



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

# Analise heurística do padrão de votação dos deputados em proposições na Câmara dos Deputados entre 2015 e 2016

Laura Solano de Carvalho

Monografia apresentada como requisito parcial  
para conclusão do Curso de Computação — Licenciatura

Orientador  
Prof. Dr. Jan Mendonça Correa

Brasília  
2016

Universidade de Brasília — UnB  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Curso de Computação — Licenciatura

Coordenador: Prof. Dr. Pedro Antônio Dourado de Rezende

Banca examinadora composta por:

Prof. Dr. Jan Mendonça Correa (Orientador) — CIC/UnB  
Prof. Dr. Wilson Henrique Veneziano — CIC/UnB  
Prof.<sup>a</sup> Dr. Pedro Antônio Dourado Rezende — CIC/UnB

### **CIP — Catalogação Internacional na Publicação**

de Carvalho, Laura Solano.

Análise heurística do padrão de votação dos deputados em proposições na Câmara dos Deputados entre 2015 e 2016 / Laura Solano de Carvalho. Brasília : UnB, 2016.

215 p. : il. ; 29,5 cm.

Monografia (Graduação) — Universidade de Brasília, Brasília, 2016.

1. *Business Intelligence*, 2. Mineração de Dados, 3. LAI,  
4. Informação, 5. Dados, 6. Câmara dos Deputados, 7. Votações,  
8. Proposições

CDU 004.4

Endereço: Universidade de Brasília  
Campus Universitário Darcy Ribeiro — Asa Norte  
CEP 70910-900  
Brasília-DF — Brasil



# Dedicatória

*Dedico este trabalho a todos os que sempre me ampararam, em especial ao meu pai Jorge Carvalho e à minha mãe Marina Solano. Ao meu pai por me ensinar a persistir e à minha mãe por investir em mim e, acima de tudo, acreditar no meu potencial.*

# Agradecimentos

*Agradeço primeiramente a Deus, pelo dom da vida, da fé e por todas as oportunidades de crescimento que Ele me proporcionou. Agradeço aos meus pais e amigos Jorge e Marina por todo apoio e investimento debruçados a mim, pela paciência, dedicação, por serem minha inspiração e, sobretudo, pelo grande amor que empreenderam ao longo de toda a minha jornada. Agradeço à minha irmã Luíza por nunca me deixar desistir de meus sonhos, me dando forças para alcançar tudo que sempre almejei. Agradeço à todos os meus amigos e colegas de trabalho da Biggdata por viabilizarem essa tão almejada formatura. Em especial, ao Alencar pelo incentivo, tutoria, força e suporte dados a mim durante todo esse processo de desenvolvimento do trabalho final. À Denise pelos anos de parceria e amizade verdadeira, pela paciência, pelo apoio e pela disposição em ajudar a tonar o meu trabalho, o melhor possível. Ao Luciano pela total disponibilidade, paciência e por todo apoio e suporte em todos os momentos em que pensei desistir. Agradeço aos professores do Departamento de Ciência da Computação, em especial ao meu orientador Jan Mendonça Correa, pelo suporte prestado durante o desenvolvimento deste trabalho. Sem vocês, nada disso teria sido possível.*

# Resumo

Com o advento da Lei de Acesso à Informação (nº 12.527) promulgada em 2012, foi regulamentado o direito constitucional de acesso às informações públicas. Tal lei se propõe a divulgar informações com o intuito de criar mecanismos que possibilitem, a qualquer pessoa, física ou jurídica, sem necessidade de apresentar motivo, o recebimento de informações públicas dos órgãos e entidades. Os órgãos públicos disponibilizam dados, muitos deles de difícil acesso, extração, compreensão e com pouca qualidade. Desta forma, por meio de processos de mineração de dados e por análises heurísticas nos dados, este trabalho visa proporcionar uma análise dos dados de atuação dos deputados eleitos para a Câmara dos Deputados em sessões de votação de proposições em plenário nos anos de 2015 e 2016.

**Palavras-chave:** *Business Intelligence*, Mineração de Dados, LAI, Informação, Dados, Câmara dos Deputados, Votações, Proposições

# Abstract

With the advent of the Access to Information Act (No. 12.527) enacted in 2012, it has regulated the constitutional right of access to public information. This law aims to disseminate information in order to create mechanisms that allow to any person or entity, without the need for reason, the receipt of public information organs and entities. Public agencies providing data, many of them difficult to access, extraction, understanding and poor quality. Thus, through data mining processes and heuristics analysis on the data, this paper aims to provide an analysis of the performance data of the elected deputies to the House of Representatives in voting sessions of proposals in plenary in the years 2015 and 2016.

**Keywords:** *Business Intelligence, Data Mining, LAI, Information, Data, Chamber of Deputies, Voting, Proposals*

# Lista de Figuras

2.1	Funções do Poder Legislativo [14, p.26] . . . . .	5
2.2	Fluxograma constitucional da tramitação legislativa [14, p.26] . . . . .	10
4.1	Componentes de um ambiente de BI [7] . . . . .	22
4.2	Componentes chave que promovem uma solução de BI ágil [46, p.115] . . . . .	25
4.3	'Quadrante Mágico do Gartner' para BI e Plataformas de Análise - Fevereiro de 2016 [34] . . . . .	28
4.4	Diferenciais do AQL [22] . . . . .	29
4.5	Fases do Processo de Extração de Conhecimento (KDD) [6, apud [27]] . . . . .	31
4.6	Tarefas da Análise Prévia e Descobrimto [18] . . . . .	35
4.7	Captura de tela inicial do Weka [70] . . . . .	37
4.8	Captura de tela da interface <i>Explorer</i> do Weka [70] . . . . .	38
4.9	Captura de tela da interface <i>KnowledgeFlow</i> do Weka [70] . . . . .	39
4.10	Captura de tela da interface <i>Experimenter</i> do Weka [70] . . . . .	40
5.1	Fluxograma da Metodologia de Pesquisa . . . . .	41
5.2	Fluxograma do planejamento de estudo da Pesquisa Descritiva . . . . .	44
5.3	Modelo de dados associativo gerado pelo QlikView . . . . .	48
5.4	Fluxograma do planejamento de estudo da Pesquisa Exploratória . . . . .	51
6.1	Distribuição das proposições por tipo de proposição . . . . .	55
6.2	Evolução dos votos por proposição . . . . .	55
6.3	Distribuição dos votos . . . . .	57
6.4	Os 10 parlamentares com maior presença em votações . . . . .	57
6.5	Os 10 parlamentares com maior número de abstenções . . . . .	58
6.6	Os 10 partidos com maior número de abstenções . . . . .	59
6.7	Ranking de deputados fiéis à orientação do partido . . . . .	61
6.8	Ranking de bancadas partidárias fiéis à orientação do partido . . . . .	63
6.9	Ranking de UFs com parlamentares fiéis à orientação do partido . . . . .	64
6.10	Os 10 Deputados em Concordância com a Maioria . . . . .	65
6.11	Os 10 Partidos em Concordância com a Maioria . . . . .	66
6.12	Evolução dos votos em apoio ao PT e ao PMDB . . . . .	67
6.13	TELA1 . . . . .	69
6.14	TELA2 . . . . .	70
7.1	Fluxograma da Metodologia de Pesquisa . . . . .	75



# Lista de Tabelas

4.1	Desvantagens do BI [46, adaptado]	24
6.1	Lista de partidos presentes nas votações	56
6.2	Distribuição das orientações por partido	60
6.3	Resultado do algoritmo <i>Simple K-means</i>	72
6.4	Distribuição de partidos por base aliada	73
6.5	Os 5 partidos com maior consistência entre os parlamentares da bancada	74
6.6	Os 5 partidos com maior facilidade de serem identificados através dos votos de seus parlamentares	74
III.1	DE-PARA Blocos Partidários	91
V.1	Dicionário de dados da tabela Resumo das Votações	97
V.2	Dicionário de dados da tabela Votações	98
V.3	Dicionário de dados da tabela Orientação	98
V.4	Dicionário de dados da tabela Orientação Partido	99

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação	2
1.2	Objetivo Geral	2
1.3	Objetivos Específicos	2
1.4	Metodologia	2
1.5	Descrição dos Capítulos	3
<b>2</b>	<b>Formalização do Poder no Brasil</b>	<b>4</b>
2.1	A tripartição de Poderes no Brasil	4
2.2	O Poder Legislativo	5
2.2.1	Câmara dos Deputados	6
2.2.2	Senado Federal	7
2.3	Processo Legislativo	8
2.3.1	Proposições Legislativas	10
2.3.1.1	Projeto de Lei Ordinária	11
2.3.1.2	Proposta de Emenda à Constituição	12
2.3.1.3	Medida Provisória	13
<b>3</b>	<b>O Papel da Informação para a Sociedade</b>	<b>15</b>
3.1	Dado, Informação e Conhecimento	15
3.2	Sociedade da Informação	16
3.3	Lei de Acesso à Informação	17
<b>4</b>	<b>Análise de Dados</b>	<b>20</b>
4.1	<i>Business Intelligence</i> (BI)	20
4.1.1	Componentes do BI	21
4.1.1.1	Dados Operacionais	22
4.1.1.2	<i>Operational Data Store</i>	22
4.1.1.3	Extração, Transformação e Carga	22
4.1.1.4	<i>Data Warehouse</i> e <i>Data Mart</i>	23
4.1.1.5	Mineração de Dados	23
4.1.1.6	Visualização dos Resultados	24
4.1.1.7	Desvantagens do BI	24
4.1.2	BI Ágil	24
4.1.3	Ferramentas de BI	27
4.1.3.1	QlikView	28
4.2	Mineração de Dados	30

