busca de padrões.

A utilização da taxonomia foi criada para montar a categoria que representa o padrão; Ou seja, os itens do padrão são lidos, sua classificação correspondente é recuperada e posta lado a lado, elas montam a classificação do padrão. A taxonomia baseada na Teoria de Classificação Facetada foi criada com essa necessidade, a de permitir um nível de diferenciação mais elaborado, permitindo a separa em padrões ordinários e extraordinários. Por exemplo, partindo-se do padrão P {CALÇA AZUL 44, BOTA 40 e CAMISA BRANCA 3}, o item Calça tem sua classificação recuperada da taxonomia como UNIFORME NÃO-OPERATIVO; o item Bota faz parte da classificação de CALÇADO OPERATIVO; e o item Camisa faz parte da classificação de UNIFORME NÃO-OPERATIVO. Juntando todas as classificações e eliminando as repetições, tem-se a categoria do padrão representada por [UNIFORME NÃO-OPERATIVO e UNIFORME NÃO-OPERATIVO]. Este procedimento é realizado para cada padrão recuperado pela mineração de padrões sequenciais. A próxima seção aplica filtros para selecionar apenas padrões interessantes.

### **3.2.5 Seleção de critérios e aplicação de Filtros**

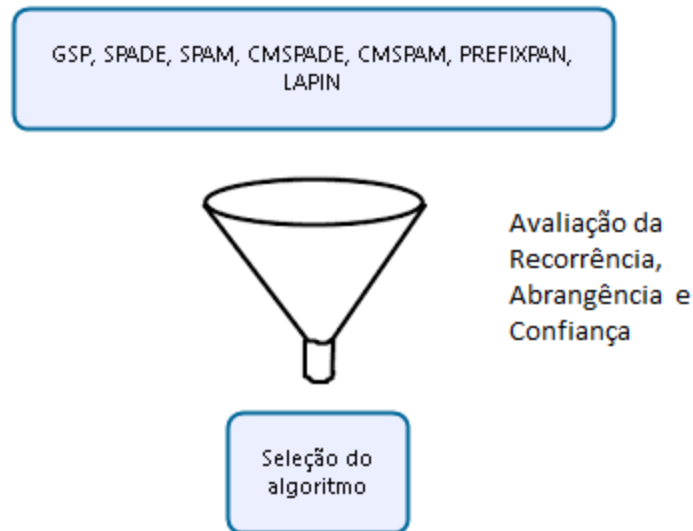
Após a etapa de cálculo de métricas, há a necessidade de aplicação de filtros e realização de seleção em dois momentos: para definir o algoritmo mais viável e com resultados mais promissores; da mesma forma, os filtros foram aplicados para selecionar os padrões mais interessantes.

Para todas as comparações realizadas, neste trabalho, foram consideradas as métricas recorrência e abrangência, interagindo da seguinte forma: primeiro, são selecionados os padrões com os maiores valores de recorrência (uma vez que a recorrência demonstra o interesse pelo padrão mencionado); se houver empate entre os valores da recorrência, a abrangência deve ser considerada; segundo, o empate deve ser decidido para o padrão que apresentar o maior valor para abrangência, fechando as avaliações.

Para definir o algoritmo mais promissor, foram selecionados os cinco padrões contendo os maiores valores para recorrência e abrangência, seguindo o procedimento do parágrafo anterior; as



médias para a recorrência e para a abrangência foram calculadas e, após isso, o algoritmo que apresentar os maiores valores serão utilizados para fornecer os padrões para a pesquisa. A idéia funciona, basicamente, pelo esquema apresentado na Figura 7, onde o funil representa a dinâmica de cálculos e comparações da recorrência e abrangência.



**Figura 6 – Representação da seleção do algoritmo**

Este procedimento deve acontecer para todos os três cenários aventados para este trabalho, ou seja, deve ser reproduzido para o cenário de curto, médio e longo prazos, revelando os padrões mais promissores para a montagem da previsão.

O segundo momento em que a aplicação de filtros deve ser realizada e na geração da previsão de demanda, onde de um conjunto de padrões, pode-se selecionar o/os mais solicitados para compor um novo processo de compra. A próxima seção apresenta a análise de cenários, definindo e apresentando cada um deles.

### 3.2.6 Análise de Cenários

Para efeitos didáticos, o espaço de tempo de 365 dias foi dividido e rotulado, com o objetivo de observar melhor o comportamento dos padrões. Assim, foram criadas três divisões, chamadas de cenários: de Curto Prazo, de Médio Prazo e de Longo Prazo.

Cada uma das divisões contém um número de dias, constituindo apenas uma divisão didática: Curto (0 a 30 dias); Médio (31 a 180 dias) e Longo (181 a 365 dias). Os cenários ajudarão a identificar padrões de saída breve, como também, de saída muito longa, de acordo com o perfil da

organização. Com os cenários identificados, resta a montagem da previsão de demanda propriamente dita, exibida na próxima seção.

### **3.2.7 Geração da Previsão de Demanda**

Ao final de todo o processo descrito, um conjunto de padrões considerados mais importantes ou solicitados é apresentado. Cada abordagem apresenta um conjunto dos três cenários propostos (Curto, Médio e Longo prazos) e, para cada um deles, uma lista de padrões é anexada. Cada padrão deve apresentar os maiores valores possíveis para as métricas desenvolvidas.

## 4 - Processos de Aquisição mais inteligentes pela análise e avaliação dos padrões de consumo

Neste capítulo, observam-se os passos para o desenvolvimento da solução, no intuito de conseguir identificar padrões relevantes que possibilitem alcançar processos de aquisição mais inteligentes.

### 4.1 Tratamento do Conjunto de dados

O primeiro passo desenvolvido, neste trabalho, foi a formatação do conjunto de dados, que envolve a construção das sequências (AGRAWAL-SKIRANT, 1995) de entrada para os diversos algoritmos da biblioteca SPMF. Os dados para a pesquisa foram recebidos na forma de planilhas eletrônicas, as quais foram tratadas, limpas e armazenadas no banco de dados do projeto. Em seguida, os dados de compras separados foram convertidos para o formato de entrada de processamento, resultando em uma estrutura similar ao da Tabela 8.

Tabela 8 – Exemplo de Sequências

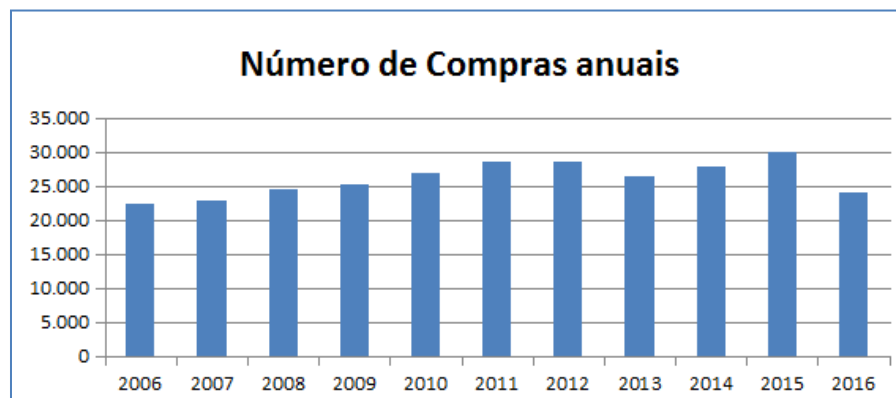
SID	Identificador	Conjunto de Compras
1	100.001	(CALÇA, SAPATO), (CINTO, BOLSA)
2	203.103	(CAMISA, BOLSA, CINTO)
3	410.345	(GORRO, BOTA, MEIA)

Na Tabela 8, três campos são importantes: o sequence id (SID), o identificador da compra, equivalendo ao número de uma nota fiscal e o conjunto de compras realizado pelo indivíduo, respeitando sua ordem de acontecimento. Como exemplo, o SID 1, tem o número de identificar 100.001 e conjunto de compras formado por Calça e Sapato, numa data qualquer, e por Cinto e Bolsa numa data posterior àquela de Calça e Sapato. Essa ordem fornece o indicativo de tempo necessário à operação das ferramentas. O conjunto de compras resultante e separados ano a ano, apresenta o comportamento observado na Tabela 9, cujo valor médio é de 26.044 sequências de indivíduos membros da organização.

**Tabela 9 – Número de sequências entre 2006 e 2016**

<b>Ano</b>	<b>Sequências</b>
2006	22.393
2007	22.869
2008	24.352
2009	25.177
2010	26.773
2011	28.461
2012	28.539
2013	26.325
2014	27.757
2015	29.926
2016	23.919

Visualizando, graficamente:

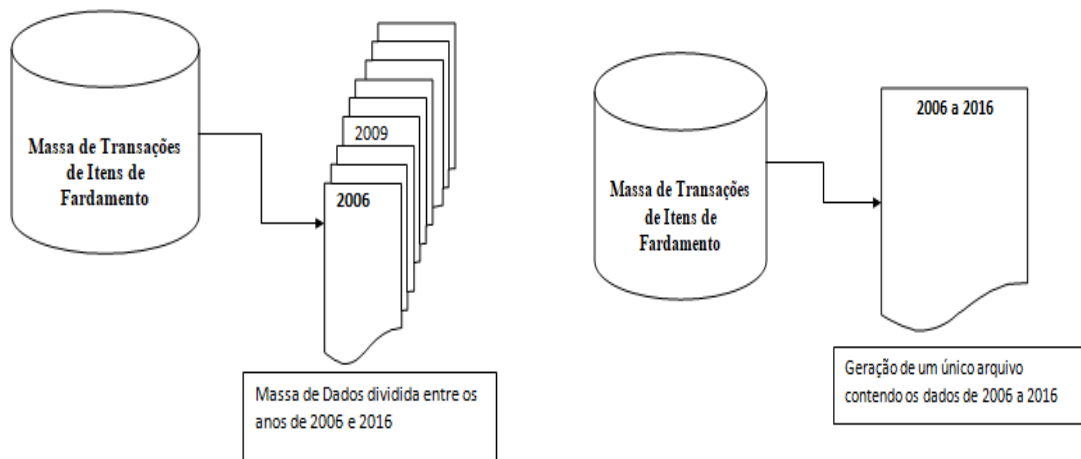


**Figura 7 – Distribuição das sequências no período de 2006 a 2016**

É possível observar um pico em 2011, 2012 e, novamente, em 2015 de número de sequências no período de 2006 a 2016, conforme Figura 8.

#### **4.2 Escolha da Abordagem: Ano a Ano ou Por 11 anos**

O segundo passo é estabelecer qual a abordagem que norteia a condução da pesquisa: Ano a Ano ou por 11 anos; Cada uma das abordagens se propõe a um objetivo diferente, onde a primeira avaliando a distribuição anual, com conseqüente seleção dos padrões com maior número de observações em anos não, necessariamente, consecutivos. A Figura 9 ilustra a abordagem.

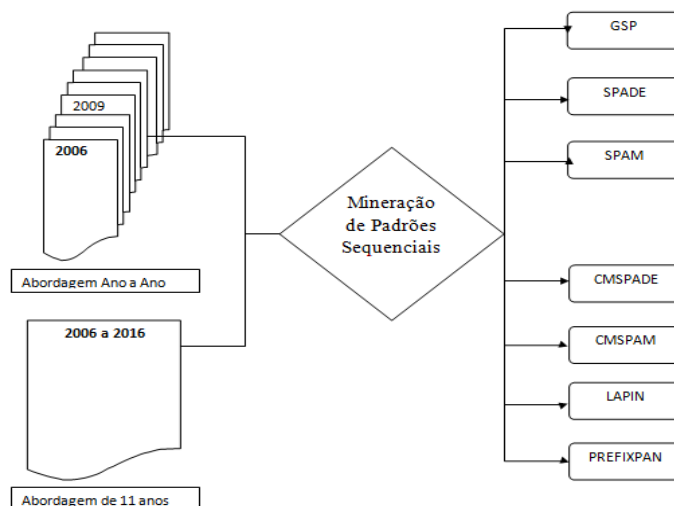


**Figura 8 – Representação das Abordagens Ano a Ano e de 11 anos**

A Abordagem de 11 anos, também representada na Figura 9, pretende realizar a exploração de padrões da massa integral dos dados, ou seja, pela reunião de todas as transações contidas em um único arquivo. A próxima seção exhibe o ambiente de apoio à realização do processo de Mineração de Padrões.

### 4.3 Busca de Padrões Sequenciais

De acordo com a literatura disponível, existem muitos algoritmos de Mineração de Dados Sequenciais desenvolvidos com propósitos diversos, porém, para os propósitos desta dissertação, serão utilizados os algoritmos disponibilizados pela biblioteca SPMF: GSP, SPADE, SPAM, CMSPADE, CMSPAM, LAPIN e PREFIXPAN.



**Figura 9 – Representação da Escolha da ferramenta de Mineração**

A Figura 10 processo representa a dinâmica da realização da Mineração de Padrões, uma vez escolhida a abordagem a ser utilizada. O resultado é um conjunto de padrões a serem analisados para verificação daqueles relevantes. A próxima seção apresenta o tratamento aplicado ao conjunto de padrões resultantes da Mineração.

#### 4.4 Tamanho Médio de Janela (TMJ)

O cálculo do tamanho médio de janela inicia o conjunto de medidas a serem obtidas logo após a execução dos algoritmos de mineração e dizem respeito ao intervalo de tempo, entre padrões semelhantes. Esta informação é relevante para efeitos de análise, quando o usuário tiver que tomar uma decisão motivada, por exemplo, por necessidades de gestão de estoque e previsão de demanda. Em seu artigo, AGRAWAL-SKIRANT (1995) definiu a base para o cálculo do tamanho da janela, que é:

$$\text{Equação 1: } \mathbf{FIM(S_i) - INÍCIO(S_i) \leq \text{Tamanho da Janela, } 1 \leq i \leq n,$$

no qual o tamanho da janela é definida pelo usuário e S é uma sequência.

O resultado da execução dos algoritmos de MDS, utilizando o framework SPMF, devolve um arquivo de texto contendo o conjunto de padrões resultantes da computação de padrões. Cada arquivo contém, também, um conjunto de SIDs (Sequence IDs), mostrando onde o padrão foi visto. Como no exemplo, abaixo:

**Padrão S:** <Arroz Feijão> <Açúcar>

**Sequence ID:** 1 3 10 17

O formato do padrão diz que, num primeiro momento, Arroz e Feijão são adquiridos; num segundo momento, Açúcar. Sabendo que para cada SID há um par de datas para correspondente a cada uma das partes do padrão, exibidos na Tabela 10.

**Tabela 10 – Lista de Padrões, SID's e datas**

<b>Padrão</b>	<b>SID</b>	<b>Data</b>
Arroz Feijão	1	10/01/2006
Açúcar	1	13/01/2006
Arroz Feijão	3	22/03/2006
Açúcar	3	30/03/2006
Arroz Feijão	10	1/06/2006
Açúcar	10	5/06/2006
Arroz Feijão	17	7/08/2006
Açúcar	17	19/08/2006

Para realizar o cálculo:

$$\text{Fórmula 3. } \mathbf{FIM}(S_{\langle \text{Arroz Feijão} \rangle} \langle \text{Açúcar} \rangle) - \mathbf{INÍCIO}(S_{\text{Açúcar}}) \leq \mathbf{Tamanho Janela}$$

A Tabela 11 apresenta a diferença entre as datas de duas compras sucessivas.

**Tabela 11 – Lista de SID e períodos**

<b>SID</b>	<b>Cálculo das diferenças das datas</b>	<b>Dias</b>
1	<b>FIM(13/01/2006) - INÍCIO(10/01/2006)</b>	3
3	<b>FIM(30/03/2006) - INÍCIO(22/03/2006)</b>	8
10	<b>FIM(05/06/2006) - INÍCIO(01/06/2006)</b>	4
17	<b>FIM(19/11/2006) - INÍCIO(07/08/2006)</b>	12

Calculando a média, o desvio padrão e o erro padrão.

$$\text{Média} = (3 + 8 + 4 + 12) / 4 = \mathbf{6.75 \text{ dias}}$$

$$\text{Desvio padrão} = \sqrt{[(3 - 6,75)^2 + (8 - 6,75)^2 + (4 - 6,75)^2 + (12 - 6,75)^2] / 4} \\ = \mathbf{3.56}$$

$$\text{erro padrão} = 3,56 / \sqrt{4} = \mathbf{1.78}$$

Dessa forma, o tamanho da janela será a média mais o desvio padrão, assim:  $6.75 \pm 3.56 = 10.31$  dias, com erro padrão de 1.78, para fins de avaliação da qualidade da amostra.

Supondo o tamanho de janela de 10 dias, definida pelo usuário, o padrão <Arroz Feijão> <Açúcar> figuraria como um padrão candidato às próximas etapas do processo, pois atende à equação 1.

#### **4.5 Cálculo de Recorrência e Abrangência**

Duas medidas importantes foram desenvolvidas, especificamente, para esta dissertação, com o objetivo de permitir a avaliação dos cenários construídos: Recorrência e Abrangência. A primeira, tem o objetivo de medir, ao longo do período de onze anos, quantas vezes um padrão específico foi observado, assim:

**Recorrência (x) = n, onde x é o padrão escolhido,**

**n o número de vezes que o padrão p aparece no período.**

Para alcançar os resultados, utiliza-se a recorrência percentual, assim:

$$\mathbf{Recorrência\% (x) = Recorrência(x) / T \text{ de anos}}$$

(nesta pesquisa, T é igual a 11 anos)

**Assim:  $0 \leq \text{Recorrência\%}(x) \leq 100\%$ , x é um padrão..**

O valor máximo representa um elemento, extremamente, recorrente. A Abrangência é uma medida espacial e/ou geográfica relacionada com a observação do padrão numa região específica do Brasil e/ou Navio. Este trabalho relacionou quinze (15) regiões mais Navio, compondo dezesseis localidades possíveis.

**Abrangência (x) = m, onde x é o padrão escolhido,**

**m o número de localidades em que o padrão p surge.**

Para alcançar os resultados, utiliza-se a abrangência percentual, assim:

**Abrangência% (x) = Recorrência(x) / L localidades**

(nesta pesquisa, L é igual a 16 anos)

**Assim:  $0 \leq \text{Abrangência\% (x)} \leq 100\%$ , x é um padrão.**

Onde o valor máximo representa um padrão, extremamente, abrangente. A próxima seção apresenta a taxonomia utilizada no trabalho.

#### **4.6 Construção da Taxonomia: Aplicação da Teoria da Classificação Facetada**

Após a análise do RUMB, cada item foi confrontado com as categorias PMEST da Teoria da Classificação Facetada, verificando-se o enquadramento nas vertentes Energia e Personalidade/Entidade, respectivamente, por fazerem menção a fenômenos da natureza ou eventos, como chuva ou frio e menção a objetos ou conjunto de coisas, como um sapato ou camisa. A partir daí, o processo de análise e síntese, embutido na teoria, é aplicado para cada item (conforme mencionado na seção 2.9) identificando o assunto básico, a idéia isolada, montando o assunto e atribuindo especificadores, para construir o assunto final. Cada faceta inclui características únicas e bastante representativas. De acordo com a Teoria existem, ainda, as categorias Matéria, (representando a propriedade da matéria), Tempo (representando a dimensão temporal) e Espaço (representando a dimensão física), que não apresentam representantes no conjunto de itens de fardamento, uma vez que não há menção a sua composição, menção a alguma ação presente ou ao aspecto geográfico.