4.2.1	Tarefas e Métodos de Mineração de Dados . . . . .	32
4.2.1.1	Tarefas de Mineração de Dados . . . . .	32
4.2.1.2	Técnicas (ou Métodos) de Mineração de Dados . . . . .	34
4.2.2	Ferramentas para Mineração de Dados . . . . .	35
4.2.3	<i>Software</i> Weka . . . . .	37
<b>5</b>	<b>Metodologia</b>	<b>41</b>
5.1	Visão Geral . . . . .	41
5.1.1	Pesquisa Descritiva . . . . .	41
5.1.1.1	Objetivo . . . . .	41
5.1.1.2	Definição do Escopo de Pesquisa . . . . .	42
5.1.1.3	Planejamento do Estudo . . . . .	43
5.1.2	Pesquisa Exploratória . . . . .	50
5.1.2.1	Objetivo . . . . .	50
5.1.2.2	Hipóteses . . . . .	50
5.1.2.3	Definição do Escopo de Pesquisa . . . . .	50
5.1.2.4	Planejamento do Estudo . . . . .	51
<b>6</b>	<b>Resultados</b>	<b>54</b>
6.1	Pesquisa Descritiva . . . . .	54
6.1.1	Visão Geral . . . . .	54
6.1.2	Análise dos Votos sob a Orientação dos Partidos . . . . .	59
6.1.3	Análise dos Votos em Comparação com a Maioria . . . . .	64
6.1.4	Análise dos Votos em Apoio às Maiores Bancadas . . . . .	67
6.1.5	Painéis de Análise . . . . .	68
6.2	Pesquisa Exploratória . . . . .	71
6.2.1	Análise dos votos dos parlamentares . . . . .	71
<b>7</b>	<b>Considerações Finais</b>	<b>75</b>
7.1	Resultados . . . . .	76
7.2	Limitações e Trabalhos Futuros . . . . .	76
	<b>Referências</b>	<b>78</b>
<b>I</b>	<b><i>Script</i> de busca dos dados de proposição</b>	<b>84</b>
<b>II</b>	<b>Script iterativo para busca dos dados no <i>Web Service</i> da Câmara dos Deputados</b>	<b>87</b>
<b>III</b>	<b>Tabela DE-PARA de tratamento de Blocos Partidários</b>	<b>91</b>
<b>IV</b>	<b><i>Script</i> de transformação de dados</b>	<b>92</b>
<b>V</b>	<b>Dicionário de Dados</b>	<b>97</b>

# Capítulo 1

## Introdução

Todo cidadão tem direito ao acesso à informação, sobretudo a informação sob a guarda do Estado que é sempre pública, devendo o acesso a ela ser restringido apenas em casos específicos. Segundo a Controladoria Geral da União (CGU), "O acesso a esses dados constitui-se em um dos fundamentos para a consolidação da democracia, ao fortalecer a capacidade dos indivíduos de participar de modo efetivo da tomada de decisões que os afeta"[16].

No Brasil, o acesso à informação pública está inscrito no capítulo I - dos Direitos e Deveres Individuais e Coletivos -, artigo 5º, § 33º da Constituição Federal (CF) de 1988 [9]. O texto constitucional diz que "todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado". Apesar dessas garantias constitucionais, não havia no Brasil uma lei unificada sobre o acesso à informação e à transparência e que pudesse ser aplicada a todos os órgãos da administração pública. Em 18 de novembro de 2011, foi promulgada a Lei nº 12.527, que entrou em vigor em 16 de maio de 2012. Denominada Lei de Acesso à Informação (LAI), essa norma vem ampliar a transparência das atividades nos três Poderes da União, Estados, Distrito Federal e Municípios, inclusive aos Tribunais de Conta e Ministério Público.

Corroborando com esta visão, o conceito de transparência ativa ganha importância no cenário brasileiro, uma vez que preza pela divulgação de dados por iniciativa do próprio setor público, tornando públicas informações, sem um requerimento direto do cidadão, utilizando principalmente a Internet para tal.

Neste contexto, observa-se um grande aumento na disponibilização de dados abertos governamentais que, segundo as normas da *World Wide Web Consortium* (W3C), devem ser "compartilhados em formato bruto e aberto, compreensíveis logicamente, de modo a permitir sua reutilização em aplicações digitais desenvolvidas pela sociedade"[71]. No entanto, esses dados precisam ser estudados e tratados para que virem informação útil para a sociedade.

## 1.1 Motivação

Ainda nos dias de hoje, mesmo com a existência da LAI, é comum que a população não tenha o acesso às informações a respeito da atuação dos parlamentares federais em votações de proposições. Dentro do que é disponibilizado pela Câmara dos Deputados, encontra-se um amontoado de dados sem tratamento e de difícil extração e interpretação.

Desta forma, a lei que deveria garantir o acesso à informação, garante apenas que a população tenha acesso a dados brutos. Sendo assim, encontra-se como motivação a análise heurística de informações a respeito da atuação dos parlamentares eleitos para a Câmara dos Deputados, após a coleta dos dados brutos e a transformação dos mesmos.

## 1.2 Objetivo Geral

O presente trabalho tem como objetivo geral a análise heurística dos resultados obtidos através de atividades e procedimentos feitos em cima de dados de votações de proposições em plenário da Câmara dos Deputados ocorridos em 2015 e 2016, realizados com ferramentas QlikView e Weka.

## 1.3 Objetivos Específicos

Os objetivos específicos são:

- Analisar os dados gerais de votações de proposições feitas em plenário no período de 2015 e 2016;
- Identificar partidos que possuem influência sobre os outros partidos;
- Identificar parlamentares fiéis aos partidos;
- Descobrir quão significativa é a influência das bancadas estaduais nas votações;

## 1.4 Metodologia

A metodologia adotada para o desenvolvimento do presente Trabalho de Conclusão de Curso consistiu, basicamente, em quatro macroetapas conforme discriminado a seguir.

A primeira etapa consistiu na definição do escopo de pesquisa do trabalho, por meio da identificação dos pontos mais relevantes, de forma a permitir que os resultados finais sejam congruentes com os objetivos levantados.

A segunda etapa constitui-se do Referencial Teórico do trabalho, que aborda assuntos relativos ao escopo definido, sendo eles: a Formalização do Poder no Brasil, o Papel da Informação para a Sociedade e a Análise de Dados. O primeiro assunto aborda tópicos que vão desde a definição de poder, passando pela divisão de poderes no Brasil com enfoque no Poder Legislativo, finalizando em uma abordagem sobre as proposições legislativas,

que serão utilizadas como fonte de dados para esse trabalho. No âmbito do Papel da Informação para a Sociedade, foram abordados temas como sociedade da informação e Lei de Acesso à Informação, incluindo definições de termos importantes para o presente trabalho. Refente à Análise de Dados, têm-se um foco maior na abordagem de análise de dados por meio do *Business Intelligence* (BI) e do *Data Mining*.

A terceira etapa foi composta por uma pesquisa descritiva baseada nos dados extraídos. Esta etapa tem como objetivo descrever as características dos dados e proporcionar uma nova visão sobre realidade de dados já existente. As atividades da presente etapa são feitas com a ferramenta de BI *in-memory*, QlikView.

A quarta - e última etapa - consistiu em uma pesquisa exploratória visando a descoberta e a elucidação de fenômenos presentes nos dados. Os resultados obtidos nessa etapa são descobertos através de algoritmos presentes na ferramenta de *Data Mining*, Weka.

## 1.5 Descrição dos Capítulos

Este trabalho está subdividido em 5 capítulos. O Capítulo 1, em questão, apresentou o contexto, a justificativa, o problema e os objetivos que nortearam o desenvolvimento desta pesquisa.

A organização textual deste trabalho segue a estrutura abaixo:

- Capítulo 2 - Formalização do Poder no Brasil: apresenta uma visão geral sobre a formalização do poder no Brasil e sua importância política. Dentre os temas tratados encontram-se: Conceituação de Poder, a Tripartição de Poderes no Brasil, o Poder Legislativo e o Processo Legislativo até chegar nas Proposições Legislativas.
- Capítulo 3 - O Papel da Informação para a Sociedade: apresenta o conceito e alguns aspectos relativos à informação no que tange a sua importância perante a sociedade e busca abordar o posicionamento do Brasil quanto à transparência de suas informações públicas para a população.
- Capítulo 4 - Análise de Dados: apresenta uma visão geral das metodologias e ferramentas de análise dados, passando por conceitos e ferramentas de BI e Mineração de Dados.
- Capítulo 5 - Análise dos Dados de Votações de Proposições na Câmara dos Deputados: aborda a metodologia utilizada para a análise dos dados e obtenção dos resultados propostos, orientado pelas metodologias de BI e Mineração de Dados.
- Capítulo 6 - Resultados: apresenta os resultados obtido através das pesquisas descritiva e exploratória.
- Capítulo 7 - Considerações finais: apresenta um resumo do que foi feito, dos resultados alcançados e das limitações e trabalhos futuros.