Analisando os itens de fardamento, nota-se que, a categoria Energia tem duas facetas bem identificadas: Frio e Chuva. Os itens japonsa, luva e blusão estão relacionados a frio e capa à chuva. A categoria Personalidade/Entidade apresenta nove facetas: Uniforme Adorno, Uniforme Operativo, Uniforme Camuflado, Uniforme de Gala, Saída de Banho, Calçado, Uniforme Não-Operativo, Uniforme Prática Esportiva e Cozinha. Tanto a categoria Energia como a



Personalidade/Entidade podem ser observadas na Figuras 11 e na Figura 12. Aplicando a noção de especificador de CAMPOS (2001), obtém-se o resultado exibido na Tabela 12.

**Tabela 12 – Categoria Energia**

Assunto Básico	Ideia Isolada	Assunto (Composto)	Especificador	2º Especificador	Assunto Final
Vestuário	Casaco	Vestimenta de Frio	Japona	Num. 42 a 58	Vestimenta de Frio Japona [Nº]
			Blusão	Tam. P,M,G,GG	Vestimenta de Frio Blusão [Nº]
Vestuário	Mão	Abrigo para Chuva	Luva	Num. 70,75,80,85,90	Vestimenta de Frio Luva [Nº]
	Capa de Chuva		-	-	Abrigo para Chuva

A categoria Energia pode ser representada pela Figura 11. Devido às características de cada item, os especificadores apresentam denominações diferentes, baseadas na estrutura montada na Figura 11. Assim, a expressão ‘Num. 42 a 58’, quer dizer que há um intervalo de números que podem figurar num determinado item. Para o caso do Blusão (Tabela 12), o assunto final é Vestimenta de Frio Blusão [Nº], onde [Nº] pode ser substituído por um dos números mencionados formando o conjunto que vai de Vestimenta de Frio Blusão 42 até Vestimenta de Frio Blusão 58. O mesmo raciocínio vale para os outros especificadores. Alguns itens não possuem tamanho variável e, portanto, não possuem especificadores, enquanto outros vão apresentar um conjunto maior deles. A Tabela 13, Tabela 14 e a Tabela 15 detalham todos os assuntos finais utilizados pelo trabalho. A categoria Personalidade/Entidade pode ser apresentada pela estrutura da Figura 12.

**Tabela 13 – Categoria Personalidade/Entidade, com um especificador**

Assunto Básico	Ideia Isolada	Assunto (Composto)	Especificador	Assunto Final
Vestuário	Ambiente de Cozinha	Vestuário de Cozinha	Avental	Vestuário de Cozinha Avental

**Tabela 14 – Categoria Personalidade/Entidade, com dois especificadores**

Assunto Básico	Ideia Isolada	Assunto (Composto)	Especificador	Especificador	Assunto Final
Adorno	Uniforme	Adorno de Uniforme	Distintivo	Direção	Adorno de Unif. Distintivo Direção
				Aviador	Adorno de Unif. Distintivo Aviador
				Comportamento	Adorno de Unif. Distintivo Comportamento
				Punho	Adorno de Unif. Distintivo Punho
				Gola	Adorno de Unif. Distintivo Gola
Vestuário	Banho	Saída de Banho	Roupão	Lenço	Adorno de Unif. Lenço
				Cinto	Adorno de Unif. Cinto [Cor]
				Cor. [preta, verde-musgo, bege, branco]	Adorno de Unif. Cinto [Cor]
Vestuário	Camuflado	Uniforme Camuflado	Gandola	Tam [P,P,P,M,G,GG]	Saída de Banho Roupão [Tam.]
				[banho, rosto]	Saída de Banho Toalha [uso]
				Num 36 a 58	Uniforme Camuflado Calça [Nº]
				Num 1 a 5	Uniforme Camuflado Gandola [Nº]
			Gorro	-	Uniforme Camuflado Gorro

**Tabela 15 – Categoria Personalidade/Entidade, com três especificadores**

Assunto Básico	Ideia Isolada	Assunto (Composto)	Especificador	Especificador	Especificador	Assunto Final
Vestuário	Operativo	Uniforme Operativo	Macação	Cinza Azul Camuflado	Num. 38 a 60 Tam. PP,P,M,G,GG,... Num. 38 a 60	Uniforme Operativo Macacção Cinza [Nº] Uniforme Operativo Macacção Azul [Nº] Uniforme Operativo Macacção Camuflado [Nº]
Vestuário	Gala	Uniforme de Gala	Dólmã Jaquetão Luva Casaco*	Tam. [curto, médio, longo] Tam. [curto, médio, longo] [branca, verde-musgo, marrom] Azul	Num. 44 a 62 Num. 44 a 54 Tam.P,M,G,GG Num. 36 a 52	Uniforme Gala Dólmã [Tam] [Nº] Uniforme Gala Jaquetão [Tam] [Nº] Uniforme Gala Luva [cor] [Tam] Uniforme Gala Casaco azul [Tam]
Vestuário	Não-Operativo	Uniforme Não-Operativo	Calça Bermuda Blusa Camisa Jaleco Saia Gandola Gorro Gorro Boné Capa Boné Caxangá Maio	Cor [Azul, Bege,Branco,Cinza,Garança] Cor [azul, cinza] Cor [azul, branca, cinza] Cor [azul, branca, cinza, bege] Num. 42 a 58 Cor [azul, branca, cinza] Cor [bege, branco, azul] Cor [azul, branca, cinza, bege] Cozinheiro Cor [bege, cinza, branco] Cor [bege, cinza, branco] Num. 52 a 61 Tam [P,M,G,GG]	Num. 36 a 58 Num 36 a 56 Num 36 a 52 Num. 1 a 7 - Num 36 a 52 Num 1 a 7 - - - - - - - - - - - -	Uniforme Não-Operativo Calça [cor] [Nº] Uniforme Não-Operativo Bermuda [cor] [Nº] Uniforme Não-Operativo Blusa [cor] [Nº] Uniforme Não-Operativo Camisa [cor] [Nº] Uniforme Não-Operativo Jaleco [Nº] Uniforme Não-Operativo Saia [cor] [Nº] Uniforme Não-Operativo Gandola [cor] [Nº] Uniforme Não-Operativo Gorro [cor] Uniforme Não-Operativo Gorro cozinheiro Uniforme Não-Operativo Boné [cor] Uniforme Não-Operativo Capa Boné [cor] Uniforme Não-Operativo Caxangá [Nº] Uniforme Prática Esportiva Maio [Tam]
	Prática Esportiva	Uniforme Prática Esportiva	Calção Bermuda Camiseta Bota Coturno Tênis Sandália	Atividade [Ginástica, Natação] Tam [PP,P,M,G,GG,XGG] Cor [verde, branca] Num 38 a 48 Cor [Selva, Marrom] Num. 33 a 48 Cor [preto, marrom]	Tam [PP,P,M,G,GG,XGG] Tam [P,M,G,GG,XGG] - Num 38 a 48 - Num. 33 a 48	Uniforme Prática Esportiva Calção [Ativ.] [Tam] Uniforme Prática Esportiva Bermuda [Tam] Uniforme Prática Esportiva Camiseta [cor] [Tam] Calçado Operativo Bota [Nº] Calçado Operativo Coturno [cor] [Nº] Calçado Prática Esportiva [Nº] Calçado Prática Esportiva [cor] [Nº]
Calçado	Operativo	Calçado Operativo	Sapato Masculino Sapato Feminino Meia Palmilha	Cor [branco, preto] Cor [branco, preto] Cor [branco, preto, verde,azul] Num. 37 a 45	Num. 33 a 52 Num. 33 a 52 - -	Calçado Não-Operativo Sapato Masculino [cor] [Nº] Calçado Não-Operativo Sapato Feminino [cor] [Nº] Calçado Conforto Meia [cor] Calçado Conforto Palmilha [Nº]
	Prática Esportiva	Calçado Prática Esportiva				
	Não-Operativo	Calçado Não-Operativo				
	Conforto	Calçado Conforto				



**Figura 10 – Categoria Energia**

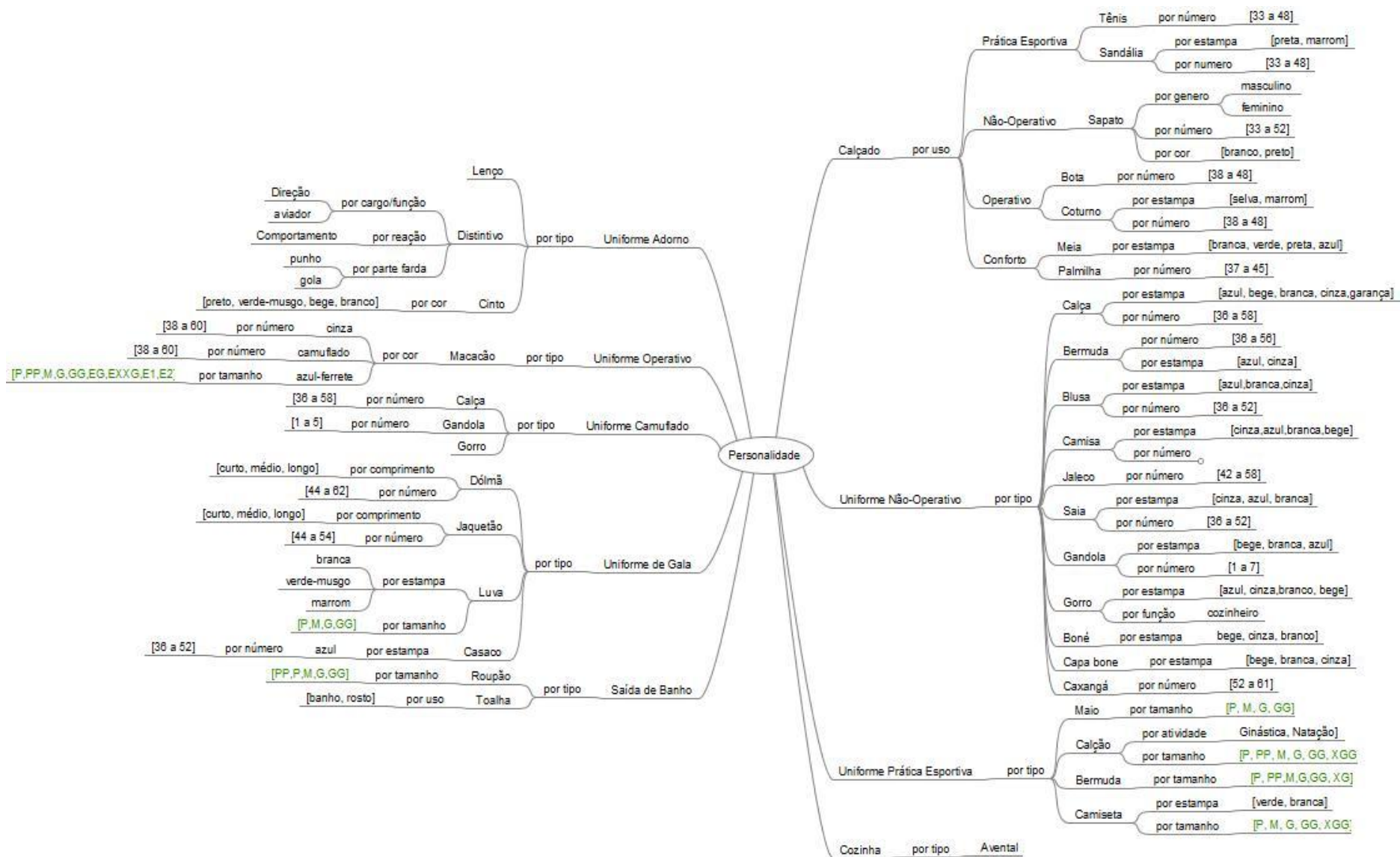


Figura 11 - Taxonomia, Teoria da Classificação Facetada, Categoria Personalidade

Neste ponto, todos os assuntos finais estão construídos, os quais servirão de base para a definição das categorias aplicadas aos padrões encontrados. As categorias serão compostas de agrupamentos de mais de um assunto final, permitindo a identificação de padrões ordinários e outros extraordinários para controle da organização pública. A próxima seção apresenta os diversos cenários identificados para a formação da previsão de demanda.

## 4.7 Seleção do algoritmo para a geração dos resultados

A aplicação do critério de escolha dos algoritmos permite a identificação da melhor ferramenta a ser utilizada para um cenário específico, em cada uma das duas abordagens. O objetivo é identificar os padrões mais demandados, em número de solicitações e em termos de localidade, assim, o critério adotado utilizou as métricas recorrência e abrangência. Dessa forma, para selecionar o algoritmo mais adequado, foi necessário identificar os padrões mais relevantes, calcular a média da métrica recorrência; em caso de empate, a maior média da métrica abrangência. O algoritmo que apresentar os maiores valores para recorrência e abrangência deve ser escolhido.

### 4.7.1 Abordagem Ano a Ano

Apresentação dos diversos algoritmos e de seus respectivos valores para a recorrência e abrangência. Para o cenário de curto prazo, de acordo com os dados da Tabela 16, o GSP apresenta os melhores valores para os padrões extraordinários e o CMSPADE para os padrões ordinários.

**Tabela 16 – Abordagem Ano a Ano, cenário de curto prazo**

Médias para recorrência e abrangência para cada algoritmo							
Algoritmos	GSP	SPADE	SPAM	LAPIN	PREFIXPAN	CMSPADE	CMSPAM
<b>Padrões extraordinários</b>							
Recorrência	69,01	56,36	58,18	67,27	54,55	58,18	65,45
Abrangência	96,25	63,75	71,25	78,75	97,5	92,5	77,5
<b>Padrões ordinários</b>							
Recorrência	43,63	56,36	54,54	61,82	52,72	65,45	60,00
Abrangência	88,75	90	86,25	93,75	98,75	91,25	96,25

Para o cenário de médio prazo, o algoritmo CMSPAM apresenta os melhores valores, de acordo com a Tabela 17, somente para os padrões extra-ordinários. Neste cenário, não foram observados resultados para os padrões ordinários, retornando valores em zero.

**Tabela 17 – Abordagem Ano a Ano, cenário de médio prazo**

Médias para recorrência e abrangência para cada algoritmo							
Algoritmos	GSP	SPADE	SPAM	LAPIN	PREFIXPAN	CMSPADE	CMSPAM
<b>Padrões extraordinários</b>							
Recorrência	0	34,54	32,7	0	0	35,54	41,81
Abrangência	0	51,25	52,5	0	0	48,75	45,00
<b>Padrões ordinários</b>							
Recorrência	0	0	0	0	0	0	0
Abrangência	0	0	0	0	0	0	0

O cenário de longo prazo tem vários algoritmos apresentando valores em zero para os padrões extra-ordinários. O melhor valor é exibido pelo SPADE, de acordo com a Tabela 18.

**Tabela 18 – Abordagem Ano a Ano, cenário de longo prazo**

Médias para recorrência e abrangência para cada algoritmo							
Algoritmos	GSP	SPADE	SPAM	LAPIN	PREFIXPAN	CMSPADE	CMSPAM
<b>Padrões extraordinários</b>							
Recorrência	0	0	0	0	0	9,09	<b>9,09</b>
Abrangência	0	0	0	0	0	8,75	<b>10</b>
<b>Padrões ordinários</b>							
Recorrência	0	0	0	0	0	0	0
Abrangência	0	0	0	0	0	0	0

Isto encerra a seleção de algoritmos para a abordagem Ano a Ano.

#### 4.7.2 Abordagem 11 anos

A abordagem de 11 anos apresenta os seguintes resultados. Para o cenário de curto prazo, nesta abordagem, o SPADE apresenta as melhores médias, apresentados na Tabela 19.

**Tabela 19 – Abordagem de 11 anos, cenário de curto prazo**

Médias para recorrência e abrangência para cada algoritmo							
Algoritmos	GSP	SPADE	SPAM	LAPIN	PREFIXPAN	CMSPADE	CMSPAM
<b>Padrões extraordinários</b>							
Abrangência	93,75	96.75	93,75	93,75	75	93,75	93,75
Confiança	100,00	97.87	100	100	100	35,84	19,45
<b>Padrões ordinários</b>							
Abrangência	87,5	100.00	93,75	93,75	87,5	93,75	93,75
Confiança	100.00	100.00	66,3	73,07	100	100	19,45

Para o cenário de médio prazo, o SPADE, também, apresenta os melhores resultados, como visto na Tabela 20.

**Tabela 20 – Abordagem de 11 anos, cenário de médio prazo**

Médias para recorrência e abrangência para cada algoritmo							
Algoritmos	GSP	SPADE	SPAM	LAPIN	PREFIXPAN	CMSPADE	CMSPAM
<b>Padrões extra-ordinários</b>							
Abrangência	0	81,25	75	75	0	31,25	50
Confiança	0	50	50	50	0	34,68	50
<b>Padrões ordinários</b>							
Abrangência	0	0	0	0	0	0	0
Confiança	0	0	0	0	0	0	0

O cenário de longo prazo exibe situação interessante, no qual nenhum resultado foi encontrado, como mostrado na Tabela 21.

**Tabela 21 – Abordagem de 11 anos, cenário de longo prazo**

Seleção do algoritmo para geração de resultados							
Algoritmos	GSP	SPADE	SPAM	LAPIN	PREFIXPAN	CMSPADE	CMSPAM
<b>Padrões extraordinários</b>							
Abrangência	0	0	0	0	0	0	0
Confiança	0	0	0	0	0	0	0
<b>Padrões ordinários</b>							
Abrangência	0	0	0	0	0	0	0
Confiança	0	0	0	0	0	0	0

Uma avaliação preliminar pode indicar que o suporte utilizado para realizar a pesquisa pode ser inadequado, requerendo um novo valor, este sendo mais de acordo com as

necessidades. A idéia inicial era avaliar o desempenho das abordagens com os mesmo valores iniciais, convencionando-se o suporte de 0.01 para uso nas duas abordagens. Como trabalho inicial, o número de resultados foi fixado em dois mil, sugerindo utilização do suporte citado. O próximo passo é a avaliação dos padrões recuperados em cada cenário mencionado.

## 4.8 Análise de padrões de consumo

Esta seção analisa os padrões recuperados para cada uma das abordagens e seus respectivos cenários, definidos para o projeto.

### 4.8.1 Abordagem Ano a Ano

A Abordagem Ano a Ano propõe-se a investigar o comportamento de consumo, dentro da organização pública, no espaço de tempo inferior a um ano (ou seja, 365 ou 366 dias), identificando os padrões de compra para os itens de fardamento mais relevantes. A intenção é a de possibilitar uma melhor tomada de decisões sobre a ação de adquirir bens.

O conjunto recuperado de padrões pode apresentar uma quantidade considerável de resultados viáveis. Nesta pesquisa, adotou-se, didaticamente, a escolha os cinco primeiros padrões mais representativos, considerados para a geração da previsão de demanda.

Os cenários são apresentados, em ordem, além de exibir os padrões ordinários e extraordinários de cada um deles, começando pelo cenário de curto prazo. Cada cenário foi construído baseado na seção 4.7.

#### 4.8.1.1 Cenário de Curto Prazo e padrões ordinários

##### a) Padrão {Sapato Preto, Blusa Azul ou Branca}

O padrão {Sapato Preto e Blusa Azul ou Branca} apresenta seis resultados importantes, com variados níveis de solicitação (Recorrência e Abrangência), conforme Tabela 22.

**Tabela 22 – Resultados para o padrão {Sapato Preto, Blusa Azul ou Branca}**

Ordem	Padrão importantes	%Rec	%Abr
1	{SAPATO PRETO 38, BLUSA AZUL 40}	63,64	68,75
2	{SAPATO PRETO 38, BLUSA AZUL 52}	63,64	75,00
3	{SAPATO PRETO 46, BLUSA AZUL 40}	72,73	93,75
4	{SAPATO PRETO 46, BLUSA AZUL 52}	18,19	43,75
5	{SAPATO PRETO 38, BLUSA BRANCA 36}	54,55	87,5
6	{SAPATO PRETO 38, BLUSA BRANCA 46}	27,27	81,25

A Figura 13 e a Figura 14 apresentam as medidas de Recorrência e Abrangência, respectivamente, em forma gráfica.

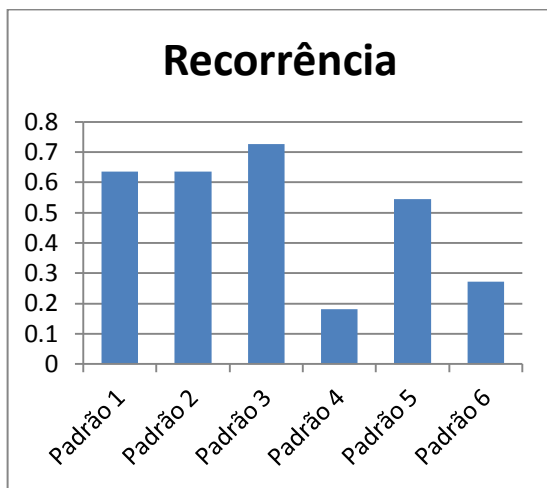


Figura 12 – Recorrência dos padrões da tabela 22

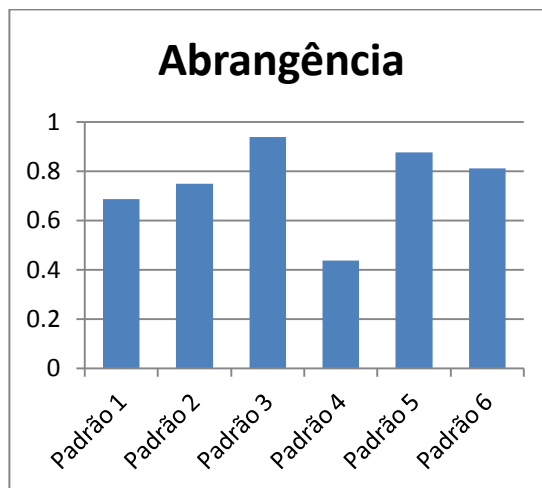


Figura 13 – Abrangência dos padrões da tabela 22

Uma rápida observação pelos padrões específicos revela que há três itens básicos presentes: Sapato Preto, Blusa Azul, Blusa Branca. O padrão 3 {SAPATO PRETO 46, BLUSA AZUL 40} apresenta o maior valor de recorrência e abrangência, sendo, dentre todos, o mais requisitado.

#### b) Padrão {Calção Azul P e Blusa Azul ou Branca}

O padrão {Calção Azul P e Blusa Azul ou Branca} apresenta três resultados importantes com variados níveis de solicitação (Recorrência e Abrangência), conforme Tabela 23.

Tabela 23 – Resultados para o padrão {Calção Azul P e Blusa Azul ou Branca}

Ordem	Padrões importantes	%Rec	%Abr
1	{CALÇÃO AZUL P, BLUSA AZUL 40}	43,75	93,75
2	{ CALÇÃO AZUL P, BLUSA AZUL 52}	50,00	81,25
3	{ CALÇÃO AZUL P, BLUSA BRANCA 36}	6,25	56,25

A Tabela 23 exhibe o conjunto de padrões, enquanto a Figura 15 e a Figura 16 apresentam os dados da Recorrência e Abrangência.



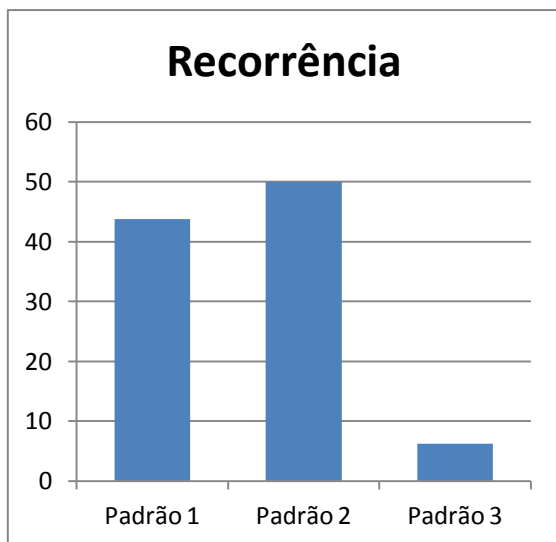


Figura 14 – Recorrência dos padrões da tabela 23

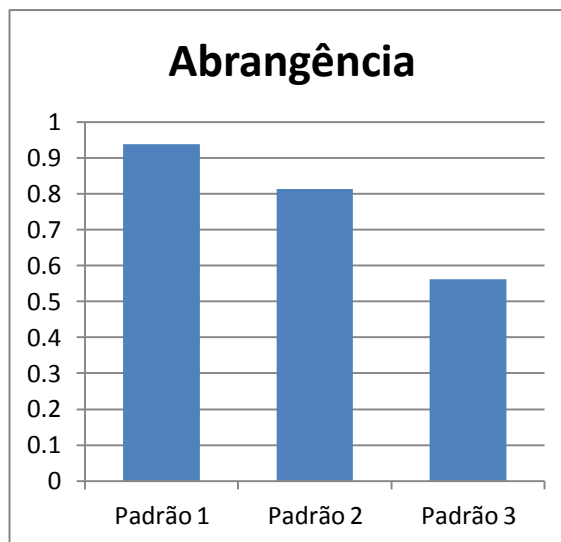


Figura 15 – Abrangência dos padrões da tabela 23

Segundo os dados da Figura 15 e da Figura 16, o padrão {CALÇÃO AZUL P, BLUSA AZUL 52} apresenta os maiores indicadores.

### c) Padrão {Sapato e Calça}

O padrão {Sapato e Calça} apresenta treze resultados importantes com variados níveis de solicitação (Recorrência e Abrangência), conforme Tabela 24.

Tabela 24 – Resultados para o padrão {Sapato e Calça}

Ordem	Padrões importantes	%Rec	%Abr
1	{SAPATO PRETO 38, CALÇA AZUL 38}	63,64	100
2	{SAPATO PRETO 38, CALÇA AZUL 54}	63,64	100
3	{SAPATO PRETO 38, CALÇA AZUL 40}	36,37	56,25
4	{SAPATO PRETO 38, CALÇA AZUL 44}	45,45	87,5
5	{SAPATO PRETO 40, CALÇA AZUL 38}	18,19	87,5
6	{SAPATO PRETO 40, CALÇA AZUL 54}	27,27	100
7	{SAPATO PRETO 46, CALÇA AZUL 54}	63,64	100
8	{SAPATO PRETO 38, CALÇA CINZA 48}	27,27	93,75
9	{SAPATO PRETO 38, CALÇA CINZA 50}	18,18	6,25
10	{SAPATO PRETO 46, CALÇA BRANCA 38}	63,64	56,25
11	{SAPATO PRETO 38, CALÇA BRANCA 38}	45,45	87,5
12	{SAPATO PRETO 38, CALÇA BEGE 48}	36,37	87,5
13	{SAPATO PRETO 38, CALÇA CAMUFLADA 48}	36,37	100

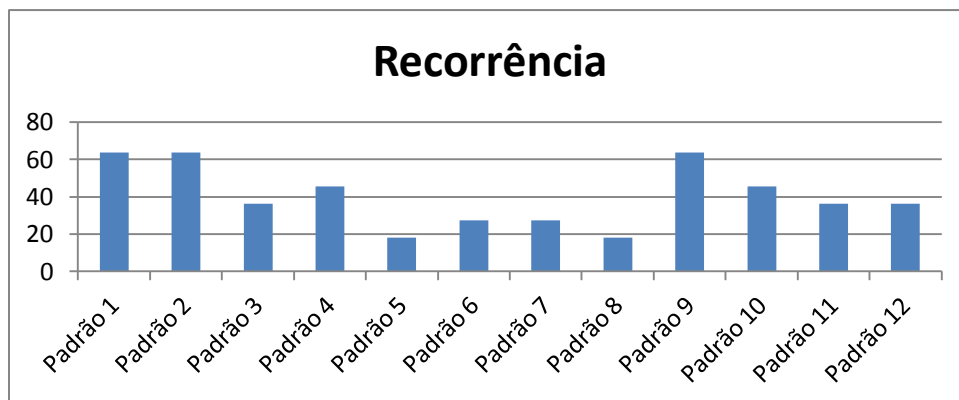


Figura 16 – Recorrência dos padrões da tabela 24

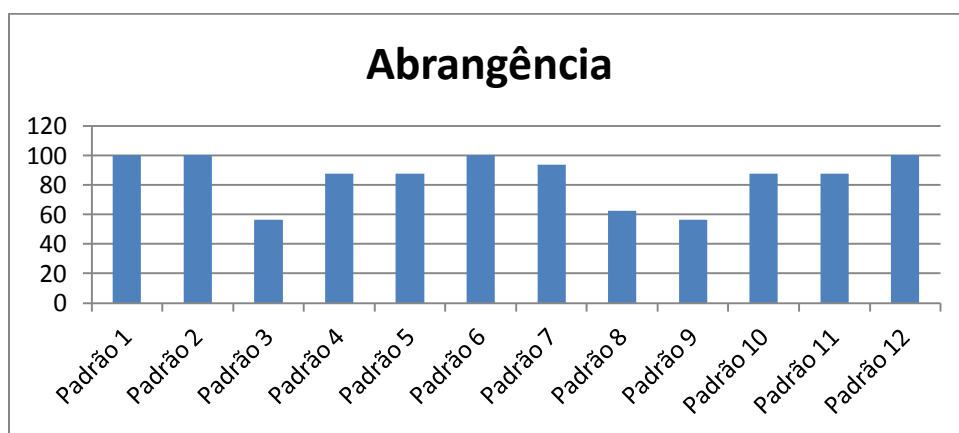


Figura 17 – Abrangência dos padrões da tabela 24

As Figura 17 e Figura 18 apresentam a distribuição dos valores da recorrência e abrangência dos diversos padrões, sendo que os padrões {SAPATO PRETO 38, CALÇA AZUL 54}, {SAPATO PRETO 38, CALÇA AZUL 38}, {SAPATO PRETO 46, CALÇA AZUL 54} apresentam as maiores ocorrências.

#### d) Padrão {Gandola e Sapato}

O padrão {Gandola e Sapato Preto} apresenta três resultados importantes com variados níveis de solicitação (Recorrência e Abrangência), conforme Tabela 25.

Tabela 25 - Resultados para o padrão {Gandola e Sapato}

Ordem	Padrões importantes	%Rec	%Abr
1	{GANDOLA AZUL 3, SAPATO PRETO 38}	54,55	87,5
2	{GANDOLA AZUL 3, SAPATO PRETO 46}	16,67	63,64
3	{GANDOLA BRANCA 2, SAPATO PRETO 38}	16,67	18,19

A distribuição dos padrões são exibidos na Figura 19 e na Figura 20.

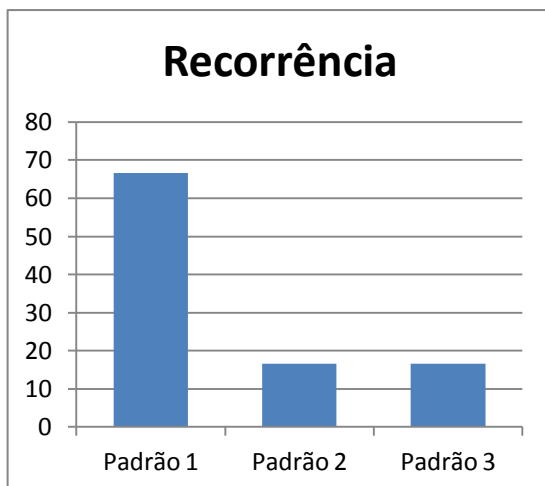


Figura 18 – Recorrência dos padrões da tabela 25

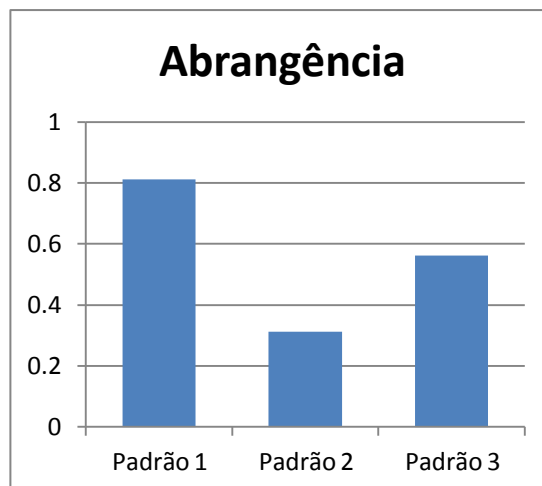


Figura 19 – Abrangência dos padrões da tabela 25

O padrão 1 {GANDOLA AZUL 3, SAPATO PRETO 38} é responsável por mais de 50% das solicitações, como mostrado na Figura 19 e na Figura 20.

**e) Padrão {Sapato Preto, Calção Azul e Calça}**

O padrão {Sapato Preto, Calção Azul e Calça} apresenta cinco resultados importantes com variados níveis de solicitação (Recorrência e Abrangência), conforme Tabela 26.

**Tabela 26 – Resultados para o padrão {Sapato Preto, Calção Azul e Calça}**

Ordem	Padrões importantes	Rec	Abr
1	{SAPATO PRETO 38, CALÇÃO AZUL P, CALÇA AZUL 38}	27,27	93,75
2	{SAPATO PRETO 38, CALÇÃO AZUL P, CALÇA AZUL 54}	27,27	81,25
3	{SAPATO PRETO 38, CALÇÃO AZUL P, CALÇA BRANCA 38}	27,27	75,00
4	{SAPATO PRETO 38, CALÇÃO AZUL P, CALÇA CINZA 48}	18,18	62,5
5	{SAPATO PRETO 38, CALÇÃO AZUL P, CALÇA AZUL 44}	4,54	18,75

A distribuição dos padrões, em termos de recorrência e abrangência, pode ser observada na Figura 21 **Erro! Fonte de referência não encontrada.** e na Figura 22.

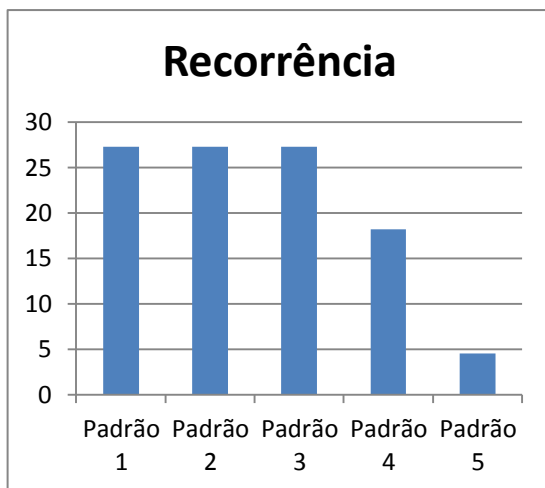


Figura 20 – Recorrência dos padrões da tabela 26

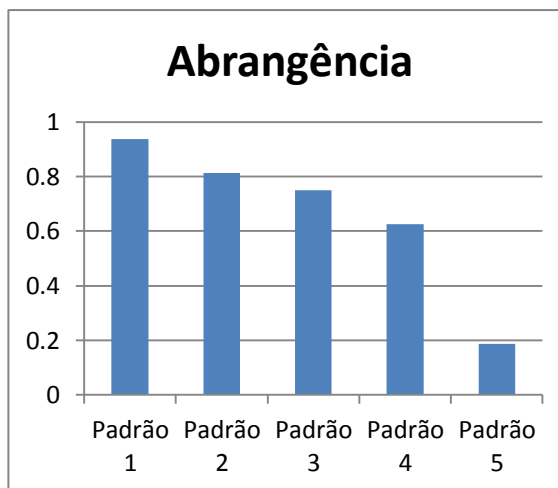


Figura 21 – Abrangência dos padrões da tabela 26

O padrões {SAPATO PRETO 38, CALÇÃO AZUL P, CALÇA AZUL 38} e {SAPATO PRETO 38, CALÇÃO AZUL P, CALÇA AZUL 54} são os mais solicitados, contendo os maiores valores de recorrência e abrangência.

#### 4.8.1.2 Cenário de Curto Prazo e Padrões Extra-ordinários

Os padrões conhecidos são fornecidos mediante os dados recuperados do algoritmo GSP (seção 2.7.1), conforme os critérios de seleção de algoritmos seção 4.7.

##### a) Padrão {Calça Azul e Bota}

O padrão {Calça Azul e Bota} apresenta oito resultados importantes com variados níveis de solicitação (Recorrência e Abrangência), conforme frequência exibida na Tabela 27.