# Capítulo 2

## Formalização do Poder no Brasil

Este capítulo apresenta uma visão geral sobre a formalização do poder no Brasil e sua importância política. A Seção 2.1 apresenta uma introdução aos diversos significados de poder e como se deu a divisão de poderes no Brasil. A Seção 2.2 aborda a importância do Poder Legislativo na legislação e fiscalização, bem como lista e diferencia as duas Casas legislativas membros do sistema bicameral instalado no país. Por fim, a Seção 2.3 conceitua o processo legislativo a partir de proposições legislativas, que serão recursos utilizados neste trabalho para pesquisa.

### 2.1 A tripartição de Poderes no Brasil

Em seu significado mais abrangente, a palavra poder (do latim *potere*) significa o direito de deliberar ou mandar, indicando a capacidade de agir ou de produzir efeito [32]. A partir da análise de um contexto sociológico, nota-se a existência de diversos tipos de poder, dentre os quais destacam-se o poder social, o poder econômico, o poder militar e o poder político. Quando associado ao campo social, o poder está conectado à possibilidade/capacidade de um indivíduo em determinar o comportamento de outro indivíduo. Desta forma, para Rodrigues (1999) [54] "o ser humano não é apenas o sujeito capaz de exercer poder, mas é também objeto sobre o qual o poder é exercido."

O poder se expressa nas diversas relações sociais e, onde existem relações de poder, existe política, e a política se expressa nas diversas formas de poder. Assim sendo, surge o poder político caracterizado pela capacidade de influenciar as ações dos sujeitos inseridos em um contexto social por meio das instituições políticas (Estado) que regimentam as relações desse espaço [54]. Segundo Britto (1982) [10], a Constituição Federal de 1988 (CF) possui três possíveis acepções para a expressão poder: a) o poder enquanto revelação da soberania (artigo 1º, parágrafo único da CF); b) o poder enquanto órgão do Estado (artigo 2º da CF); c) o poder enquanto função (artigos 44º, 76º e 92º da CF).

Analisando o poder enquanto órgão do Estado, observa-se que a separação dos poderes no Brasil, comumente chamada de 'tripartição dos poderes', é prevista pela atual CF, em seu artigo 2º, sendo discriminados três poderes: o Executivo, o Legislativo e o Judiciário. Temer (2014) [62] considera a utilização da expressão 'tripartição dos poderes' um equívoco, uma vez que considera o poder uma unicidade, sendo a distinção entre os

poderes dada através dos órgãos desempenhantes de funções.

O modelo de sistematização do poder que se aplica no Brasil fora idealizado por Montesquieu que propôs a criação de órgãos distintos e independentes uns dos outros e classificou os três poderes do Estado em: poder Legislativo (aquele que faz as leis); poder Executivo das coisas que dependem do direito das gentes (aquele que faz a paz ou a guerra, envia ou recebe embaixadas, instaura a segurança, previne invasões, dentre outros), que é o poder Executivo do Estado; poder Executivo das coisas que dependem do direito civil (aquele que pune os crimes, julga conflitos entre os indivíduos, dentre outros), ou seja, o poder de julgar, que é o poder Judiciário [43].

De acordo com Cantergi (2007) [12], a concepção da separação dos poderes tem por finalidade destacar três funções estatais: legislar, administrar e julgar, distribuindo tais funções a órgãos distintos, assim, como destaca Moreira (2006) [45], visando que cada um exerça sua atribuição com autonomia sem que haja qualquer meio de subordinação entre os mesmos.

## 2.2 O Poder Legislativo

O Poder Legislativo no Brasil é exercido pelo Congresso Nacional, cabe a este órgão deliberar e fiscalizar os projetos de lei de competência da União. Segundo Cantergi (2007) [12], o Congresso adota um sistema bicameral, de tal modo que a função legislativa em âmbito federal é exercida por duas câmaras distintas. Desta forma, de acordo com o artigo 44º da CF (1988) [9], o Congresso Nacional divide-se em Senado Federal, composto pelos representantes dos Estados da Federação e do Distrito Federal e em Câmara dos Deputados, composta pelos representantes do povo. Assim sendo, o Senado Federal e a Câmara dos Deputados figuram as duas casas legislativas do Brasil.

Legislar e fiscalizar não são as únicas funções do Poder Legislativo. Cada poder exerce funções típicas (próprias) e atípicas (impróprias).

São funções típicas do Poder Legislativo legislar e fiscalizar. A função legiferante envolve a elaboração de leis e atos normativos, e a função fiscalizadora pode se efetivar por meio de propostas de fiscalização e controle. No que tange às funções atípicas, o Poder Legislativo exerce as funções de administrar e julgar [14, p.25].

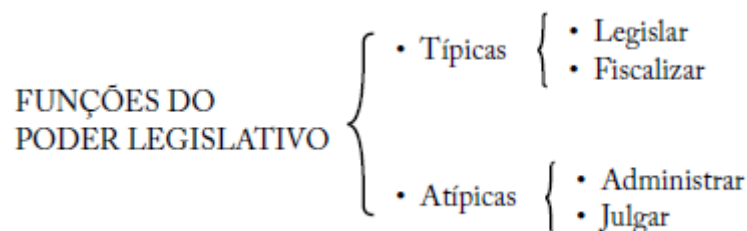


Figura 2.1: Funções do Poder Legislativo [14, p.26]



De acordo com Carneiro e Santos (2014) [14], "cabe mencionar outra função primária do Poder Legislativo: a de representar o povo, a nação". Nesse sentido, a atuação das Casas Legislativas vai além das competências típicas de legislar e fiscalizar a aplicação de recursos públicos como mencionado anteriormente. Cabe a ele, por exemplo, "debater temas importantes, o que, muitas vezes, não necessariamente resulta na inserção de uma nova norma no ordenamento jurídico ou em uma fiscalização de determinada entidade ou agente público, o que denota atuação inerente ao exercício da atividade parlamentar".

Segundo Temer (2014) [62], com o intuito de atingir o desempenho da função legislativa, as Casas se reúnem separada ou conjuntamente. Desta maneira, verificam-se reuniões da Câmara dos Deputados, do Senado e do Congresso Nacional, tendo cada Casa um papel importante no âmbito legislativo.

### 2.2.1 Câmara dos Deputados

Cantergi (2007) [12] afirma que "a Câmara dos Deputados é composta por representantes do povo, os quais são eleitos, por voto secreto e direto, pelo sistema de representação proporcional, em cada Estado, em cada Território e no Distrito Federal", revelando-se uma Casa legislativa plural, a serviço da sociedade brasileira.

Nesse contexto, a Câmara dos Deputados, autêntica representante do povo brasileiro, exerce atividades que viabilizam a realização dos anseios da população, mediante discussão e aprovação de propostas referentes às áreas econômicas e sociais, como educação, saúde, transporte, habitação, entre outras, sem descuidar do correto emprego, pelos Poderes da União, dos recursos arrecadados da população com o pagamento de tributos[25].

O número de deputados eleitos em cada Estado será proporcional à sua população, ou seja, cada unidade da Federação terá um determinado número de representantes populares de acordo com a sua quantidade de habitantes. O artigo 45º da CF determina esse o número, procedendo-se aos ajustes necessários, no ano anterior às eleições, para que nenhuma das unidades da Federação tenha menos de oito ou mais de setenta Deputados. A Lei Complementar nº 78, de 30 de dezembro de 1993 [20], estabelece que o número de Deputados não pode ultrapassar quinhentos e treze.

Segundo a CF (1988) [9], a Fundação Instituto Brasileiro de Geografia e Estatística (IBGE) fornece os dados estatísticos para a efetivação do cálculo. Feitos os cálculos, o Tribunal Superior Eleitoral encaminha aos Tribunais Regionais Eleitorais e aos partidos políticos o número de vagas a serem disputadas. Além do número mínimo de representantes, a lei determina que cada Território Federal será representado por quatro Deputados Federais.

Temer (2014) [62] ressalta que os requisitos para a elegibilidade para o cargo de deputado federal de acordo com a CF são: "ser brasileiro nato ou naturalizado, maior de vinte e um e anos e estar em pleno gozo dos direitos políticos. Os deputados são eleitos para um mandato de quatro anos, sendo permitida a reeleição".

No Brasil, o sistema eleitoral utilizado para eleger deputados federais é o sistema proporcional, onde os partidos registram vários candidatos para concorrerem ao mesmo cargo. Nesse caso, os votos oferecidos a cada candidato são direcionados para o partido, que necessita atingir determinada quantidade mínima de votos para garantir a eleição de pelo menos um de seus candidatos.

De acordo com Moraes (2004) [44, p.428], "esse tipo de eleição, no Brasil, consiste na divisão do número total de votos válidos em candidatos pelo número de cargos em disputa". O resultado dessa operação aritmética denomina-se quociente eleitoral. A partir disso, "divide-se o total de votos obtidos por cada uma das legendas pelo quociente, chegando-se, conseqüentemente, ao número de cadeiras obtidas por cada legenda".

Dessa forma, um candidato que não conseguiu votos equivalentes ao quociente eleitoral poderá ser eleito deputado, situação corriqueira em que o partido pelo qual concorreu consegue o quociente que viabiliza essa situação; por outro lado, o candidato mais votado dentre todos os participantes do pleito eleitoral poderá não ser eleito, evento mais difícil de ocorrer, que se dá caso o somatório de votos do partido pelo qual concorreu seja inferior ao quociente eleitoral [14, p.29].

## 2.2.2 Senado Federal

O Senado Federal está definido no artigo 46º da CF e nele estão os representantes dos Estados e do Distrito Federal. Como pontua Temer (2014) [62], os senadores são eleitos por voto direto e secreto, segundo o princípio majoritário em único turno, ou seja, será considerado vencedor o candidato que obtiver o maior número de votos. Cada unidade da Federação elege três senadores, cada um com dois suplentes. Assim sendo, o Senado Federal é composto atualmente por oitenta e um senadores. Cantergi (2007) [12] afirma que "os Estados e o Distrito Federal elege o mesmo número de representantes, pois o ordenamento jurídico pátrio adota o princípio da igualdade de representação dos Estados no sistema federativo."

O mandato dos senadores é de oito anos. Contudo, o Senado renova-se parcialmente de quatro em quatro anos, alternadamente, por um e dois terços (artigo 46, § 2º, CF). Os requisitos de elegibilidade para o Senado Federal são: ser brasileiro nato ou naturalizado, maior de trinta e cinco anos e estar em pleno gozo dos direitos políticos [12, p.7].

De acordo com a CF (1988) [9], os requisitos de elegibilidade para o Senado Federal são: ser brasileiro nato ou naturalizado, maior de trinta e cinco anos e estar em pleno gozo dos direitos políticos.

Os senadores são eleitos com a idade mínima de 35 anos e para mandatos de 8 anos, com renovação de um terço das cadeiras em uma eleição e dois terços nas eleições seguintes. Isso explica por que em algumas eleições você vê seu estado eleger dois novos senadores e em outras vezes vê apenas um novo senador [12].

Compete privativamente ao Senado Federal (artigo 52º da CF e Emendas Constitucionais nº 19/1998, 23/1999, nº 42/2003 e nº 45/2004):

- Processar e julgar o presidente e o vice-presidente da República, os ministros de Estado, os ministros do Supremo Tribunal Federal (STF), o procurador-geral da República e o advogado-geral da União nos crimes de responsabilidade;
- Aprovar, previamente, a indicação do presidente da República de magistrados, ministros do Tribunal de Contas da União, governador de território, presidente e diretores do Banco Central, procurador-geral da República, chefes de missão diplomática e titulares de outros cargos que a lei determinar;
- Autorizar operações externas de natureza financeira de interesse dos entes federados;
- Dispor sobre a regulamentação das agências executivas e reguladoras;
- Suspender a execução, no todo ou em parte, de lei declarada inconstitucional por decisão definitiva do Supremo Tribunal Federal (artigo 52º da Constituição Federal e Emendas Constitucionais nº 19/1998 e nº 23/1999).

Sendo, também, o Senado Federal um agente fundamental no processo de criação de leis, uma vez que, de acordo com Blume (2016) [8] "senadores podem elaborar projetos de lei, a serem analisados e votados tanto no Senado quanto na Câmara, assim como devem analisar, avaliar e aprovar ou rejeitar projetos de lei de autoria da Câmara ou do Presidente da República".

## 2.3 Processo Legislativo

Uma das principais funções do Congresso Nacional e de suas Casas (Câmara dos Deputados e Senado Federal) é a de elaborar normas legais. Para tal, faz-se necessária a existência do processo legislativo. Para Cantergi (2007) [12], Processo Legislativo pode ser entendido de duas maneiras: juridicamente, como um conjunto de atos necessários para a constituição das leis e atos normativos a serem observados pelos órgãos legislativos e, sociologicamente, como um mecanismo pelo qual o Poder Legislativo atende sua função primordial de legislar.

Moraes (2004) [44], analisando o processo legislativo sob a ótica jurídica, define-o como “o conjunto coordenado de disposições que disciplinam o procedimento a ser obedecido pelos órgãos competentes na produção das leis e atos normativos que derivam diretamente da própria constituição”, salientando que o próprio texto constitucional estabelece uma sequência de atos a serem cumpridos com o objetivo de formar as espécies normativas previstas no artigo 59º da CF.

Corroborando com a visão de Moraes (2014), Silva (2006) [57, p.42] elabora uma noção complexa, conjugando aspectos objetivos e subjetivos desse processo afirmando que o “processo legislativo é o conjunto de atos (iniciativa, emenda, votação, sanção) realizados pelos órgãos legislativos e órgãos cooperadores para o fim de promulgar leis”.

As regras do processo legislativo, no Congresso Brasileiro, são ditadas pela CF (1988) [9], pela Lei Complementar nº 95, de 26 de fevereiro de 1988 [19] (com as alterações promovidas pela Lei Complementar nº 107 [21], de 26 de abril de 2001), que dispõe sobre

a elaboração, a redação, a alteração e a consolidação das leis, pelos Regimentos Internos da Câmara dos Deputados (Resolução da Câmara dos Deputados nº 17, de 1989 [26], e alterações posteriores) e do Senado Federal (Resolução do Senado Federal nº 93, de 1970 [31] e alterações posteriores) e pelo Regimento Comum (Resolução do Congresso Nacional nº 1, de 1970 [47]).

A CF (1988) [9], em seu artigo 59º, dispõe que o processo legislativo compreende a elaboração de: emendas à Constituição, leis complementares, leis ordinárias, leis delegadas, medidas provisórias, decretos legislativos e resoluções.

O processo legislativo no Brasil, na esfera federal, é bicameral, ou seja, envolve a manifestação de vontade das duas câmaras legislativas para a produção de normas jurídicas. Assim, as espécies de normas jurídicas enumeradas no artigo 59 da CF são submetidas à manifestação de vontade das duas Casas do Congresso Nacional, separadamente ou em conjunto, como no caso dos projetos relativos às leis orçamentárias. Os vetos presidenciais aos projetos de lei também são apreciados em sessão conjunta das duas Casas [45, p.29].

Tanto o Senado Federal quanto a Câmara dos deputados podem fazer o papel de Casa iniciadora e revisora. Segundo a CF (1988) [9], em seu artigo 65º, um projeto de lei aprovado por uma Casa será revisto pela outra, em turno único de discussão e votação, e enviado à sanção ou promulgação, se a Casa revisora o aprovar, ou arquivado, se o rejeitar. Caso o projeto seja emendado, volta a Casa iniciadora. Entretanto, de acordo com o que afirma Moreira (2006) [45], é comumente dito que a Câmara dos deputados funciona como a casa iniciadora e o Senado Federal, como revisora. Tal ideia decorre dos mandamentos constitucionais, pois em casos específicos, a CF determina que Câmara seja a Casa iniciadora e o Senado, a revisora. No que se refere ao papel da Câmara dos Deputados quanto casa iniciadora, está sua participação nas discussões e votações dos projetos, uma vez que ao Senado, quanto casa revisora, compete o papel de ‘revisar’ a decisão de aprovação tomada pela primeira casa (a iniciadora).

O fluxograma definido na Figura 2.2, a seguir, mostra resumidamente o modelo de processo legislativo bicameral.

## Fluxograma constitucional da tramitação legislativa

O processo legislativo federal  
(CF, arts. 65 e 66)

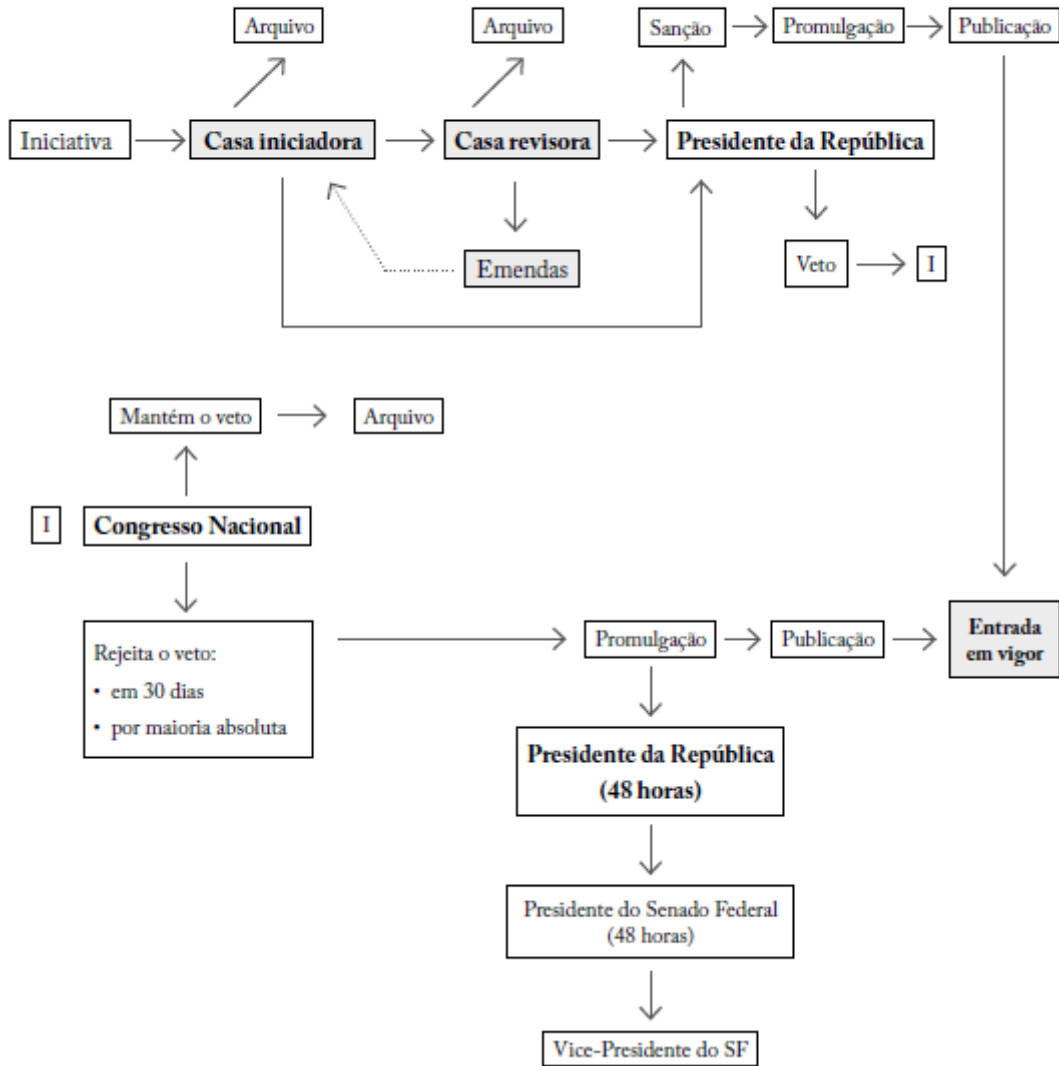


Figura 2.2: Fluxograma constitucional da tramitação legislativa [14, p.26]