Tabela 27 – Resultados para o padrão {Calça Azul e Bota}

Ordem	Padrões importantes	%Rec	%Abr
1	{CALCA AZUL 54, BOTA 40}	90,9	100
2	{CALCA AZUL 54, BOTA 43}	90,9	100
3	{CALCA AZUL 38, BOTA 40}	63,63	93,75
4	{CALCA AZUL 42, BOTA 40}	36,36	100
5	{CALCA AZUL 42, BOTA 43}	9,09	81,25
6	{CALCA AZUL 38, BOTA 43}	36,36	100
7	{CALCA CAMUFLADA 42, BOTA 43}	9,09	75,00
8	{CALCA CAMUFLADA 42, BOTA 40}	27,27	100

A Figura 23 e a Figura 25 apresentam os valores recorrência e abrangência para o padrão {Calça Azul e Bota}

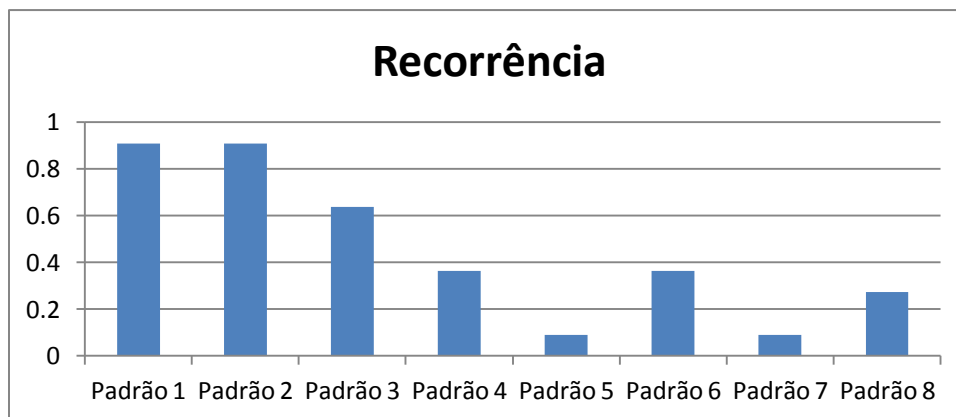


Figura 22 – Recorrência dos padrões da tabela 27

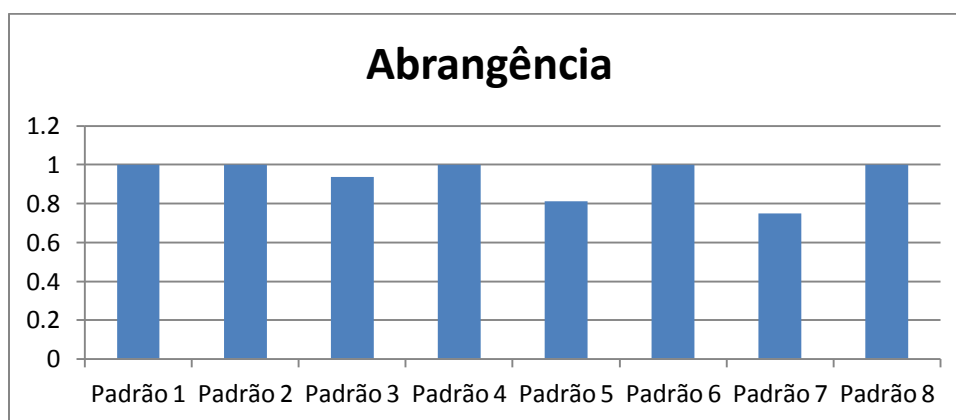


Figura 23 – Abrangência dos padrões da tabela 27

Os padrões {CALCA AZUL 54, BOTA 40} e {CALCA AZUL 54, BOTA 43} são os mais requisitados.

#### b) Padrão {Sapato Preto e Bota}

O padrão {Sapato Preto e Bota} apresenta três resultados importantes com variados níveis de solicitação (Recorrência e Abrangência), conforme frequência exibida na Tabela 28.

Tabela 28 – Resultados para o padrão {Sapato Preto e Bota}

Ordem	Padrões importantes	%Rec	%Abr
1	{SAPATO PRETO 38, BOTA 43}	31,57	100
2	{SAPATO PRETO 38, BOTA 40}	47,36	100
3	{SAPATO PRETO 46, BOTA 40}	21,05	93,75

A Figura 25 e a Figura 26 exibem a distribuição dos valores de recorrência e abrangência para o padrão {Sapato Preto e Bota}.

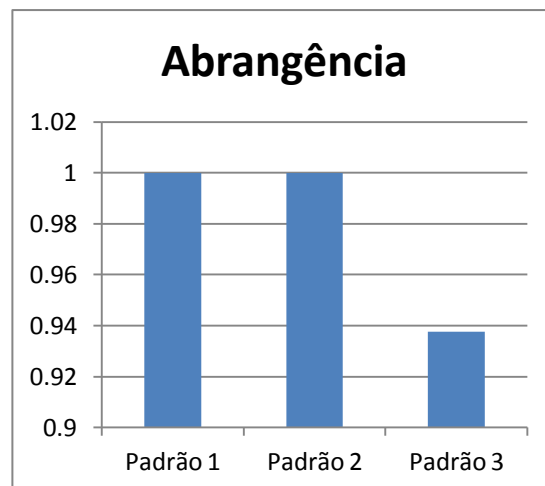
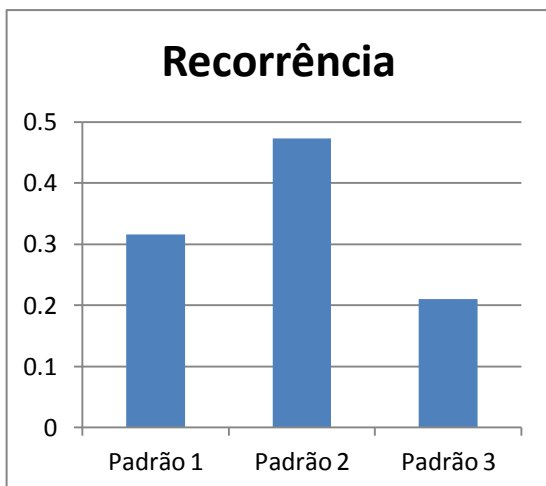


Figura 24 – Recorrência dos padrões da tabela 28

Figura 25 – Abrangência dos padrões da tabela 28

O padrão {SAPATO PRETO 38, BOTA 40} é o mais requisitado, tendo o maior valor para a recorrência.

**c) Padrão {Calça Azul e Dólmã}**

O padrão {Calça Azul e Dólmã} apresenta dezoito resultados com variados níveis de solicitação (Recorrência e Abrangência), conforme a Tabela 29 exibida a seguir.

**Tabela 29 – Resultados para o padrão {Calça Azul e Dólmã}**

Ordem	Padrões importantes	%Rec	%Abr
1	{CALCA AZUL 38, DÓLMÃ 48}	9,09	81,25
2	{CALCA AZUL 38, DÓLMÃ LONGO 46}	9,09	43,75
3	{CALCA AZUL 38, DÓLMÃ LONGO 56}	18,18	68,75
4	{CALCA AZUL 38, DÓLMÃ MÉDIO 52}	9,09	87,5
5	{CALCA AZUL 38, DÓLMÃ MÉDIO 50}	9,09	75,00
6	{CALCA AZUL 42, DÓLMÃ 46}	9,09	68,75
7	{CALCA CAMUFLADA 42, DÓLMÃ 46}	18,18	93,75
8	{CALCA CAMUFLADA 42, DÓLMÃ 48}	9,09	87,5
9	{CALCA AZUL 38, DÓLMÃ MÉDIO 52}	18,18	87,5
10	{CALCA AZUL 38, DÓLMÃ 46}	27,27	87,5
11	{CALCA AZUL 54, DÓLMÃ 46}	63,63	100
12	{CALCA AZUL 54, DÓLMÃ 48}	9,09	81,25
13	{CALCA AZUL 54, DÓLMÃ 50}	18,18	75,00
14	{CALCA AZUL 54, DÓLMÃ LONGO 46}	9,09	50,00
15	{CALCA AZUL 54, DÓLMÃ LONGO 56}	9,09	50,00
16	{CALCA AZUL 54, DÓLMÃ MÉDIO 50}	9,09	50,00
17	{CALCA AZUL 54, DÓLMÃ MÉDIO 52}	27,27	87,5
18	{CALCA AZUL 54, DÓLMÃ 56}	18,18	62,5

A Figura 27 e Figura 28 apresentam a distribuição da recorrência e abrangência para o padrão {CALCA AZUL e DÓLMÃ}.

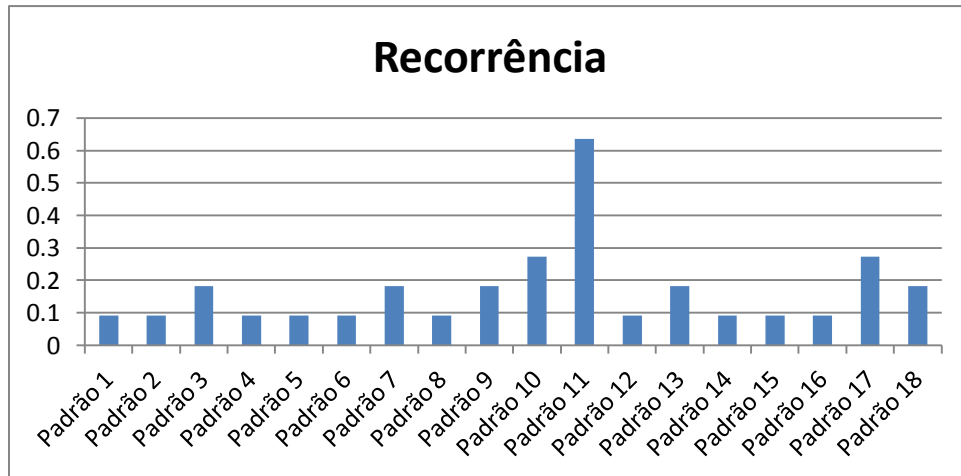


Figura 26 – Recorrência dos padrões da tabela 29

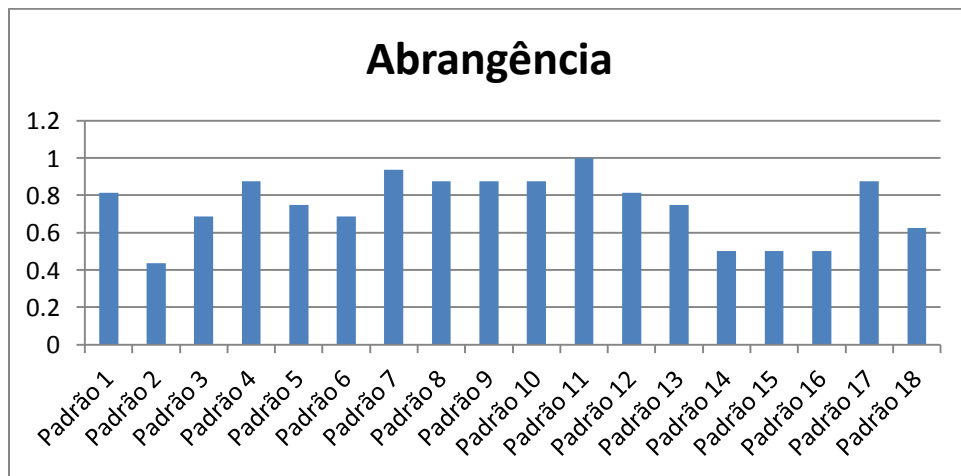


Figura 27 – Abrangência dos padrões da tabela 29

O padrão {CALCA AZUL 54, DÓLMÃ 46} é o mais requisitado, apresentando valores para recorrência e abrangência significativos.

**d) Padrão {Bota e Caxangá}**

O padrão {Bota e Caxangá} apresenta 1 resultado com características específicas, na forma de (Bota 40, Caxangá 59). Ou seja, 100% das solicitações desse padrão têm a associação do número 40 da Bota com o número 59 do Caxangá.

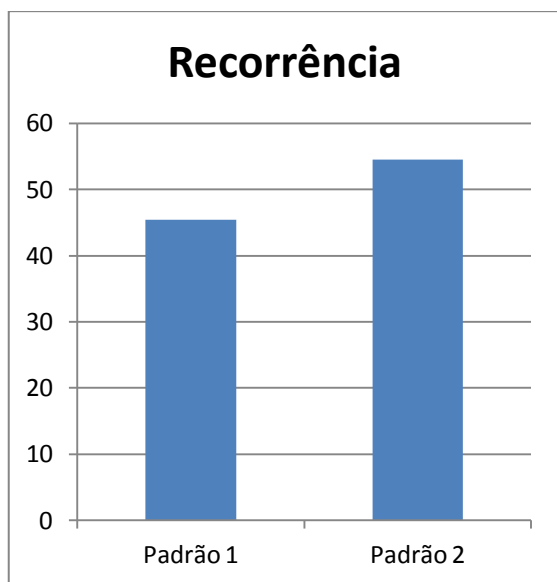
**e) Padrão {Calção Azul e Bota}**

O padrão {Calção Azul e Bota} apresenta dois resultados importantes com variados níveis de solicitação (Recorrência e Abrangência), conforme aTabela 30, exibida a seguir.

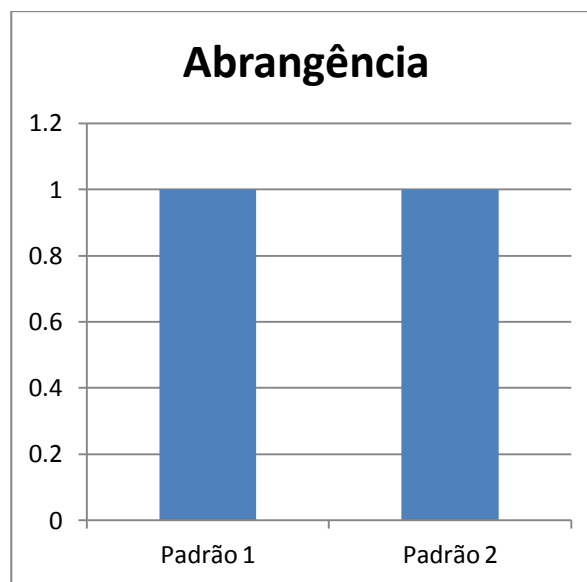
**Tabela 30 – Resultados para o padrão {Calça Azul e Bota}**

Ordem	Padrões importantes	%Rec	%Abr
1	{CALCAO AZUL P, BOTA 43}	45,45	100
2	{CALCAO AZUL P, BOTA 40}	54,54	100

A Figura 29 e Figura 30 apresentam a distribuição dos valores de recorrência e abrangência para o padrão {Calção Azul e Bota}.



**Figura 28 – Recorrência dos padrões da tabela 30**



**Figura 29 – Abrangência dos padrões da tabela 30**

O padrão {CALCAO AZUL P, BOTA 40} é o mais requisitado, de acordo com os valores exibidos.

#### **4.8.1.3 Cenário de Médio Prazo e Padrões Novos**

Os padrões extraordinários são fornecidos mediante os dados recuperados do algoritmo CMSPAM (seção 2.7.1), conforme os critérios de seleção de algoritmos seção 4.7.

##### **a) Padrão {Calça Azul, Sapato Preto e Blusa Azul}**

O padrão {Calça Azul, Sapato Preto e Blusa Azul} apresenta 1 resultado, conforme frequência exibida na Tabela 31.

**Tabela 31 – Resultados para o padrão {Calça Azul, Sapato Preto e Blusa Azul}**

Ordem	Padrões importantes	%Rec	%Abr
1	{CALCA AZUL 38, SAPATO PRETO 46, BLUSA AZUL 40}	75,00	31,25



O padrão em questão é bastante recorrência, porém, com abrangência de pouco mais de 30%.

#### 4.8.1.4 Cenário De Longo Prazo e padrões novos

Os padrões extraordinários são fornecidos mediante os dados recuperados do algoritmo SPADE (seção 2.7.1), conforme os critérios de seleção de algoritmos seção 4.7.

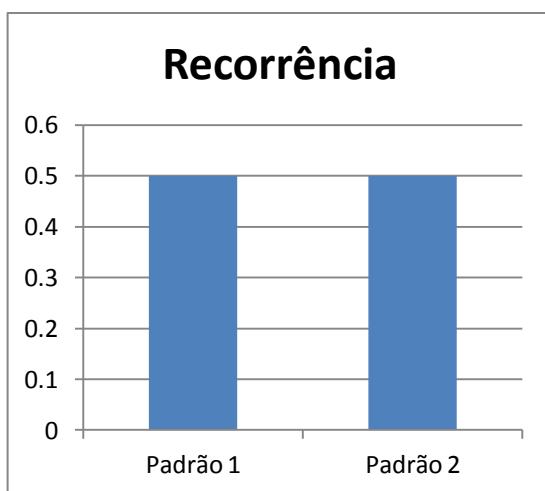
##### a) Padrão {Sapato, Calção, Bota e Gorro}

O padrão {Sapato, Calção, Bota e Gorro} apresenta dois resultados importantes com variados níveis de solicitação (Recorrência e Abrangência), conforme Tabela 32 apresenta a seguir.

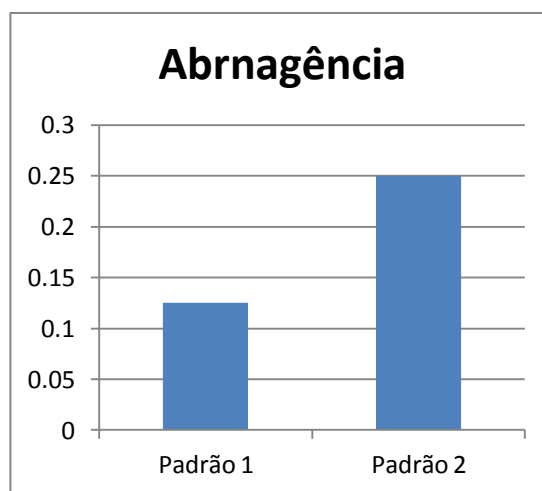
**Tabela 32 – Resultados para o padrão {Sapato, Calção, Bota e Gorro}**

Ordem	Padrões importantes	%Rec	%Abr
1	{SAPATO PRETO 38, CALÇÃO AZUL P, BOTA 43, GORRO BRANCO 59}	50,00	12,5
2	{SAPATO PRETO 38, CALÇÃO AZUL P, BOTA 40, GORRO BRANCO 59}	50,00	25,00

As Figura 31 e a Figura 32 apresentam os resultados para a recorrência e para a abrangência do padrão {Sapato Preto, Calção Azul, Bota e Gorro Branco}. De acordo com os dados de recorrência e abrangência, o padrão {SAPATO PRETO 38, CALÇÃO AZUL P, BOTA 40, GORRO BRANCO 59} é o mais solicitado.



**Figura 30 – Recorrência dos padrões da tabela 32**



**Figura 31 – Abrnagência dos padrões da tabela 32**

## 4.8.2 Abordagem de 11 anos

Esta seção apresenta todos os resultados recuperados para a abordagem de 11 anos. A forma de avaliação sofreu uma pequena mudança, apresentando como elementos de comparação a abrangência e a confiança para cada padrão, uma vez que a análise passa a ser de todo o período, sem especificar os anos, deixando a métrica recorrência inviabilizada. Para a abordagem de 11 anos, a quantidade de padrões recuperados é menor, dado o período maior de observação e os critérios aplicados.

### 4.8.2.1 Cenário de curto prazo e padrões ordinários

#### a) Padrão {Calça e Camisa}

O padrão {Calça e Camisa} apresenta um resultado importante, conforme Tabela 33 apresenta a seguir.