7

### 2.3.1 Proposições Legislativas

Proposições legislativas são os instrumentos por meios dos quais os deputados e senadores apresentam suas ideias de novas leis (ou de alteração das existentes) e exercem sua função legisladora [30]. De acordo o artigo 100º do RICD [23], proposição é toda matéria sujeita à deliberação da Câmara e constitui um tema muito importante para a compreensão do processo legislativo. Para Carneiro e Santos (2014) [14], "trata-se de ideias organizadas de forma escrita para o exercício parlamentar e o alcance dos anseios da população brasileira".

As proposições utilizadas para iniciar o processo legislativo, uma vez aprovadas, originam normas jurídicas contempladas no artigo 59º da CF (1988) [9], nestes termos:

- (I) Emendas à Constituição;
- (II) Leis complementares;
- (III) Leis ordinárias;
- (IV) Leis delegadas;
- (V) Medidas provisórias;
- (VI) Decretos legislativos;
- (VII) Resoluções.

De acordo com Pacheco e Mendes (2015) [50], proposições podem consistir em variadas espécies como propostas de emenda à Constituição, projetos – de lei, de decreto legislativo ou de resolução –, emendas, indicações, requerimentos, pareceres, recursos ou propostas de fiscalização e controle, e, cada uma das espécies possui um processo de tramitação específico. A seguir serão abordadas as proposições que serão tratadas neste trabalho.

### **2.3.1.1 Projeto de Lei Ordinária**

O volume de proposições que tramita no âmbito da Câmara dos Deputados, em sua grande maioria, refere-se a projetos. De acordo com Carneiro e Santos (2014) [14], "Eles são tão importantes que o legislador constituinte permitiu que o cidadão exercesse a democracia direta mediante a apresentação de projeto de lei de iniciativa popular".

Segundo Pacheco e Mendes (2015) [50], os Projetos de lei são proposições destinadas a criar novas leis ou a alterar outras já em vigor, podendo ser apresentados, usualmente, tanto por parlamentares, individual ou coletivamente, quanto pelas comissões da Câmara, do Senado ou das duas Casas em conjunto, ou ainda pelo presidente da República – "que dispõe de iniciativa concorrente com a dos parlamentares sobre temas em geral, mas detém competência privativa para a apresentação de projetos sobre certas matérias definidas pela CF".

"A CF de 1988 deu abrigo ainda à iniciativa popular de leis, conferindo aos cidadãos o direito de apresentar projetos ao Congresso Nacional desde que atendida a exigência de subscrição mínima de um por cento do eleitorado nacional, distribuído por pelo menos cinco estados, com não menos de três décimos por cento dos eleitores de cada um deles. Como o atingimento desse número mínimo de subscritores é bastante difícil e complicado, têm sido muito poucos os projetos de lei de iniciativa popular apresentados até hoje"[50, p.30].

Os projetos de lei admitem dois tipos, os Projetos de Lei Ordinária (PL) e os Projetos de Lei Complementar (PLP). As diferenças entre os dois tipos estão pautadas em dois critérios, quanto ao quórum de aprovação e quanto à matéria. O PLP, segundo o artigo

69º da CF (1988) [9], é aprovado por maioria absoluta e exigido em matérias específicas da Constituição. Já o PL, objeto de estudo deste tópico, de acordo com o artigo 47º da CF (1988) [9], é aprovado por maioria simples e é exigido de modo residual, nos casos em que não houver a expressa exigência de lei complementar.

O processo de criação de uma lei ordinária se inicia com um PL. De acordo com Mendes et al. (2009) [42], esse processo é iniciado quando algum ente (membro do Congresso Nacional, Presidente de República, cidadãos, etc.), toma a iniciativa de propor uma nova lei. Dada a fase inicial, o projeto de lei será debatido nas comissões e nos plenários da Câmara dos Deputados e do Senado Federal, dando início à Fase de Discussão. O PL deverá ser apreciado nas duas Casas do Congresso Nacional (Casa Iniciadora e Revisora), separadamente, e em um turno de discussão e votação (no plenário), necessitando de maioria simples em cada uma delas. Podem ser formuladas emendas que alterem os projetos, podendo estas sofrer restrições. Tendo a aprovação das duas Casas, a sanção é dada pelo Presidente da República, conforme artigo 66º da CF (1988) [9]. Com a sanção do PL, dá-se início à Fase Complementar do processo: promulgação e publicação da lei. Temer (2014) [62] ressalva a hipótese em que haja veto na sanção dada pelo Presidente e subsequente rejeição desse veto pelo Senado.

"Nesse caso, se a lei não for promulgada dentro de 48 horas pelo Presidente da República, o Presidente do Senado a promulgará e, se este não o fizer no prazo de 48 horas (contadas a partir do escoamento do prazo de 48 horas conferido ao Presidente da República), a promulgação competirá ao Vice-Presidente do Senado, conforme aponta do artigo 66º, §7º da CF. O mesmo procedimento será seguido no caso de sanção tácita, se após, 48 horas, o Presidente não promulgar a lei"[62, p. 151].

### 2.3.1.2 Proposta de Emenda à Constituição

As Propostas de Emenda à Constituição (PEC), como o nome indica, são proposições destinadas a promover alterações no texto da Constituição vigente, sem a necessidade de convocar uma nova assembleia constituinte.

"Para serem recebidas e processadas, têm de estar assinadas, no caso de iniciativa dos parlamentares, por no mínimo um terço do total de membros da Casa, o que, na Câmara, equivale à assinatura de 171 deputados. Sua apresentação, entretanto, pode se dar ainda por parte de agentes externos ao Congresso Nacional, como o presidente da República e as assembleias legislativas das unidades da federação. No caso das assembleias legislativas, a iniciativa da apresentação só será válida se contar com o apoio de mais da metade delas, cada uma tendo tomado a decisão por deliberação da maioria de seus membros". [50, p.28]

As PECs deverão ser recebidas pela Comissão de Justiça e Cidadania (CCJ), que ficará encarregada de examinar a admissibilidade das mesmas. Se rejeitada nesta fase, a proposta deve ser arquivada. Se admitida, a PEC pode ser encaminhada para outra comissão temporária criada pela CCJ, que examinará o conteúdo da PEC, podendo propor emendas, que também devem ser submetidas ao exame de admissibilidade, ou pode ser encaminhada para deliberação e revisão. De acordo com o artigo 60º, §2º da CF (1988)



[9], a PEC é discutida e votada em cada Casa do Congresso, em dois turnos e considera-se aprovada se obtiver, em ambos, três quintos dos votos dos membros de cada uma das Casas.

Votado e aprovado o projeto, passa-se à promulgação, efetivada pelas Mesas da Câmara dos Deputados e do Senado Federal. Segundo Temer (2014) [62, p. 148], não há sanção nem veto do Presidente em caso de emenda constitucional, diferentemente do que ocorre com os PLs, pois a Constituição é de competência do povo, cujos representantes são os parlamentares. Para o autor, o Texto Constitucional não especifica sobre a responsabilidade da publicação da PEC. Entretanto, "há de entender-se que essa competência é do Congresso Nacional".

As PECs, conforme apontam Pacheco e Mendes (2015) [50], "só podem ter andamento se suas disposições não tiverem impacto sobre as chamadas 'cláusulas pétreas', que são as normas constitucionais não modificáveis, tidas como verdadeira espinha dorsal da CF (artigo 60º, §4º)". São elas:

- A forma federativa do Estado – que envolve, no Brasil, a existência de três esferas autônomas de organização político-administrativa: a União, os estados e o Distrito Federal e os municípios, cuja capacidade de autogoverno e autogestão deve ser assegurada;
- O direito universal, isto é, de todos, a escolher seus representantes no Governo e no Legislativo por meio do voto direto e secreto, exercido periodicamente;
- A independência e a harmonia entre os três poderes da República, não se permitindo o domínio de um sobre o outro;
- Os direitos e garantias fundamentais dos cidadãos.

### 2.3.1.3 Medida Provisória

A Medida Provisória (MPV) é um ato unipessoal do Presidente da República e possui força imediata de lei. Para Temer (2014) [62], MPV não é lei, pois não nasce do Legislativo. Para o autor, uma MPV é um ato que tem a 'força da lei'. Segundo Pacheco e Mendes (2015) [50], MPVs são "atos normativos com características de lei e de projeto ao mesmo tempo". São, na verdade, "leis de caráter precário, tendo força e aplicabilidade imediata, mas ficando em vigor por tempo limitado – sessenta dias, com possibilidade de prorrogação por mais sessenta, contados da publicação no Diário Oficial".

"Uma vez editada uma medida provisória, o presidente do Congresso Nacional terá o prazo de 48 horas para criar uma comissão mista de deputados e senadores, destinada a emitir parecer sobre a matéria. O parecer da comissão mista deverá abordar tanto os aspectos de admissibilidades constitucional, jurídica e orçamentária quanto de mérito da medida provisória, podendo concluir por sua aprovação ou rejeição ou ainda pela apresentação de um projeto de lei de conversão, por meio do qual são propostas alterações ao texto original da medida provisória"[50, p.63].

Emendas poderão ser feitas, por parte de deputados e senadores, perante a comissão mista, dentro dos primeiros seis dias seguintes à publicação da medida provisória no Diário



Oficial da União (DOU). Emitido o parecer da comissão mista, a medida provisória deverá ser encaminhada ao plenário da Câmara para discussão e votação.

"Caso a medida provisória não tenha sua apreciação encerrada na Casa em até 45 dias da data inicial de sua entrada em vigor, “tranca-se” a pauta de deliberações da Câmara, não se podendo mais votar nenhuma matéria enquanto não for ultimada a votação da medida provisória"[50, p.64].

Uma vez encerrado o processo de apreciação na Câmara, a MPV seguirá para revisão no Senado Federal, onde poderá vir a ser alterada, devendo, nesse caso, retornar à Câmara para apreciação das modificações lá aprovadas. Em sendo aprovadas pelo Congresso Nacional, as MPVs convertem-se definitivamente em lei, passando a integrar a ordem jurídica nessa condição.

# Capítulo 3

## O Papel da Informação para a Sociedade

O presente capítulo apresenta o conceito e alguns aspectos relativos à informação no que tange a sua importância perante a sociedade e busca abordar o posicionamento do Brasil quanto à transparência de suas informações públicas para a população. A Seção 3.1 contextualiza o momento atual em que a sociedade se encontra, bem como conceitua termos importantes como dado, informação e conhecimento. A Seção 3.2 aborda a sociedade da informação através da descrição de suas características, desafios e anseios, e a Seção 3.3 finaliza o capítulo tratando da Lei de Acesso à Informação (LAI) e como ela cumpre com seu papel na democratização e transparência da informação.

### 3.1 Dado, Informação e Conhecimento

O princípio do século XXI, segundo Titão e Viapiana (2008) [63], vem sendo marcado por algumas características: a globalização e a emergência de uma nova sociedade, que se convencionou chamar de sociedade da informação. Para Silva *et al.* (2013) [58], as tecnologias digitais estão cada vez mais avançadas, sendo cada vez mais atualizadas em um intervalo de tempo cada vez menor, permitindo que o mundo e suas culturas estejam ao alcance de todos.

Estas mudanças ocorridas na sociedade nos últimos anos, ocasionadas pela globalização, pela velocidade das informações compartilhadas, pela redefinição de tempo e espaço, pela descoberta do virtual e do digital, alterou completamente o cotidiano das pessoas e conseqüentemente do ambiente em que vivemos [63, p.1].

Os vocábulos: *dado*, *informação* e *conhecimento*, estão cada vez mais presentes na rotina da sociedade. De acordo com Sene (2008) [55], no senso comum e, às vezes, mesmo nos meios científicos, esses conceitos são utilizados imprecisamente e, muitas vezes, de forma intercambiável, embora tenham conceitos diferentes.

Siqueira (2005) [60] descreve dado como sendo a menor unidade, ou unidade primária que compõe a informação. Cardoso e Machado (2008) [13] identificam dados como "fatos, imagens ou sons que podem ou não ser úteis ou pertinentes para uma atividade particular. São abstrações formais quantificadas, que podem ser armazenadas e processadas por

computador". Já Simon (1999) [59], acredita que "dado é uma sequencia de símbolos, é um ente totalmente sintático, não envolve semântica como na informação".

No que tange ao conceito de informação, para Sene (2008) [55], é evidente que para existir informação é necessário haver comunicação, sendo assim, a autora afirma que a informação é o dado com algum significado, com algum sentido. Segundo Azevedo e Santos (2015) [6], "a informação é o resultado do conjunto de unidades de dados, dotados de significado e relevância para quem a utiliza". São dados contextualizados, que segundo Cardoso e Machado (2008) [13], possuem forma e conteúdo apropriados para um uso particular. "São abstrações informais (não podem ser formalizadas segundo uma teoria matemática ou lógica) que representam, por meio de palavras, sons ou imagens, algum significado para alguém."

No que diz respeito ao conceito de conhecimento, para Cardoso e Machado (2008) [13], "pode ser entendido como uma combinação de instintos, ideias, informações, regras e procedimentos que guiam ações e decisões; tem embutido em si valores como sabedoria e *insights*. *É a inteligência obtida pela experiência*".

O conhecimento, por sua vez, pode ser subdividido em duas categorias: o explícito, que é aquele que pode ser codificado e apresentado por meio de uma linguagem formal, sendo mais plausível de ser estruturado, manipulado e transmitido; e o tácito, que é o conhecimento abstrato, subjetivo a cada pessoa. Pode ser composto por crenças, valores, perspectivas, vivências e experiências do indivíduo, que são variáveis intangíveis [6, p. 27 apud [13]].

## 3.2 Sociedade da Informação

A sociedade da informação, como é comumente chamada, não é um modismo e está sendo concebida em diversos países. Segundo Werthein (2000) [69], "a expressão '*sociedade da informação*' passou a ser utilizada, nos últimos anos desse século, como substituto para o conceito complexo de '*sociedade pós-industrial*'. Para Takahashi (2000) [61], o termo '*sociedade da informação*' é "um fenômeno global, com elevado potencial transformador das atividades sociais e econômicas, uma vez que a estrutura e a dinâmica dessas atividades inevitavelmente serão, em alguma medida, afetadas pela infra-estrutura de informações disponível". Representa uma profunda mudança na organização da sociedade e da economia, havendo quem a considere um novo paradigma técnico-econômico.

Tal paradigma, segundo Werthein (2000) [69, apud Castells (2000) [15]] possui as seguintes características fundamentais:

- **A informação é sua matéria-prima:** as tecnologias se desenvolvem para permitir o homem atuar sobre a informação propriamente dita, ao contrário do passado quando o objetivo dominante era utilizar informação para agir sobre as tecnologias, criando implementos novos ou adaptando-os a novos usos.
- **Os efeitos das novas tecnologias têm alta penetrabilidade:** a informação é parte integrante de toda atividade humana, individual ou coletiva

e, portanto, todas essas atividades tendem a serem afetadas diretamente pela nova tecnologia.

- **Predomínio da lógica de redes:** esta lógica, característica de todo tipo de relação complexa, pode ser, graças às novas tecnologias, materialmente implementada em qualquer tipo de processo.
- **Flexibilidade:** a tecnologia favorece processos reversíveis, permite modificação por reorganização de componentes e tem alta capacidade de reconfiguração.
- **Crescente convergência de tecnologias:** principalmente a microeletrônica, telecomunicações, optoeletrônica, computadores, mas também, e crescentemente, a biologia. O ponto central aqui é que trajetórias de desenvolvimento tecnológico, em diversas áreas do saber, tornam-se interligadas e transformam as categorias segundo as quais pensamos todos os processos.

Para Werthein (2000) [69], é preciso reconhecer que muitas das promessas desse paradigma tecnológico foram e estão sendo realizadas. Entretanto, "os desafios da sociedade da informação são inúmeros e incluem desde os de caráter técnico e econômico, cultural, social e legal, até os de natureza psicológica e filosófica". Para o autor, um desses desafios está ligado ao aumento do volume de informação de qualidade e de domínio público disponível na Internet nos idiomas de expressão da população de cada sociedade. "Isso envolverá convencer o governo e centros produtores de conhecimento financiados por recursos públicos a tornarem disponíveis ao público as informações produzidas".

Na era da Internet, o Governo deve promover a universalização do acesso e o uso crescente dos meios eletrônicos de informação para gerar uma administração eficiente e transparente em todos os níveis. A criação e manutenção de serviços equitativos e universais de atendimento ao cidadão contam-se entre as iniciativas prioritárias da ação pública [61].

Takahashi (2000) [61] conclui que "o advento da sociedade da informação é o fundamento de novas formas de organização e de produção em escala mundial, redefinindo a inserção dos países na sociedade internacional e no sistema econômico mundial". Os países economicamente desenvolvidos, bem como boa parte dos emergentes, já adotam políticas e iniciativas voltadas para a sociedade da informação. Ao Brasil urge acelerar o processo de articulação efetiva de um programa nacional no sentido de assegurar a perspectiva de que seus benefícios efetivamente alcancem a todos os brasileiros, de forma a terem como consequência, o surgimento de novas demandas dirigidas ao Poder Público no que diz respeito ao seu próprio funcionamento.

### 3.3 Lei de Acesso à Informação

Conforme abordado na seção 3.2, o caminho rumo à sociedade da informação é repleto de desafios em todos os países. Contudo, cada um enfrenta um desafio diferente, refletindo, assim, em uma combinação singular de oportunidades e de riscos. Para Takahashi (2000) [61], todos os países caminham, voluntária ou involuntariamente, rumo à sociedade da informação. Entretanto, compete a cada um encontrar sua rota e suas prioridades.