**Tabela 33 – Resultados para o padrão {Calça e Camisa}**

<b>Ordem</b>	<b>Padrões importantes</b>	<b>%Abr</b>	<b>%Conf</b>
1	{CALÇA AZUL 38, CAMISA CINZA 3}	100	100

O padrão {CALÇA AZUL-MESCLA 38, CAMISA CINZA 3} tem representação em todas as 16 localidades mencionadas neste trabalho.

#### b) Padrão {Sapato e Calção}

O padrão {Sapato e Calção} apresenta um resultado, conforme Tabela 34 apresenta a seguir.

**Tabela 34 – Resultados para o padrão {Camisa e Calção}**

<b>Ordem</b>	<b>Padrão Importante</b>	<b>%Abr</b>	<b>%Conf</b>
1	{SAPATO PRETO 38, CALÇÃO AZUL P}	100	89,37

O padrão {SAPATO PRETO 38, CALÇÃO AZUL P} tem representação em todas as 16 localidades mencionadas neste trabalho.

#### c) Padrão {Camisa e Sapato}

O padrão {Camisa e Sapato} apresenta um resultado, conforme Tabela 35 apresenta a seguir.

**Tabela 35 – Resultados para o padrão {Camisa e Sapato}**

<b>Ordem</b>	<b>Padrões importantes</b>	<b>%Abr</b>	<b>%Conf</b>
1	{CAMISA CINZA 3, SAPATO PRETO 38}	100	89,37

O padrão {CAMISA CINZA 3, SAPATO PRETO 38} tem representação em todas as 16 localidades mencionadas neste trabalho.

**d) Padrão {Calção Azul, Calça Azul}**

O padrão {Calção Azul, Calça Azul} apresenta um resultado, conforme Tabela 36 apresenta a seguir.

**Tabela 36 – Resultados para o padrão {Calção Azul e Calção Azul}**

<b>Ordem</b>	<b>Padrões importantes</b>	<b>%Abr</b>	<b>%Conf</b>
1	{ CALCAO AZUL P, CALCA AZUL 54 }	93.75	100

O padrão {CALCAO AZUL P, CALCA AZUL 54} tem representação em 15 localidades mencionadas neste trabalho.

**e) Padrão {Sapato, Calção e Calça}**

O padrão {Sapato, Calção e Calça} apresenta um resultado, conforme Tabela 37 apresenta a seguir.

**Tabela 37 – Resultados para o padrão {Sapato, Calção e Calça}**

<b>Ordem</b>	<b>Padrões importantes</b>	<b>%Abr</b>	<b>%Conf</b>
1	{SAPATO PRETO 38, CALCAO AZUL P, CALCA AZUL 54}	87.5	100

O padrão {SAPATO PRETO 38, CALCAO AZUL P, CALCA AZUL 54} tem representação em 14 localidades mencionadas neste trabalho.

**4.8.2.2 Cenário de curto prazo e padrões extraordinários**

**a) Padrão {Calça Camuflada e Calça}**

O padrão {Calça Camuflada e Calça} apresenta um resultado, conforme Tabela 38 apresenta a seguir.

**Tabela 38 – Resultados para o padrão {Calça Camuflada e Calça}**

<b>Ordem</b>	<b>Padrões importantes</b>	<b>%Abr</b>	<b>%Conf</b>
1	{ CALÇA CAMUFLADA 42, CALÇA AZUL 38 }	100	100

O padrão {SAPATO PRETO 38, CALCAO AZUL P, CALCA AZUL 54} tem representação em 14 localidades mencionadas neste trabalho.

### **b) Padrão {Calção e Bota}**

O padrão {Calção e Bota} apresenta um resultado, conforme Tabela 39 apresenta a seguir.

**Tabela 39 – Resultados para o padrão {Calção e Bota}**

<b>Ordem</b>	<b>Padrão Específico</b>	<b>%Abr</b>	<b>%Conf</b>
1	{ CALCAO AZUL P e BOTA 40}	100	100

O padrão {CALCAO AZUL P e BOTA 40} tem representação em 14 localidades mencionadas neste trabalho.

### **c) Padrão {Bota e Distintivo}**

O padrão {Bota e Distintivo} apresenta um resultado, conforme Tabela 40 apresenta a seguir.

**Tabela 40 – Resultados para o padrão {Bota e Distintivo}**

<b>Ordem</b>	<b>Padrão Específico</b>	<b>%Abr</b>	<b>%Conf</b>
1	{BOTA 40, DISTINTIVO GOLA}	100	100

O padrão {BOTA CONVÉS 40, DISTINTIVO GOLA} tem representação em todas as 16 localidades mencionadas neste trabalho.

### **d) Padrão {Bota e Gorro}**

O padrão {Bota e Gorro} apresenta um resultado, conforme Tabela 41 apresenta a seguir.

**Tabela 41 – Resultados para o padrão {Bota e Gorro}**

<b>Ordem</b>	<b>Padrões importantes</b>	<b>%Abr</b>	<b>%Conf</b>
1	{ BOTA 40, GORRO BEGE 57}	100	100

O padrão {BOTA DE CONVÉS 40, GORRO BEGE 57} tem representação em todas as 16 localidades mencionadas neste trabalho.

#### 4.8.2.3 Cenário de médio prazo e padrões extraordinários

##### a) Padrão {Calça e Gorro}

O padrão {Calça e Gorro} apresenta um resultado, conforme Tabela 42 apresenta a seguir.

Tabela 42 – Resultados para o padrão {Calça e Gorro}

Ordem	Padrões importantes	%Abr	%Conf
1	{CALCA AZUL 54, GORRO BRANCO 59}	81.25	50

O padrão {CALCA AZUL-MESCLA 54, GORRO BRANCO 59} tem representação em 13 localidades mencionadas neste trabalho.

#### 4.8.2.4 Cenário de longo prazo e padrões extraordinários

Os resultados para a abordagem de 11 anos e cenário de longo prazo não foram gerados, conforme os critérios de seleção de algoritmos seção 4.7. A próxima seção apresenta a previsão de demanda para as duas abordagens.

### 4.9 Previsão de demanda gerada

A previsão de demanda visa a montagem de uma lista de compras, em prioridade, como sugestão a um novo processo de compra.

#### 4.9.1 Abordagem Ano a Ano, cenário de curto Prazo e padrões ordinários

Os padrões, exibidos na Tabela 43, representam os oito padrões mais requisitados dentre os padrões ordinários.

Tabela 43 – Previsão de demanda para padrões ordinários

Ordem	Padrões ordinários
1	{SAPATO PRETO 46, BLUSA AZUL 40}
2	{CALÇÃO AZUL P, BLUSA AZUL 52}
3	{SAPATO PRETO 38, CALÇA AZUL 38}
4	{SAPATO PRETO 46, CALÇA AZUL 54}
5	{SAPATO PRETO 38, CALÇA AZUL 54}
6	{GANDOLA AZUL 3, SAPATO PRETO 38}
7	{SAPATO PRETO 38, CALÇÃO AZUL P, CALÇA AZUL 38}
8	{SAPATO PRETO 38, CALÇÃO AZUL P, CALÇA AZUL 54}

A Figura 33 exibe a distribuição encontrada nos padrões de curto prazo da Tabela 43, onde o item Sapato Preto é o mais presente.

#### 4.9.2 Abordagem Ano a Ano, cenário de curto prazo e padrões extraordinários

O padrões exibidos na Tabela 44 representam os mais requisitados dentre aqueles recuperados para os padrões extraordinários.

Tabela 44 – Previsão de demanda para padrões extraordinários

Ordem	Padrões extraordinários
1	{CALÇA AZUL 54, BOTA 40}
2	{CALÇÃ AZUL 54, BOTA 43}
3	{SAPATO PRETO 38, BOTA 40}
4	{CALÇA AZUL 54, DÓLMÃ 46}
5	{BOTA 40, CAXANGÁ 59}
6	{CALÇÃO AZUL P, BOTA 40}

A Figura 34 exibe a distribuição encontrada nos padrões de curto prazo da Tabela 44, o item Calça Azul é o mais presente.

#### 4.9.3 Abordagem Ano a Ano, cenário de médio prazo e padrões extraordinários

Há apenas um padrão recuperado para o cenário médio e padrões extraordinários, exibido na Tabela 45.

Tabela 45 – Previsão de demanda para padrões extraordinários

Ordem	Padrões extraordinários
1	{CALÇA AZUL 38, SAPATO PRETO 46, BLUSA AZUL 40}

#### 4.9.4 Abordagem Ano a Ano, cenário de longo prazo e padrões extra-ordinários

Para o cenário longo, também, há apenas um padrão recuperado, exibido na Tabela 46.

Tabela 46 – Previsão de demanda para padrões extraordinários

Ordem	Padrões extraordinários
1	{SAPATO PRETO 38, CALÇÃO AZUL P, BOTA 40, GORRO BRANCO 59}

#### 4.9.5 Abordagem de 11 anos, cenário de curto prazo e padrões ordinários

Os padrões, exibidos na Tabela 47, representam os oito padrões mais requisitados dentre os padrões ordinários.

Tabela 47 – Previsão de demanda para padrões ordinários

Ordem	Padrões ordinários
1	{CALÇA AZUL 38, CAMISA CINZA 3}
2	{SAPATO PRETO 38, CALÇÃO AZUL P}
3	{CAMISA CINZA 3, SAPATO PRETO 38}

4	{ CALCAO AZUL P, CALCA AZUL 54 }
5	{SAPATO PRETO 38, CALCAO AZUL P, CALCA AZUL 54}

A Figura 35 exibe a distribuição dos itens que constam da Tabela 47, onde os itens Calção e Sapato são os mais frequentes.

#### 4.9.6 Abordagem de 11 anos, cenário de curto prazo e padrões extraordinários

Os padrões, exibidos na Tabela 48, representam os quatro mais requisitados dentre os padrões extraordinários.

**Tabela 48 – Previsão de demanda para padrões extraordinários**

<b>Ordem</b>	<b>Padrões extraordinários</b>
1	{CALÇA CAMUFLADA 42, CALÇA AZUL 38}
2	{CALCAO AZUL P e BOTA 40}
3	{BOTA 40, DISTINTIVO GOLA}
4	{BOTA DE CONVÉS 40, GORRO BEGE 57}

A Figura 36 apresenta a distribuição dos padrões representados na Tabela 48.

#### 4.9.7 Abordagem de 11 anos, cenário de médio prazo e padrões extraordinários

A Tabela 49 representa o padrão mais requisitado, dentre os extraordinários.

**Tabela 49 – Previsão de demanda para padrões extraordinários**

<b>Ordem</b>	<b>Padrões extraordinários</b>
1	{CALCA AZUL 54, GORRO BRANCO 59}

O cenário médio mostra a relação entre Calça Azul e Gorro Branco.

#### 4.9.8 Abordagem de 11 anos, cenário de longo prazo e padrões extraordinários

Não há padrões para a geração da previsão de demanda, neste cenário. A seção seguinte verifica a validade dos padrões identificados, confrontando-os com as licitações realizadas entre os anos de 2014 e 2016. O objetivo é mostrar a redução do número de licitações realizadas, caso fossem aplicadas as técnicas aqui descritas para identificar padrões de consumo.

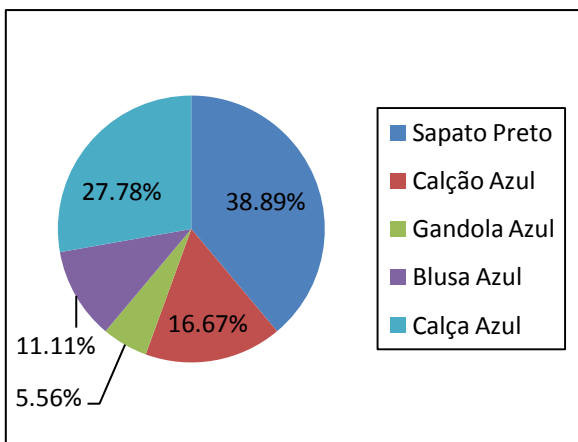


Figura 32 – Distribuição dos itens do cenário de curto prazo

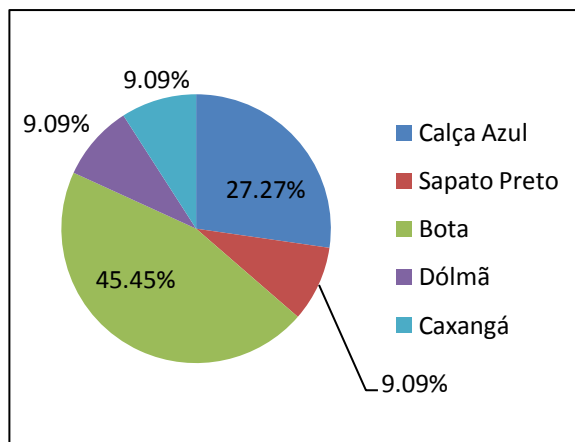


Figura 33 – Distribuição dos itens para o cenário de curto prazo

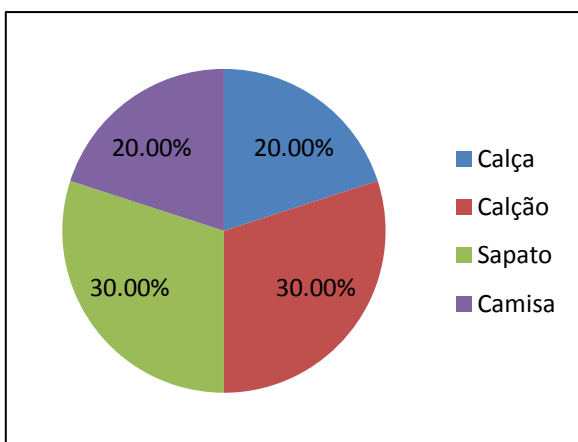


Figura 34 – Distribuição dos itens para o cenário de curto prazo, padrões ordinaries

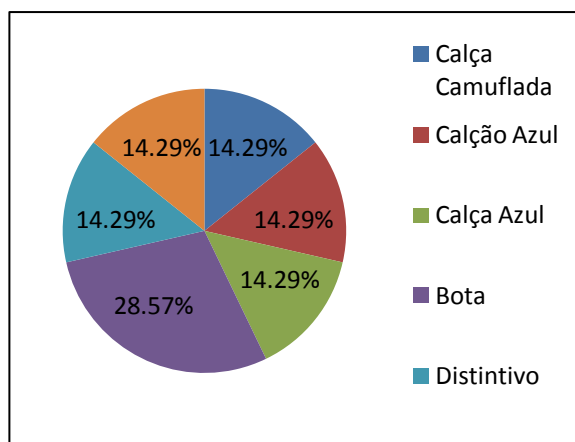


Figura 35 – Distribuição dos itens para o cenário de curto prazo, padrões extraordinários

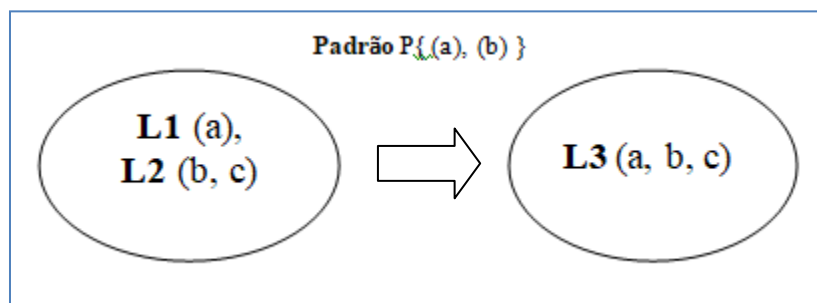
#### 4.10 Mundo real: reflexo sobre as Licitações

De posse dos padrões de consumo, é possível avaliar o fator de impacto para redução do número de licitações, realizadas, através da identificação dos padrões revelados pela mineração de dados. Esta comparação propõe-se a verificar o ganho em termos de redução do número de licitações, caso fosse adotada a compra baseada em padrões de consumo. Inicialmente, buscou-se dados para todas as licitações realizadas no período de 11 anos, porém, somente foi possível a recuperação do conjunto de licitações no período de 2014 a 2016. A quantidade de licitações, foi igual a 18 processos no ano de 2014, 12 em 2015 e 15 em 2016. Um fato importante é a forma como a organização executa suas licitações,



separando-as por categorias de itens; ou seja, há uma licitação para sapatos e outra para calças. A estrutura de apresentação dos dados seguirá a formatação adotada neste trabalho, mencionando as abordagens e os três prazos, para depois separá-los em padrões extraordinários e ordinários.

O objetivo é evidenciar o impacto no número de licitações com a utilização do planejamento de compras baseado na identificação de padrões de consumo. Para checar o impacto causado, parte-se da seguinte situação: um padrão específico P, composto dos itens (a) e (b), (notação  $P \{(a), (b)\}$ ) e licitações L1 composta do item (a) e L2 composta do itens (b) e (c). O passo seguinte é verificar quais licitações possuem o item (a) e quais possuem o item (b), lembrando que a organização pública realiza a compra de itens em separado, como mencionado no parágrafo anterior. Para a situação proposta, percebe-se que situação L1 possui o item (a) e a licitação L2 possui o item (b), além do (c), sugerindo a possibilidade de junção de L1 e L2 em uma única licitação, neste caso, L3. Dessa forma, a licitação L3 substituiria L1 e L2, contendo os itens (a), (b) e (c) e permitindo a realização de apenas um processo licitatório, podendo consumir menor tempo e, com certeza, menor quantidade de recursos financeiros. Em resumo, as licitações L1 e L2 saem de cena, surgindo a licitação, L3, conforme Figura 37.



**Figura 36 – Representação do processo de redução do número de licitações**

A Figura 37 exibe a transformação das licitações L1 e L2 na licitação L3, contendo todos os itens. Na próxima seção, são apresentados os impactos nas licitações dos anos de 2014, 2015 e 2016.

#### **4.10.1 Abordagem Ano a Ano**

Os resultados para o cenário de curto prazo e padrões extraordinários, para os anos de 2014, 2015 e 2016 são apresentados na Figura 38, na Figura 39 e na Figura 40.