No Brasil, há a busca por uma administração transparente e centrada no cidadão, ou seja, a busca por um governo ao alcance de todos. Para tal, necessita-se de uma administração pública mais transparente, eficaz e voltada para a prestação de informações e serviços à população.

Emissão de documentos, prestação de informações ligadas aos serviços públicos, acompanhamento das ações de governo e condução dos negócios públicos, acesso aos governantes e representantes eleitos são exemplos das possibilidades do uso das tecnologias de informação e comunicação pela máquina administrativa pública. A tecnologia pode ainda ser largamente aplicada para aperfeiçoar a própria gestão do governo – coordenação, planejamento, execução e controle de ações, contabilidade pública etc. – e suas transações comerciais com o setor privado. A possibilidade de acesso aos serviços, de participação nas decisões e acompanhamento dos atos governamentais por parte de todos os cidadãos, portanto, impõe a adoção de meios e métodos digitais por parte do governo, em todos os poderes constituídos e níveis governamentais, do emprego das tecnologias de informação e comunicação em benefício da eficácia, responsividade, transparência e governança [61].

Abdala e Nascimento [3, apud CF (1988) [9]] apontam que "a Constituição Brasileira em seu artigo 5º, define que todas as pessoas têm o direito de receber informações de qualquer órgão público de interesse pessoal ou coletivo desde que esses dados não comprometam a segurança da sociedade e do Estado". O artigo 37º da mesma Carta dispõe que "toda a administração pública direta, indireta ou fundacional, de qualquer dos Poderes da União, do Estado, do Distrito Federal e dos Municípios obedecerá ao princípio da legalidade, impessoalidade, moralidade, publicidade e eficiência". Além do artigo 216º, § 2º, da Lei Fundamental, constituindo, desse modo, o marco tutelador do acesso à informação pública.

Apesar dessas garantias constitucionais, não havia no Brasil uma lei unificada sobre o acesso à informação e à transparência e que pudesse ser aplicada a todos os órgãos da administração pública.

Em 18 de novembro de 2011 foi promulgada a Lei nº 12.527 e entrou em vigor em 16 de maio de 2012. Essa norma representa um grande avanço na trajetória democrática brasileira. Denominada Lei de Acesso à Informação, essa norma vem ampliar a transparência das atividades nos três Poderes da União, Estados, Distrito Federal e Municípios, inclusive aos Tribunais de Conta e Ministério Público. Entidades privadas sem fins lucrativos também são obrigadas a dar publicidade a informações referentes ao recebimento e à destinação dos recursos públicos por elas recebidos.

Segundo o Governo Federal [29], alguns aspectos da LAI a serem destacados são:

- **Acesso é a regra** e o sigilo é a exceção (divulgação máxima)
- Requerente **não precisa dizer por que e para que** deseja a informação (não exigência de motivação)
- **Hipóteses de sigilo são limitadas** e legalmente estabelecidas (limitação de exceções)

- **Fornecimento gratuito** de informação, salvo custo de reprodução (gratuidade da informação)
- **Divulgação proativa** de informações de interesse coletivo e geral (transparência ativa)
- Criação de **procedimentos e prazos** que facilitam o acesso à informação (transparência passiva)

De acordo com a Cartilha de Orientação ao Cidadão (2012) [24] disponibilizada pela Câmara dos Deputados, os objetivos dessa lei são de fomentar o desenvolvimento de uma cultura de transparência e controlar socialmente a administração pública. Para isso, a divulgação de informações de interesse público ganha procedimentos, a fim de facilitar e agilizar o acesso de qualquer pessoa, inclusive com o uso da tecnologia da informação (TI).

O acesso à informações, nos termos da Lei nº 12.527/11, tem duas vertentes:

1. Disponibilização para consulta: situação em que a informação é solicitada pelo interessado
2. Divulgação: publicação em sítios oficiais na internet, obrigatória para informações de interesse coletivo ou geral

Para que o direito de acesso seja respeitado, o Governo Federal [28] indica que foram estabelecidos dois responsáveis para garantir o cumprimento da LAI no Poder Executivo Federal.

A primeira é a **autoridade de monitoramento** (prevista na LAI, no artigo 40<sup>o</sup>), responsável por verificar o cumprimento da LAI no ente público a que pertence. Cada órgão e entidade deve indicar um dirigente para desempenhar essa atribuição. Cabe também à autoridade de monitoramento recomendar medidas para aperfeiçoar as normas e procedimentos necessários à efetividade do acesso à informação na instituição.

A outra é a **Controladoria-Geral da União (CGU)**, responsável pelo monitoramento da Lei em todo Poder Executivo Federal. Cabe à CGU o fomento à cultura da transparência e a conscientização sobre o direito de acesso à informação. É de responsabilidade também da CGU publicar informações estatísticas sobre a implementação da Lei de Acesso e preparar relatório anual a ser encaminhado ao Congresso Nacional.

A LAI prevê exceções ao acesso à informação nos casos em que as informações são classificadas como sigilosas pelas autoridades competentes, relacionadas às demais hipóteses legais de sigilo ou quando são informações pessoais, relativas à intimidade, vida privada, honra e imagem.

# Capítulo 4

## Análise de Dados

Neste capítulo será apresentado uma visão das metodologias e ferramentas de análise dados. A Seção 4.1 aborda a criação do termo *Business Intelligence* (BI), bem como seu conceito, componentes, e ferramentas, com ênfase no QlikView. A Seção também apresenta o conceito de BI ágil, nova tendência de implementação de BI. A Seção 4.2 apresenta o conceito de *data mining* e as etapas do processo de extração do conhecimento, lista também algumas sugestões de programas (*softwares*) computacionais disponibilizados para análise e mineração de dados com foco no Weka.

### 4.1 *Business Intelligence* (BI)

O conhecimento tem sido reconhecido como um dos mais importantes recursos para uma sociedade ou uma organização, como visto no no Capítulo 2. A partir do conhecimento, torna-se possível o surgimento de ações inteligentes nos planos organizacional e individual, induzindo inovações e capacidade de criar produtos e serviços excelentes em termos de complexidade, flexibilidade e criatividade.

De acordo com Cardoso e Machado (2008) [13], o "processo de gestão do conhecimento abrange toda a forma de gerar, armazenar, distribuir e utilizar o conhecimento, tornando necessária a utilização de tecnologias de informação para facilitar o processo, devido ao grande aumento no volume de dados". Corroborando com essa visão, Goldschmidt e Passos (2005) [35] apontam que a análise de grandes volumes de dados pelo homem é inviável sem o auxílio de ferramentas computacionais apropriadas. Portanto, "torna-se imprescindível o desenvolvimento de ferramentas que auxiliem o homem, de forma automática e inteligente, na tarefa de analisar, interpretar e relacionar esses dados para que se possa desenvolver e selecionar estratégias de ação em cada contexto de aplicação".

Entretanto, Cardoso e Machado (2008) [13] alertam que a coleta e o armazenamento de dados, por si só, não contribuem para melhorar a estratégia da organização.

"É necessário que sejam feitas análises sobre essa grande quantidade de dados, estabelecendo-se indicadores para descobrir padrões de comportamento implícitos nos dados, assim como relações de causa e efeito. Processar e analisar as informações geradas pelas enormes bases de dados atuais de forma correta estão entre os requisitos essenciais para uma boa tomada de decisão."

Nesse contexto, em 1989 surgiu o termo *Business Intelligence* (BI), criado e difundido pela empresa *Gartner Group*. De forma direta, o termo 'Inteligência de Negócio' ou BI, é definido por Barbieri (2001) [7, p.5 apud Angeloni e Reis (2006) [5]] como:

"Um guarda-chuva conceitual, visto que se dedica à captura de dados, informações e conhecimentos que permitam às empresas competirem com maior eficiência em uma abordagem evolutiva de modelagem de dados, capazes de promover a estruturação de informações em depósitos retrospectivos e históricos, permitindo sua modelagem por ferramentas analíticas. Seu conceito é abrangente e envolve todos os recursos necessários para o processamento e disponibilização da informação ao usuário".

Segundo Primak (2008) [52], BI é "o processo inteligente de coleta, organização, análise, compartilhamento e monitoração de dados [...], gerando informações para o suporte à tomada de decisões no ambiente de negócios". Vieira (2001) [68, apud Shariat e Hightower (2007) [56]] complementa a visão de Primak quando aponta que o BI tem como objetivo fornecer suporte estratégico, tático e operacional necessário ao processo de tomada de decisão.

Haberkorn (2004) [37] esclarece que o BI "transforma dados em informações, flexibilizando o modo de visualizá-las de acordo com as solicitações dos usuários". Nota-se, desta forma, que o BI age em prol das necessidades dos usuários. Contudo a inteligência advém do conhecimento dos tomadores de decisão, conforme aponta Vanti (2003) [66, p.72, tradução minha] a seguir:

"BI é uma ferramenta capaz de automatizar a inteligência. Porém a inteligência não é do BI, mas dos executivos que trabalham com os fatores macro e micro econômicos e que impactam no negócio. Esta inteligência está na criatividade de cada organização. O diferencial é o que faz se faz através do BI e que gera informação de valor".

Segundo Noronha (2013) [48], "a grande vantagem do conceito de *Business Intelligence* é justamente a capacidade que o sistema possui para 'traduzir' os dados armazenados em uma linguagem de fácil assimilação pelo corpo gerencial das empresas".