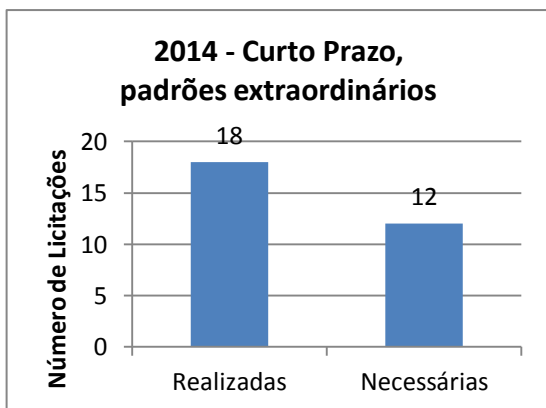


Figura 37 – Curto Prazo ano de 2014, padrões extraordinários

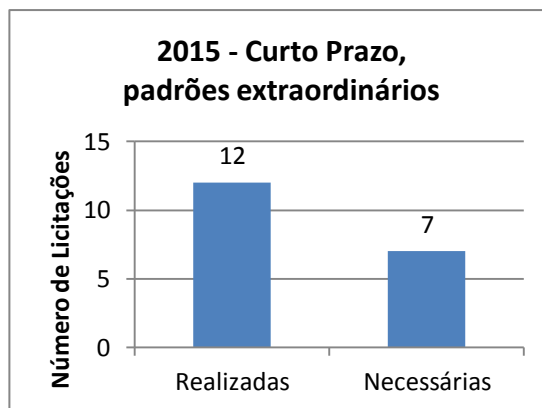


Figura 38 – Curto Prazo ano de 2015, padrões extraordinários

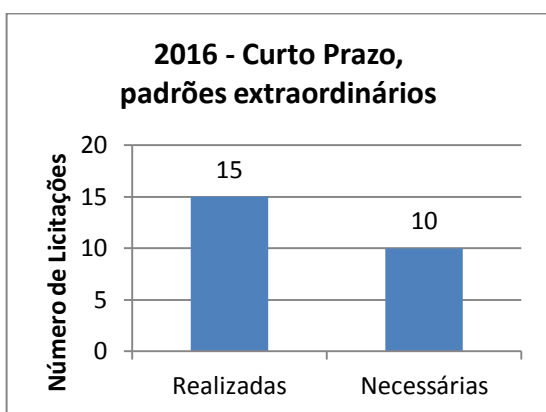


Figura 39 – Curto Prazo ano de 2016, padrões extraordinários

Realizando uma rápida avaliação, para os anos de 2014 e de 2016, há uma redução de 33,34% no número de licitações realizadas no ano. Em 2015, 41,67%, figurando como o ano de maior redução

Para o cenário curto e padrões ordinários, o impacto foi significativo, transformando as 18 licitações de 2014 em 10; as 12 de 2015 em 8; e as 15 de 2016 em 9, conforme a Figura 41, a Figura 42 e a Figura 43..

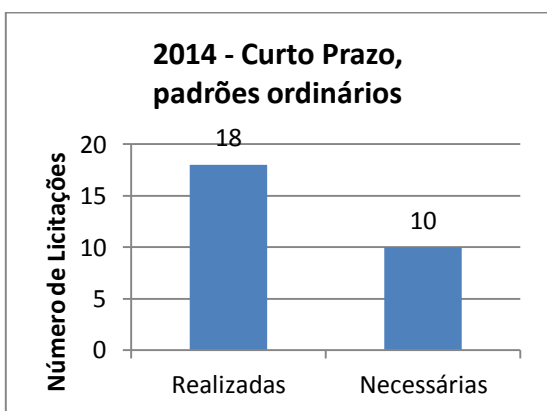


Figura 40 – Curto Prazo ano de 2014, padrões ordinários

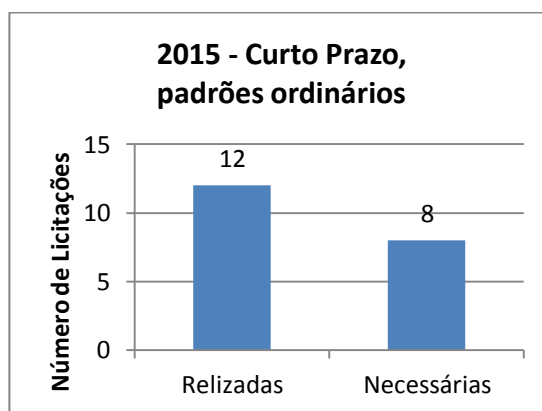
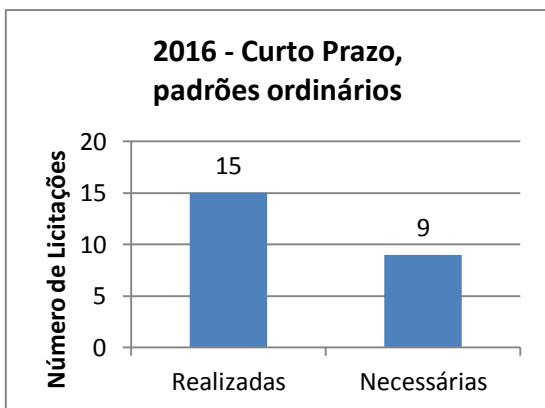


Figura 41 – Curto Prazo ano de 2015, padrões ordinários

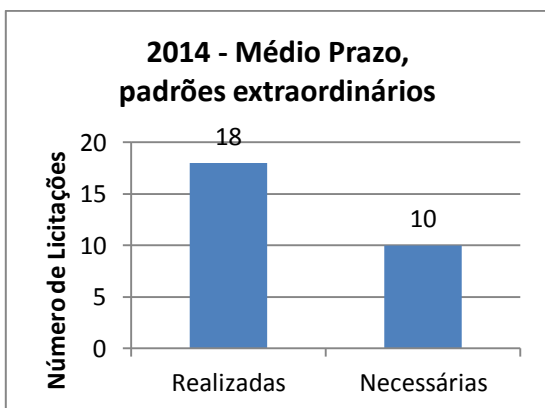
Os padrões ordinários apresentam a seguinte



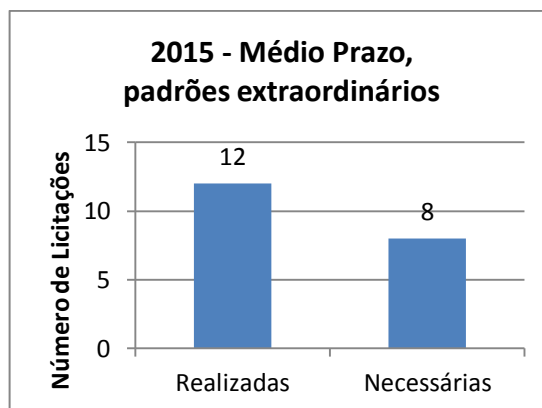
**Figura 42 – Curto Prazo ano de 2016, padrões ordinários**

redução no período: em 2014, o fator de redução é de 44,45%; em 2015, 33,34%; e em 2016, 40% de redução no número de licitações anuais.

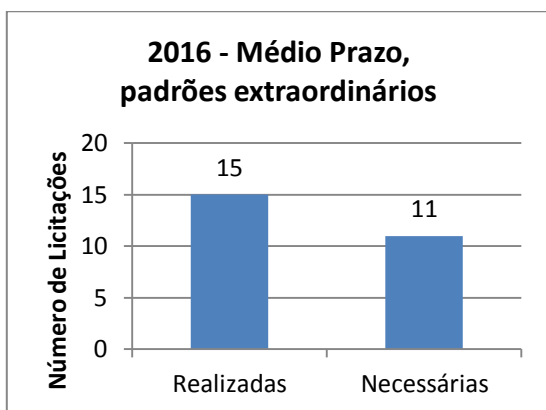
A Figura 44, a Figura 45 e a Figura 46 apresentam os resultados do médio prazo, padrões extraordinários.



**Figura 43 – Médio Prazo ano de 2014, padrões extraordinários**



**Figura 44 – Médio Prazo ano de 2015, padrões extraordinários**



**Figura 45 – Médio Prazo ano de 2016, padrões extraordinários**

Os percentuais de redução do possível número de licitações são: para o ano de 2014 de 44,45%; para 2015 de 33,34% e para o ano de 2016, a redução fica em torno de 26,67%.

A Figura 47, a Figura 48 e a Figura 49 apresentam os resultados do médio prazo, padrões ordinários.

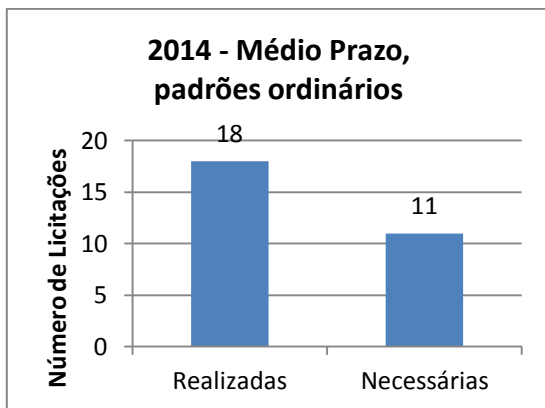


Figura 46 – Médio Prazo ano de 2014, padrões ordinários

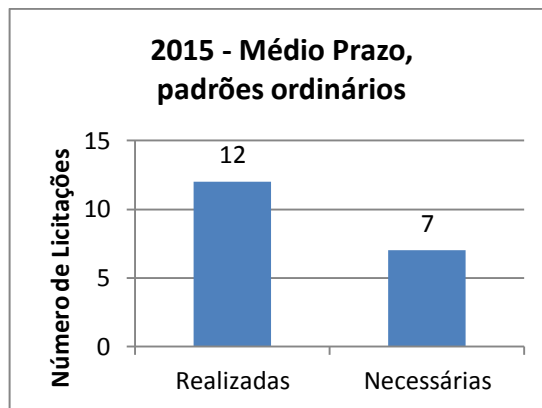


Figura 47 – Médio Prazo ano de 2015, padrões ordinários

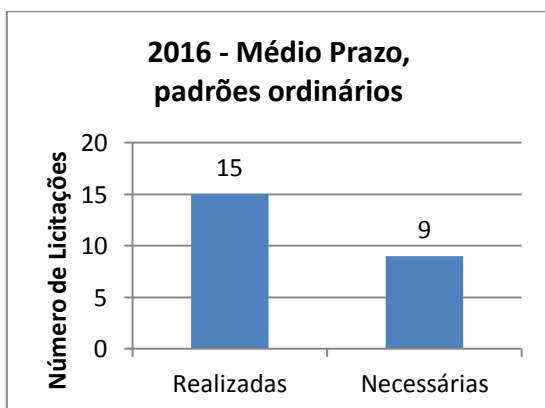


Figura 48 – Médio Prazo ano de 2016, padrões ordinários

A Figura 47, a Figura 48 e a Figura 49 apresentam os resultados para o médio prazo, para padrões ordinários, sendo que para o ano de 2014 a redução é de 38,89%; para o ano de 2015, 41,67% e para o ano de 2016, 40%.

A Figura 50, a Figura 51 e a Figura 52 apresentam os resultados para o longo prazo, padrões extraordinários.

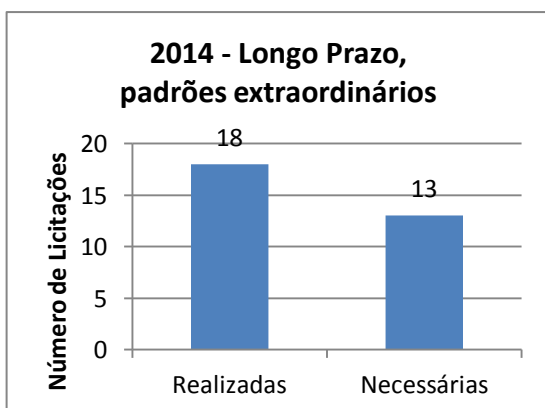


Figura 49 – Longo Prazo ano de 2014, padrões extraordinários

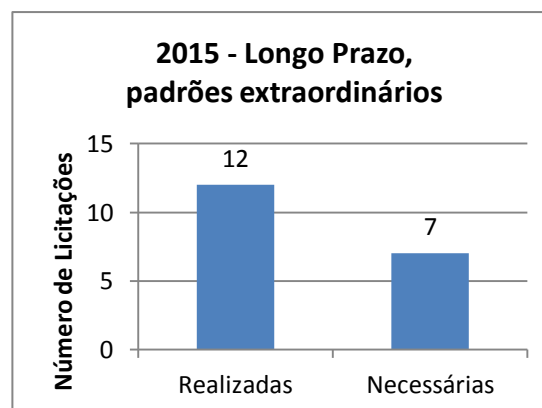


Figura 50 – Longo Prazo ano de 2015, padrões extraordinários



Figura 51 – Longo Prazo ano de 2016, padrões extraordinários

Os fatores de redução são: para o ano de 2014, 27,78%; para o ano de 2015, 41,67% e para o ano de 2016, 40%.

A Figura 53, a Figura 54 e a Figura 55 apresentam os resultados para o longo prazo, padrões ordinários.

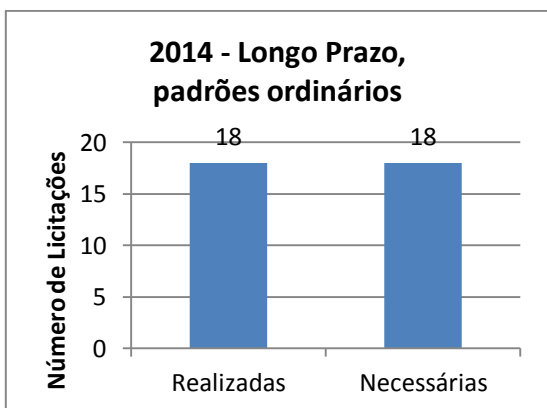


Figura 52 – Longo Prazo ano de 2014, padrões ordinários

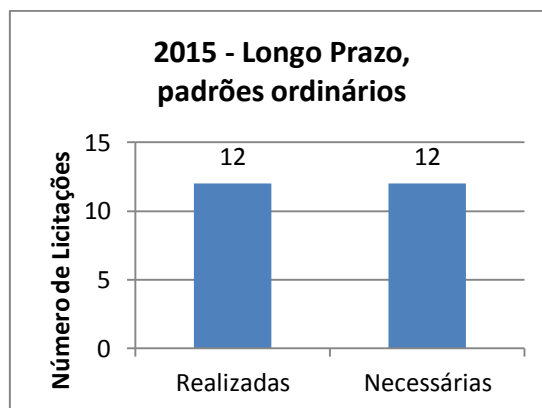


Figura 53 – Longo Prazo ano de 2015, padrões ordinários

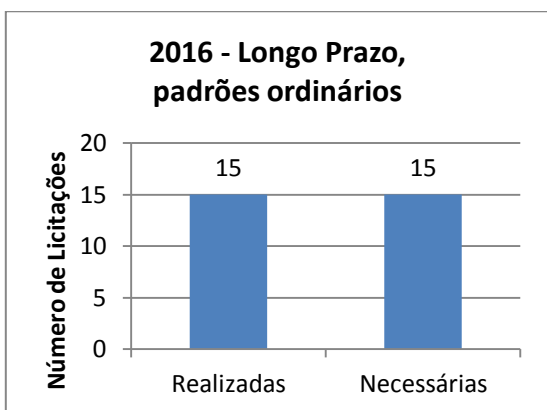


Figura 54 – Longo Prazo ano de 2016, padrões ordinários

Para o longo prazo e padrões ordinários não há percentual de redução, uma vez que não foram encontrados padrões de consumo.

Por último, dentre os prazos curto, médio e longo, o ano de 2015 apresenta o maior número de reduções do tempo de licitação, dentre os três anos.

#### 4.10.2 Abordagem de período de 11 anos

A mesma formatação para apresentação dos impactos de redução de licitação serão utilizados para a abordagem de 11 anos.



Figura 55 – Curto Prazo ano de 2014, padrões extraordinários

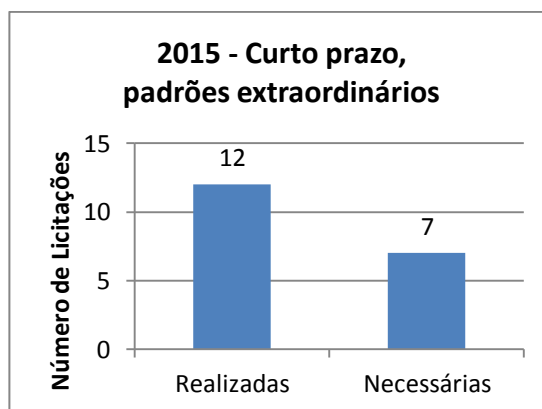


Figura 56 – Curto Prazo ano de 2015, padrões extraordinários

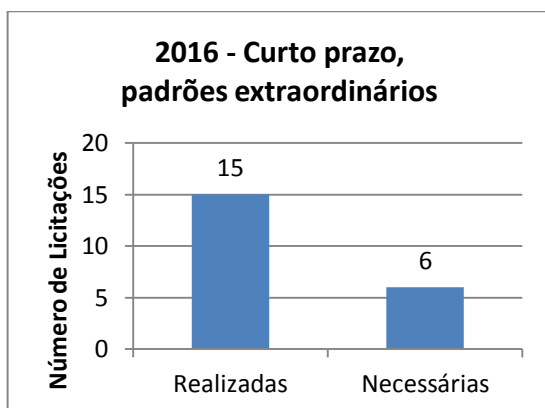


Figura 57 – Curto Prazo ano de 2016, padrões extra-ordinários

A Figura 56, a Figura 57 e a Figura 58, respectivamente, dos anos 2014, 2015 e 2016 apresentam os seguintes fatores de redução do número final de licitações a serem realizadas: 38,9%, 41,67% e 60%.

A Figura 59, a Figura 60 e a Figura 61 apresentam os resultados para o curto prazo, padrões ordinários.

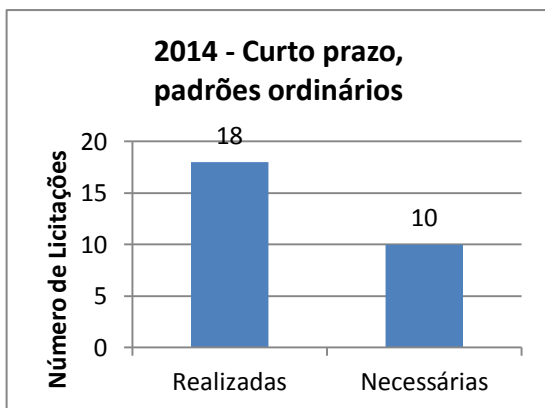


Figura 58 – Curto Prazo ano de 2014, padrões ordinários

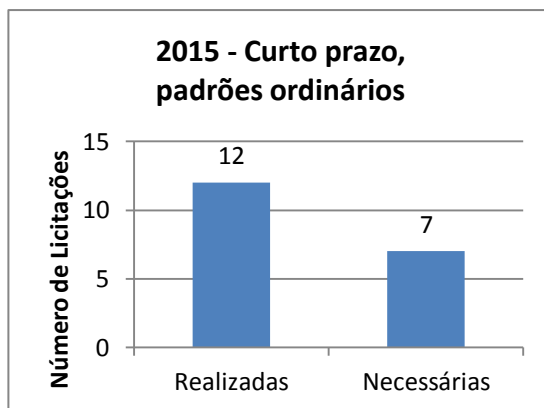


Figura 59 – Curto Prazo ano de 2015, padrões ordinários

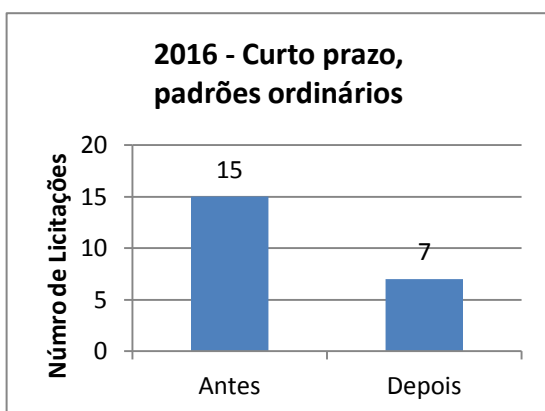


Figura 60 – Curto Prazo ano de 2016, padrões ordinários

A Figura 59, a Figura 60 e a Figura 61 relativas, respectivamente, aos anos 2014, 2015 e 2016 apresentam os seguintes fatores de redução do número final de licitações a serem realizadas: 44,45%, 41,67% e 53,34%.

A Figura 62, a Figura 63 e a Figura 64 apresentam os resultados para os padrões extraordinários de médio prazo.

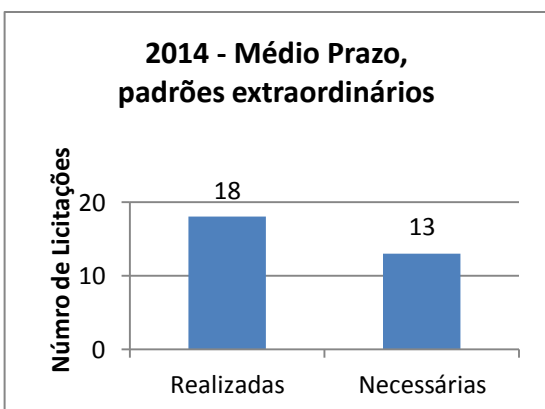


Figura 61 – Médio Prazo ano de 2014, padrões extraordinários

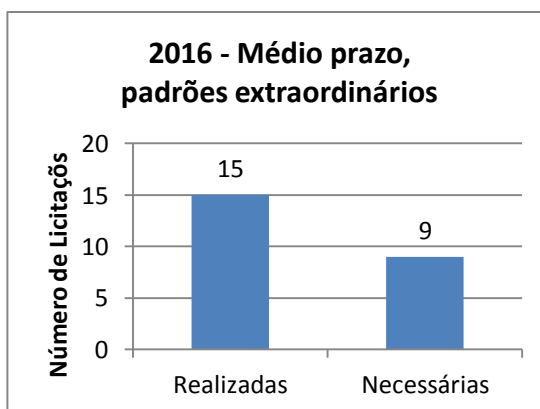
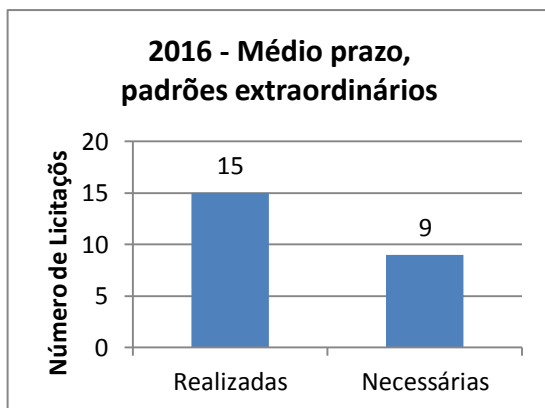


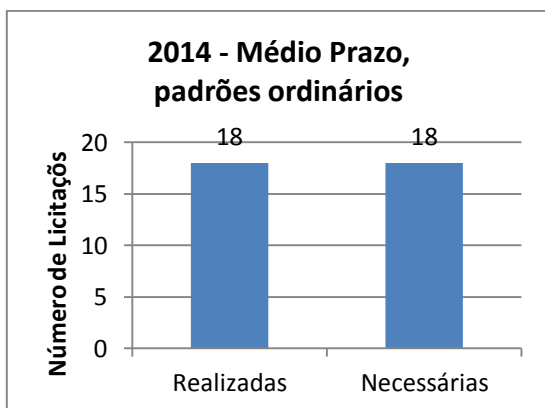
Figura 62 – Médio Prazo ano de 2015, padrões extraordinários



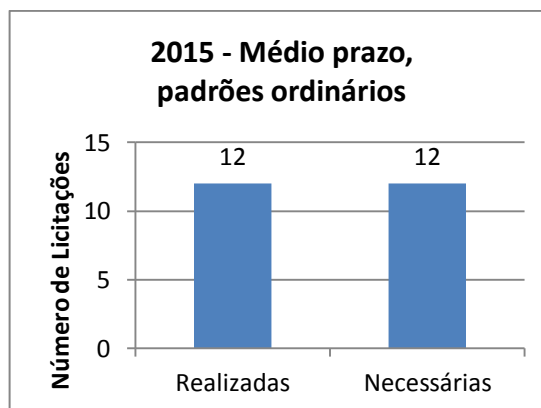
**Figura 63 – Médio Prazo ano de 2016, padrões extraordinários**

A Figura 62, a Figura 63 e Figura 64 relativas, respectivamente, aos anos 2014, 2015 e 2016 apresentam os seguintes fatores de redução do número final de licitações a serem realizadas: 27,78%, 25% e 40%.

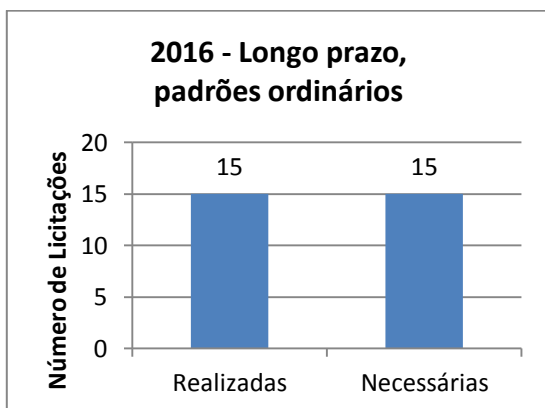
A Figura 62, a Figura 63 e a Figura 64 apresentam os resultados para os padrões extraordinários de médio prazo.



**Figura 64 – Médio Prazo ano de 2014, padrões ordinários**



**Figura 65 – Médio Prazo ano de 2015, padrões ordinários**



**Figura 66 – Médio Prazo ano de 2016, padrões ordinários**

A Figura 65, a Figura 66 e a Figura 67 relativas, respectivamente, aos anos 2014, 2015 e 2016 apresentam os seguintes fatores de redução do número final de licitações a serem realizadas: 0%. Isto se deve ao fato de não haver padrões de consumo detectados no período.



A Figura 68, a Figura 69 e a Figura 70 apresentam os resultados para os padrões ordinários de longo prazo.

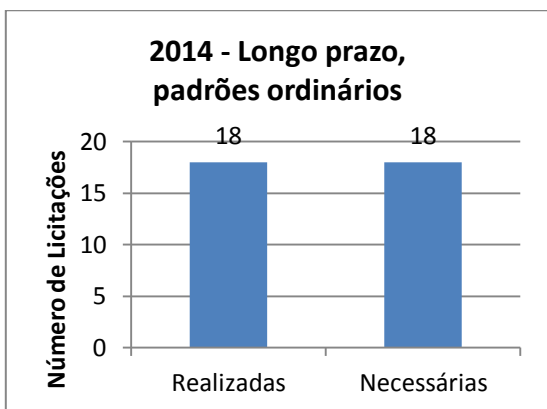


Figura 67 – Longo Prazo ano de 2014, padrões ordinários

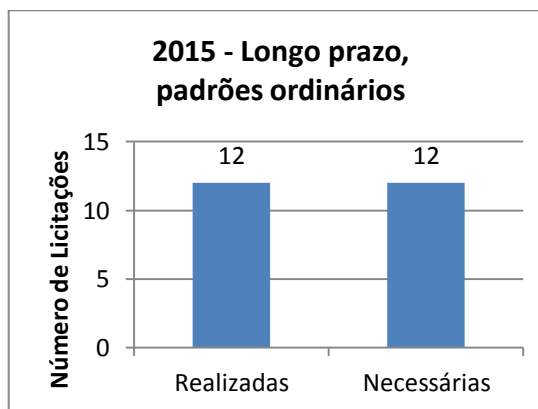


Figura 68 – Longo Prazo ano de 2015, padrões ordinários

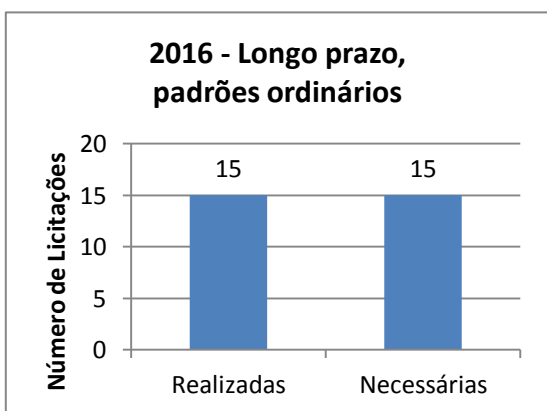


Figura 69 – Longo Prazo ano de 2016, padrões ordinários

As Figura 68, a Figura 69 e a Figura 70 relativas, respectivamente, aos anos 2014, 2015 e 2016 apresentam os seguintes fatores de redução do número final de licitações a serem realizadas: 0%. Isto se deve ao fato de não haver padrões de consumo detectados no período.

A Figura 71, a Figura 72 e a Figura 73 apresentam os resultados para os padrões extraordinários de longo prazo.

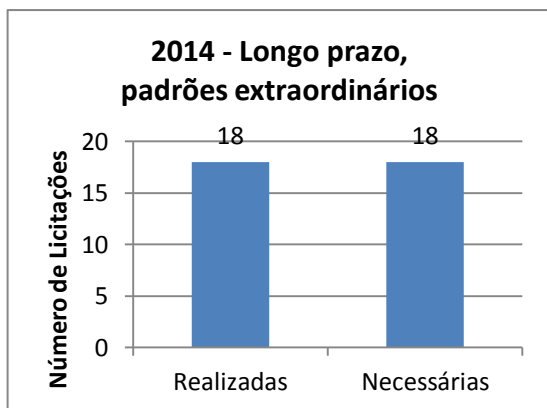


Figura 70 – Longo Prazo ano de 2014, padrões extraordinários

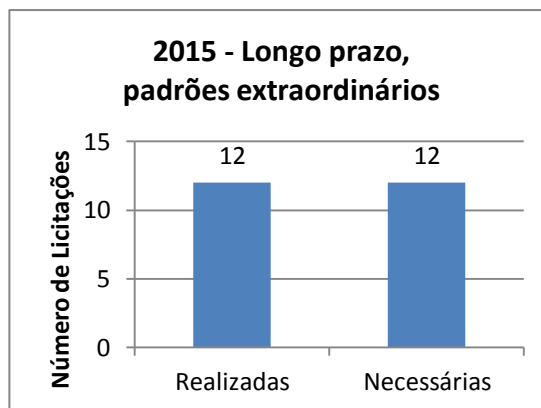


Figura 71 – Longo Prazo ano de 2015, padrões extraordinários



Figura 72 – Longo Prazo ano de 2016, padrões extraordinários

As Figura 71, a Figura 72 e a Figura 73 relativas, respectivamente, dos anos 2014, 2015 e 2016 apresentam os seguintes fatores de redução do número final de licitações a serem realizadas: 0%. Isto se deve ao fato de não haver padrões de consumo detectados no período. Por ultimo, entre os prazos curto, médio e de

longo, o ano de 2014 é o que apresenta o maior número de reduções do tempo de licitação, dentre os três cenários. Ou seja, com o emprego de técnicas de mineração de dados, surgiu a possibilidade de identificação de padrões relevantes suficientes para indicar a dinâmica de comportamento de consumo da organização, facilitando o processo de elaborar uma previsão de demanda. Em geral, para a abordagem ano a ano, do conjunto de licitações para o ano de 2014, há uma redução de 37,78% no número de licitações realizadas, com a introdução dos padrões sequenciais; para o ano de 2015, uma redução de 38,34%; e para o ano de 2016, de 36% no número de licitações realizadas no ano. Para a abordagem de 11 anos, para 2014, há 37,12% de redução; 36,11% para o ano de 2015; e 51,11% para o ano de 2016. Esses resultados podem contribuir, efetivamente, para uma melhor utilização dos recursos públicos da Marinha do Brasil.

#### **4.11 Resultados e Discussões**

O objetivo desta dissertação é propor uma alternativa à geração de previsão de demanda tradicional da Marinha do Brasil. Com este fito, alguns resultados importantes foram alcançados, possibilitando o desenvolvimento de uma solução. Dentre eles, a pesquisa e o desenvolvimento de uma taxonomia baseada na Teoria da Classificação Facetada, por meio da qual foi possível a categorização de padrões de itens de fardamento e a revelação (de forma sólida), do que foi chamado, por este trabalho, de padrões extraordinários e padrões ordinários. Do total de 48 classificações ou assuntos finais, 44 foram atribuídas à categoria Personalidade/Entidade e 4 à categoria Energia. A previsão de demanda, então, pode ser elaborada com base nos padrões recuperados, selecionando os de melhor resultado quanto à recorrência e à abrangência. Foram 16 padrões para a previsão de demanda da abordagem ano a ano e 10 padrões para a abordagem de 11 anos.

O passo seguinte foi a observação dos padrões recuperados diante de dados reais de compras por licitações realizadas pela organização pública. Para tornar a atividade viável, foram feitas buscas por licitações realizadas, considerando os itens comprados, num período estabelecido. O período fornecido foi entre os anos de 2014 e 2016, contendo todas as necessidades de materiais e itens de fardamento da organização, resultando em 18 licitações para o ano de 2014; 12 para o ano de 2015; e 15 para o ano de 2016. A seção 4.10 apresenta todos os resultados discriminados por abordagem, cenário e padrão extraordinário ou ordinário, onde é exibido o número de licitações realizadas e o número de licitações que seriam necessárias se a abordagem aqui proposta tivesse sido utilizada. Assim, na abordagem ano a ano, para o ano de 2014 há uma redução (simulando uma compra baseada nos padrões recuperados) de 37,78%, envolvendo todos os cenários; para o ano de 2015, uma redução de 38,34%; e para o ano de 2016, de 36% no número de licitações realizadas no ano. Para a abordagem de 11 anos, para um período de observação maior, tem-se 37,12% de redução para o ano de 2014; 36,11% para o ano de 2015; e 51,11% para o ano de 2016. Os resultados favoráveis mostram que a abordagem proposta nesta dissertação pode contribuir efetivamente menores gastos de dinheiro público com processos administrativos ligados à previsão de demanda, além de poupar tempo pela redução do número de licitações.

## 5 - Conclusões e Trabalhos Futuros

Esta dissertação construiu uma alternativa ao processo convencional de previsão de demanda adotado pela Marinha do Brasil, analisando comportamento de padrões de consumo de itens de fardamento, utilizando a Mineração de Padrões Sequenciais, com o objetivo de auxiliar na aquisição de novos produtos (licitações). A análise envolvendo a dimensão tempo foi fundamental para identificar os padrões de consumo da organização, de modo mais eficiente, utilizando a técnica de Mineração de Padrões Sequenciais. Os algoritmos avaliados foram o GSP, SPADE, SPAM, CMSPADE, CMSPAM, LAPIN e PREFIXPAN, desenvolvidos e fornecidos pela biblioteca SPMF.

Duas abordagens, Ano a Ano e período de 11 (onze) anos, foram propostas para estudar em maiores detalhes os padrões de consumo, assim como a construção de três cenários diferentes, porém complementares. A construção de taxonomia baseada na Teoria da Classificação Faceta, permitiu identificar e diferenciar os padrões ordinários e extraordinários, ampliando o nível de informação da organização sua dinâmica de compras; e, por último, a simulação de montagem de licitações, utilizando o conhecimento de padrões descobertos. Os algoritmos que apresentaram os melhores resultados foram: Abordagem Ano a Ano – GSP, CMSPADE e CMSPAM; Para a Abordagem de 11 anos – SPADE.

A principal contribuição da dissertação é proporcionar impacto positivo no tempo e no custo da renovação de um estoque. Os resultados obtidos mostram que, a abordagem proposta pode efetivamente reduzir o número de licitações para compra de itens.

Quanto às questões de pesquisa:

1. “Permitir um melhor agrupamento dos itens de uma licitação, guiado pelo relacionamento estabelecido pela Mineração de Dados. Os padrões encontrados pelo processo de Mineração devem permitir que itens deixem de ser comprados em separado, para serem comprados em conjunto, reduzindo custos de maneira geral”.

Na abordagem convencional, até então adotada, o agrupamento de itens segue um critério aleatório, sem considerar qualquer relacionamento entre eles; Na abordagem proposta, o agrupamento é baseado, primeiramente, no relacionamento entre os itens; segundo, na sua urgência, ou seja, no nível de solicitação; e, terceiro, na sua dispersão pelo território nacional. O conjunto de padrões recuperado permite a montagem de

processos mais inteligentes, obtendo como resultados interessantes o impacto na redução no número de licitações, com redução média de 37,78% para 2014, de 38,34% para 2015 e de 36% para 2016, na abordagem ano a ano; e de 37,12% para 2014, 36,11% para o ano de 2015; e 51,11% para o ano de 2016, abordagem de 11 anos.

2. “Permitir a ampliação do conhecimento dos padrões ocultos na organização, através da construção de categorias, baseadas em taxonomia”.

Uma taxonomia foi desenvolvida baseada na Teoria da Classificação Facetada (seção 2.9), permitindo a categorização dos padrões e a identificação da subdivisão em padrões ordinários e padrões extraordinários, ampliando o conhecimento da dinâmica de aquisição de itens da organização.

3. “Permitir a seleção dos padrões prioritários, visando à montagem de previsões de demanda com itens de maior saída ou necessidade”.

Dois métricas foram desenvolvidas para permitir a identificação dos padrões mais solicitados e/ou importantes, a Recorrência e a Abrangência, que servem de base para a geração da previsão de demanda.

4. “Identificar padrões que ocorrem dentro do espaço de tempo de 365 dias e, se possível, aqueles maiores que 365 dias. Ou seja, identificar os intervalos médios de tempo em que um padrão é observado”.

Na verdade, esta foi a razão principal do emprego da Mineração de Padrões Sequenciais: a dimensão tempo. Para realizar a análise, três cenários didáticos foram construídos para avaliar o período de 365 dias: curto prazo, médio prazo e longo prazo. Estes cenários permitem revelar alguma informação sobre o seu intervalo de solicitação, variando dos mais solicitado (curto prazo), para os menos solicitados (longo prazo). Para os valores de suporte fixados na pesquisa, não foi observada a ocorrência de padrões com intervalos maiores que 365 dias, embora haja a possibilidade de, pela variação dos valores de suporte, alguns padrões possam ser observados. A fixação do valor de suporte foi definida em razão da necessidade de estabelecer um elemento comparativos entre as duas abordagens (Ano a Ano e de 11 anos), avaliando os resultados sob um mesmo valor de suporte inicial.

5. “O conjunto de dados obtidos da Marinha do Brasil contém dados em quantidade suficiente para delinear o comportamento de consumo e, assim, representar a realidade para sugerir padrões a serem adquiridos? Ou seja, a questão diz respeito à necessidade de mais dados, além, daqueles já adquiridos”.

A pesquisa revelou uma grande quantidade de padrões para as abordagens e cenários criados, apresentando padrões relevantes, resultando, ao final, em uma significativa redução no número de licitações, sugerindo que o volume de dados obtido é suficiente para impactar positivamente a previsão de demanda da organização. A seção 4.11 apresenta a simulação dos fatores de redução no número de licitações a serem realizadas, pela introdução dos padrões de consumo.

Por fim, alguns sugestões podem ser desenvolvidas para possíveis trabalhos futuros, dentre eles:

- 1 – Uma bateria de testes foi executada para dugerir um valor de suporte mínimo, para o projeto, afixado no valor de  $10^{-2}$ , ou seja, de 0.01, rendendo uma média de dois mil resultados por ferramenta. O valor foi suficiente para a realização do trabalho inicial, com padrões de até cinco itens. Para trabalhar com padrões em quantidades maiores e contendo maior número de itens, será preciso utilizar um valor maior de suporte mínimo, além de considerável recurso computacional.

- 2 – Esta dissertação envolveu a busca de padrões sequenciais, da área de Mineração de Padrões Sequenciais, uma subárea da Mineração de Dados. Acredita-se que para a Abordagem Ano a Ano, esta abordagem tenha se constituído na escolha ideal. Porém, para a busca de outros aspectos interessantes, o emprego de outras subáreas pode agregar valor considerável, como a utilização do Agrupamento (*Clustering*) e de Redes Neurais, oferecendo uma gama de resultados mais rica.

- 3 – Esta dissertação não contém pesquisas de regras que impõem restrições do tipo: padrões começando com determinada classificação como, por exemplo: com Calçado Operativo ou com o Dólmã. Assim, tanto é possível a utilização de técnicas com imposição de restrições, como a implementação de consultas, em um ambiente específico, exibindo as mencionadas regras.

# Referências Bibliográficas

ADAMO, J.-M. *Data Mining for Association Rules and Sequential Patterns*. Springer-Verlag, New York, NY, 2001.

AGANETTE, Elisangela; ALVARENGA, Lídia; SOUZA, Renato Rocha. “Elementos constitutivos do conceito de Taxonomia”. *Informação & Sociedade*, v. 20, n. 3, 2010.

AGRAWAL, Rakesh; IMIELIŃSKI, Tomasz; SWAMI, Arun. “Mining association rules between sets of items in large databases”. In: *Acm sigmod record*. ACM, p. 207-216, 1993.

AGRAWAL, Rakesh et al. “Fast algorithms for mining association rules”. In: *Proc. 20th int. conf. very large data bases, VLDB*. p. 487-499, 1994

AGRAWAL, Rakesh; SRIKANT, Ramakrishnan. “Mining sequential patterns”. In: *Data Engineering, 1995. Proceedings of the Eleventh International Conference on*. IEEE, p. 3-14, 1995.

AYRES, Jay et al. “Sequential pattern mining using a bitmap representation”. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2002. p. 429-435.

AHLEMEYER-STUBBE, Andrea; COLEMAN, Shirley. *A practical guide to data mining for business and industry*. John Wiley & Sons, 2014.

ALENCAR, Aretha Barbosa, 2007, *Mineração e visualização de coleções de séries temporais*. Tese de D.Sc., Universidade de São Paulo, São Paulo, SP, Brasil.

BALLOU, R.H. *Logística Empresarial*. São Paulo: Atlas, 1995.

BARBOSA, Alice Príncipe. “Classificações facetadas”. *Ciência da Informação*, Rio de Janeiro, 1(2):73-81, 1972.

BARROS, A. J. S. e LEHFELD, N. A. S. *Fundamentos de Metodologia: Um Guia para a Iniciação Científica*. 2 Ed. São Paulo: Makron Books, 2000.

BERTAGLIA, Paulo Roberto. *Logística e gerenciamento da cadeia de abastecimento*. Editora Saraiva, 2000.

BOUMAN, Roland; VAN DONGEN, Jos. *Pentaho solutions: business intelligence and data warehousing with Pentaho and MySQL*. Wiley Publishing, 2009.

BOWERSOX, D.J.; CLOSS, D.J. *Logística Empresarial*. São Paulo: Atlas, 2001.

BLACKBURN, B. “*Taxonomy design types*”. *AIIIM E-doc Magazine*, Maryland, USA. v.20, n.3, p.14-16, maio/jun. 2006.

BRIN, Sergey et al. *Dynamic itemset counting and implication rules for market basket data*. In: *ACM SIGMOD Record*. ACM, 1997. p. 255-264.

BRUNO, Denise; RICHMOND, Heather. *The truth about taxonomies*. *Information Management*, v. 37, n. 2, p. 44, 2003.

CAMPOS, M. L. A. “*A organização de unidades do conhecimento em hiperdocumentos: o modelo conceitual como um espaço comunicacional para realização da autoria*”. 2001. 190f. Tese D.Sc. - CNPq/IBICT - URFJ/ECO, Rio de Janeiro, 2001.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. “*Taxonomia e classificação: o princípio de categorização*”. *DataGramZero: revista de Ciência da Informação*, v. 9, n. 4, 2008.

CALIL, Leonardo Aparecido de Almeida et al. *Mineração de dados e pós-processamento em padrões descobertos*. 2008.

CHINO, Daniel Yoshinobu Takada. “*Mineração de padrões frequentes em séries temporais para apoio à tomada de decisão em agrometeorologia*”. 2014. Tese de D.Sc. Universidade de São Paulo, São Paulo, SP, Brasil.

COMPUTERWORLD, *Exemplos de aplicações de data mining no mercado brasileiro*. Disponível em: <http://computerworld.com.br/exemplos-de-aplicacoes-de-data-mining-no-mercado-brasileiro>, acesso em 17 de novembro de 2017.

CONCEIÇÃO, Samuel V.; QUINTÃO, Ronan T. “*Avaliação do desempenho logístico da cadeia brasileira de suprimentos de refrigerantes*”. *Gestão e Produção*, v. 11, n. 3, p. 441-453, 2004.

CONWAY, Susan; SLIGAR, Char. *Unlocking knowledge assets*. Microsoft press, 2002.

CONWAY, Susan; SLIGAR, Char. *Building taxonomies*. In.: \_\_\_\_\_. *Unlocking knowledge assets*. Redmont: Microsoft Press, 2002. Cap. 6.

DA COSTA CÔRTEZ, Sérgio; PORCARO, Rosa Maria; LIFSCHITZ, Sérgio. “*Mineração de dados-funcionalidades, técnicas e abordagens*”. PUC, 2002.

DAHER, Cecílio Elias; SILVA, Edwin Pinto de La Sota; FONSECA, Adelaida Pallavicini. *Logística reversa: oportunidade para redução de custos através do gerenciamento da cadeia integrada de valor*. 2006.

DAHLBERG, I. “*Fundamentos teóricos-conceituais da classificação*”. *Revista de Biblioteconomia de Brasília*, Brasília, v.6, n.1, p.9-21, jan./jul. 1978.



DAHLBERG, I. “Teoria da classificação, ontem e hoje”. In: *CONFERÊNCIA BRASILEIRA DE CLASSIFICAÇÃO BIBLIOGRÁFICA*, 1972, Rio de Janeiro. Anais... Rio de Janeiro: IBICT/ABDF, 1979. v.1, p.352-370.

DA SILVA, Danielle Santos; MARTINS, Patrícia. “Um estudo sobre a mineração de padrões seqüenciais”.

DAVEY, Brian A.; PRIESTLEY, Hilary A. *Introduction to lattices and order*. Cambridge university press, 2002.

DAVENPORT, Thomas; HARRIS, Jeanne; MORISON, Robert. *Inteligência analítica nos negócios: como usar a análise de informações para obter resultados superiores*. 2010.

DA SILVA, Alexandre Procaci.,2004, *Regras de associação quantitativas em intervalos não contínuos*. Tese de M.Sc. Universidade Federal de Minas Gerais, Minas Gerais, MG, Brasil.

DE AMO, Sandra. “Técnicas de mineração de dados”. *Jornada de Atualização em Informatica*, 2004.

DEMO, Pedro. “Ambivalências da sociedade da informação”. *Ciência da Informação*, v. 29, n. 2, p. 37-42, 2000.

LIMA, S. H. O. et al. “Design Science: Perspectivas paradigmáticas e comparações com estudo de caso e pesquisa-ação”. *VII Encontro de Estudos Organizacionais da ANPAD. Gramado-RS: ANPAD.[GS Search]*, 2014.

DEVEZA, CECÍLIA HENRIQUES. *Minerando Padrões Sequenciais para Base de Dados de Lojas Virtuais*. Monografia (Curso de Bacharelado em Ciência da Computação), Universidade Federal de Ouro Preto (UFOP), Minas Gerais, MG, 2011.

DEVMEDIA, Artigo da SQL Magazine 26 - *Data mining e a descoberta de associações em dados*. Disponível em: <http://www.devmedia.com.br/artigo-da-sql-magazine-26-data-mining-e-a-descoberta-de-associacoes-em-dados/6155>, acessado em 17 de novembro de 2017

DI PIETRO, Maria Sylvia Zanella et al. *Direito Privado Administrativo*. São Paulo: Atlas, 2013.

FAYYAD, Usama M.; HAUSSLER, David; STOLORZ, Paul E. “KDD for Science Data Analysis: Issues and Examples”. In: *KDD*. p. 50-56. 1996a.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. “The KDD process for extracting useful knowledge from volumes of data”. *Communications of the ACM*, v. 39, n. 11, p. 27-34, 1996b.

FARIA, Evandro Rodrigues de et al. Estudo comparativo de redução de custos e tempo nas modalidades de licitação por pregão Presencial e presencial. In: **CONGRESSO USP DE CONTROLADORIA E CONTABILIDADE**. 2008.

FIGUEIREDO, Auriluci de Carvalho et al. *Probabilidade condicional: um enfoque de seu ensino-aprendizagem*. 2000.

FOURNIER-VIGER, Philippe et al. “SPMF: a Java open-source pattern mining library”. *The Journal of Machine Learning Research*, v. 15, n. 1, p. 3389-3393, 2014a.

FOURNIER-VIGER, Philippe et al. “Fast vertical mining of sequential patterns using co-occurrence information”. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer International Publishing, p. 40-52, 2014b.

FURTADO, Daniel Antônio. “Mineração de Padrões Sequenciais Múltiplos”. *Conjunto*, v. 2, n. 3, p. 56, 2005.

GLOBO, “Estudo da EMC prevê que volume de dados virtuais armazenados será seis vezes maior em 2020”. Disponível em: <<https://oglobo.globo.com/sociedade/tecnologia/estudo-da-emc-preve-que-volume-de-dados-virtuais-armazenados-sera-seis-vezes-maior-em-2020-12147682>>, acesso em 24 maio de 2017.

GRAEF, Jean. *Introduction to business taxonomies*. 2001

GOEBEL, Michael; GRUENWALD, Le. “A survey of data mining and knowledge discovery software tools”. *ACM SIGKDD explorations newsletter*, v. 1, n. 1, p. 20-33, 1999.

GORUNESCU, Florin. *Data Mining: Concepts, models and techniques*. Springer Science & Business Media, 2011.

HALL, Mark et al. “The WEKA data mining software: an update”. *ACM SIGKDD explorations newsletter*, v. 11, n. 1, p. 10-18, 2009.

HAN, Jiawei et al. “FreeSpan: frequent pattern-projected sequential pattern mining”. In: *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. p. 355-359. 2000a.

HAN, Jiawei; PEI, Jian; YIN, Yiwen. “Mining frequent patterns without candidate generation”. In: *ACM Sigmod Record*. ACM. p. 1-12. 2000b.

HAN, Jiawei et al. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth”. In: *proceedings of the 17th international conference on data engineering*. p. 215-224. 2001.

HANLEY, Susan. “Everything you ever wanted to know about taxonomies... but were afraid to ask”. *Susan Hanley LLC*, disponível em: <<http://cloud.snappages.com>>

com/b8898dc2c08e137d03449de65b9e82e108c15658/taxonomies.pdf> acesso em: 14 de outubro de 2011), 2005.

HANKE, John E. et al. *Estadística para negocios*. IRWIN,, 1995.

HIPP, Jochen; GÜNTZER, Ulrich; NAKHAEIZADEH, Gholamreza. “Algorithms for association rule mining—a general survey and comparison”. *ACM sigkdd explorations newsletter*, v. 2, n. 1, p. 58-64, 2000.

YANG, Zhenglu; WANG, Yitong; KITSUREGAWA, Masaru. “LAPIN: Effective sequential pattern mining algorithms by last position induction”. 2005.

LAKATOS, E. M.; MARCONI, M. de A. *Fundamentos de metodologia científica*. 6. ed. 5. reimp. São Paulo: Atlas, 2007.

LANGRIDGE, Derek. *Classificação: abordagem para estudantes de biblioteconomia*. Interciência, 1977.

LU, Yijun et al. “Concept hierarchy in data mining: specification, generation and implementation”. *Simon Fraser University*, 1998.

LOGISTICA DESCMPPLICADA , Disponível em: <http://www.logisticadescomplicada.com/gestao-da-cadeia-de-suprimentos-%E2%80%93-93-co-nceitos-tendencias-e-ideias-para-melhoria/>, acesso em 23 de maio de 2017.

MAYER-SCHONBERGER, Viktor; CUKIER, Kenneth. *Big data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana*. Elsevier Brasil, 2014.

MANNILA, Heikki; TOIVONEN, Hannu; VERKAMO, A. Inkeri. “Discovering frequent episodes in sequences extended abstract”t. In: *1st Conference on Knowledge Discovery and Data Mining*. 1995.

MARCONI, M. A. & LAKATOS, E. M. *Metodologia científica*. São Paulo: Atlas, 2000.

MARTINEZ, A. et al. “Las categorias o facetas fundamentales: una metodologia para el diseño de taxonomias corporativas de sitios Web argentinos”. *Ciência da Informação*, v.33, n.2, p.106-111, 2004.

MEIRELLES, Hely Lopes. *Direito administrativo brasileiro*. atualizada por Eurico de Andrade Azevedo, Délcio Balestero Aleixo e José Emmanuel Burle Filho. São Paulo: Malheiros, p. 08-2005, 2001.

MENZIES, Tim; HU, Ying. “Data mining for very busy people”. *Computer*, v. 36, n. 11, p. 22-29, 2003.

MORETTIN, Pedro Alberto; BUSSAB, Wilton O. *Estatística básica*. Editora Saraiva, 2000.

MUNDO EDUCAÇÃO. Variância e Desvio Padrão. Disponível em: <<http://mundoeducacao.bol.uol.com.br/matematica/variancia-desvio-padrao.htm>>, acesso em 26 de maio de 2017

NAVARRO, LEILA – Disponível em: <<http://www.leilanavarro.com.br/artigos/6>>, acesso em 24 maio de 2017.

NETO, Edson Serejo. *Organização do conhecimento em ambientes web com base na teoria da classificação facetada*. 2014, Dissertação, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.

NOVO, Hildenise Ferreira. “A elaboração de taxonomia: princípios classificatórios para domínios interdisciplinares”. 2007. 172 f. Dissertação (Mestre em Ciência da Informação) – Universidade Federal Fluminense, Instituto de Arte e Comunicação Social, Niterói, Rio de Janeiro, 2007.

PIATETSKY-SHAPIRO, Gregory. “Discovery, analysis and presentation of strong rules”. *Knowledge discovery in databases*, p. 229-248, 1991.

PROBEST. Probabilidade e Estatística UFRJ. <<http://www.im.ufrj.br/probest/>>, acessado em 29 de maio de 2017.

PRODANOV, Cleber Cristiano; DE FREITAS, Ernani Cesar. *Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico-2ª Edição*. Editora Feevale, 2013.

RANGANATHAN, Shiyali Ramamrita. *Prolegomena to library classification*. Asia Publishing House (New York), 1967.

RANGANATHAN, Shiyali Ramamrita et al. *Colon classification*. Sarada Ranganathan endowment for library Science, 1989.

ROSINI, Alessandro Marco et al. “MOOCS. PARA ONDE CAMINHA O PROCESSO DE APRENDIZAGEM E O USO DE RECURSOS INFORMÁTICOS”. São Paulo, RJ, Brasil, 2014.

SILVA, Marcelino Pereira dos Santos. “Mineração de dados: Conceitos, aplicações e experimentos com weka”. *Sociedade Brasileira de Computação*, v. 1, 2004.

SILVA, Márcio Bezerra da et al. *A teoria da classificação facetada na modelagem de dados em banco de dados computacionais*. 2011.

SRIKANT, Ramakrishnan; AGRAWAL, Rakesh. “Mining sequential patterns: Generalizations and performance improvements”. In: *International Conference on Extending Database Technology*. Springer Berlin Heidelberg, 1996. p. 1-17.

STANK, Theodore P.; KELLER, Scott B.; DAUGHERTY, Patricia J. “Supply chain collaboration and logistical service performance”. *Journal of Business logistics*, v. 22, n. 1, p. 29-48, 2001.

TERRA, José Cláudio Cyrineu. “Gestão do conhecimento: o grande desafio empresarial”. 2005.

TRUJILLO FERRARI, A. *Metodologia da ciência*. 3. ed. Rio de Janeiro: Kennedy, 1974.

ZAKI, Mohammed J. “Sequence mining in categorical domains: Algorithms and applications”. In: *Sequence Learning*. Springer Berlin Heidelberg, 2000. p. 162-187.

ZAKI, Mohammed J. “SPADE: An efficient algorithm for mining frequent sequences”. *Machine learning*, v. 42, n. 1-2, p. 31-60, 2001.

ZELENY, Milan. “Management support systems: towards integrated knowledge management”. *Human systems management*, v. 7, n. 1, p. 59-70, 1987.

VIANA, Reinaldo. “Mineração de dados: Introdução e Aplicações”. *SQL Magazine*, Ed. 10, Ano, v. 1, 2004.

SIMPLE DATA MINING. Disponível em: <<http://simpledatamining.blogspot.com.br/2015/03/generalized-sequential-pattern-gsp.html>>, acesso em 17 de novembro de 2017.

KNIME. Disponível em: <<https://www.knime.com/>>, acesso em 17 de novembro de 2017.

VAN HOEK, R. I. “Measuring the unmeasurable – measuring and improving performance in the supply chain”. *SupplyChain Management*, Bradford, v. 3, n. 4, p. 187-192, 1998.

VASCONCELOS, Fernanda. “Licitação pública: análise dos aspectos relevantes do Pregão”. *Prima Facie-Direito, História e Política*, v. 4, n. 7, p. 151-163, 2005.

WANG, Jason TL et al. “Introduction to data mining in bioinformatics”. In: *Data Mining in Bioinformatics*. Springer London, 2005. p. 3-8.