### 4.1.1 Componentes do BI

Como evidenciado na Seção 4.1, BI refere-se ao processo de coleta, organização, análise, compartilhamento e monitoramento de informações que oferecem suporte a gestão de negócio, sendo um conjunto de teorias, metodologias, processos, estruturas e tecnologias que transformam, de forma flexível, uma grande quantidade de dados brutos em informação útil para tomadas de decisões estratégicas.

Desta forma, o BI é composto por ferramentas que realizam trabalhos de armazenamento, mineração e análise dos dados. A Figura 4.1 representa essas ferramentas e a interação entre elas.



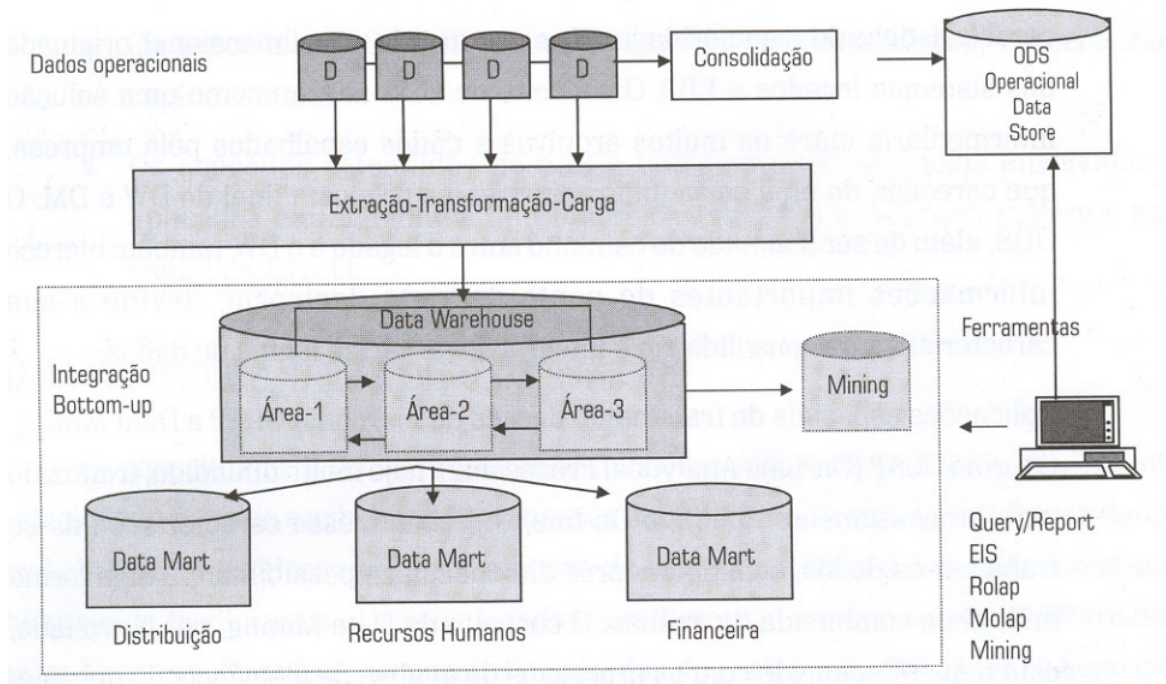


Figura 4.1: Componentes de um ambiente de BI [7]

A seguir será feita a análise individual dos elementos presentes na Figura 4.1.

#### 4.1.1.1 Dados Operacionais

Segundo Barbieri (2001) [7], o primeiro passo para a implementação de um BI é possuir os dados operacionais armazenados em, pelo menos, uma base de dados. "Estes dados originam-se do processo operacional da empresa e são armazenados pelo(s) sistema(s) de informação utilizado(s) pela organização". São a matéria-prima do BI.

#### 4.1.1.2 *Operational Data Store*

Primak (2008) [52] aponta que o *Operational Data Store* (ODS) tem a função de realizar "[...] o armazenamento de dados operacionais de forma consolidada, porém não possui características dimensionais". Barbieri (2001) [7] ressalta a importância que o ODS possui no processo de análise de dados pelas ferramentas de BI, "pois fornece importantes informações no processo decisório, devido a suas características de consolidação e integração com as possíveis várias fontes de dados operacionais da organização, de forma centralizada e íntegra".

#### 4.1.1.3 Extração, Transformação e Carga

Segundo Vassiliadis e Simitsis (2009) [67], Extração, Transformação e Carga (ETL) são processos responsáveis pelas operações que ocorrem no contexto de uma arquitetura

de *Data Warehouse*.

Em uma descrição de alto nível de um processo de ETL, os dados, em um primeiro momento, são extraídos das bases de dados fonte através de um processamento *on-line* de transações (OLTP), através do acesso à sistemas legados, páginas *web*, vários tipos de documentos (planilhas e documentos de texto) ou dados provindos de um *streaming*. Normalmente, apenas os dados que diferem da execução anterior de um processo de ETL (dados recém-inseridos, atualizados e excluídos) devem ser extraídos a partir das fontes, ou seja, não há necessidade de buscar dados já conhecidos e tratados novamente.

Em segundo lugar, os dados extraídos são propagados para uma área com fins especiais de armazenamento, chamado de área de teste de dados (DSA), onde sua transformação, homogeneização e limpeza ocorrem.

Finalizando, o ETL estrutura fisicamente os dados e, postumamente, os carrega dentro da camada de apresentação fazendo a entrega ao DW.

#### 4.1.1.4 *Data Warehouse e Data Mart*

*Data Warehouses* (DW) e *Data Marts* (DM) são estruturas especiais de armazenamento de informações com o intuito de serem utilizadas na tomada de decisão como elementos diferenciais e competitivos. Resumidamente, DW e DM são um conjunto de dados disponíveis, organizados por temas de interesse e integrados por data, facilitando o gerenciamento de grandes quantidades de dados e modelando-os para auxiliar a tomada de decisão.

Primak (2008) [52] diferencia um DW/DM de um banco de dados comum. Segundo ele, um banco de dados comum é uma "[...] coleção de dados operacionais armazenados e utilizados pelo sistema de aplicação de uma empresa específica". Já DW e DM, segundo Primak (2008) [52], é "[...] uma coleção de dados derivados dos dados operacionais para sistema de suporte a decisão. Esses dados derivados são, muitas vezes, referidos como dados gerenciais, informacionais e analíticos".

Vale salientar que tanto o DW quanto o DM possuem as mesmas características, o que os diferencia é o escopo que delimita a atuação de ambos. Para Barbieri (2001) [7], DW é um armazém de dados que deve ser "capaz de sustentar a camada de inteligência da empresa e possível de ser aplicada aos negócios, como elementos de diferenciais e competitivos". Já para Inmon (1997) [40], DW é "um conjunto de dados, não volátil, orientado a tópicos, integrado, que varia com o passar do tempo e que serve de suporte para o processo de tomada de decisão da gerencia". O DM, por sua vez, é um DW departamental, um DW reduzido, e não fornece informação para toda a empresa, mas somente para o seu departamento. Geralmente um DW é um conjunto de *Data Marts*.

#### 4.1.1.5 *Mineração de Dados*

Para que o uso de DW e DM seja interessante, é necessário utilizar boas ferramentas de mineração de dados (*Data Mining*). Desta forma, a mineração de dados é fator decisivo

para o processo de BI. Primak (2008) [52] menciona que o *Data Mining* está mais relacionado com processos de análise de inferência do que com os de análise dimensional de dados. Desta forma, a mineração de dados representa uma forma de busca de informação baseada em algoritmos que objetivam o reconhecimento de padrões escondidos nos dados e não necessariamente revelados pelas outras abordagens analíticas.

A mineração de dados será abordada mais profundamente na Seção 4.2 deste trabalho.

#### 4.1.1.6 Visualização dos Resultados

Por fim, após todo o processo descrito acima, o resultado obtido precisa ser visualizado e acompanhado de forma flexível e amigável pelo usuário do BI. Para tal, faz-se necessário o uso de ferramentas que auxiliem nessa visualização.

#### 4.1.1.7 Desvantagens do BI

O BI, assim como qualquer outra tecnologia possui algumas desvantagens. Muntean e Surcel (2013) [46] cita os principais pontos, conforme demonstrado na tabela 4.1.

Tabela 4.1: Desvantagens do BI [46, adaptado]

Desvantagens	Justificativas
Enorme quantidade de dados duplicados	-Todas as mudanças já feitas requerem uma mudança extra de dados duplicados -Inconsistência nos dados -Riscos em assegurar a qualidade de dados
Usar diferentes ferramentas para diferentes tarefas	-Especificações não compartilhadas de metadados -Resultados inconsistentes
Modelos de dados relacionais ou multidimensionais rígidos	-Flexibilidade limitada às mudanças -Suporte limitado para análise de dados não estruturados e externa

### 4.1.2 BI Ágil

Conforme abordado na Seção 4.1.1, conclui-se que o estilo do BI é predominantemente controlado, dirigido e entregue pela TI corporativa. Segundo Muntean e Surcel (2013) [46], o BI, muitas vezes, apresenta apenas pontos de vista estáticos dos dados disponíveis e quaisquer alterações ou aperfeiçoamentos devem ser feitos pela organização de TI. Estas características estão em confronto com a realidade, uma vez que, a frequência com que há mudanças nos requisitos de negócios é alta, além de não abordar a análise de *'big data'*.

Um estudo feito pela TDWI *Research* [53], mostrou que muito dos sistemas de BI não são ágeis. Segundo o estudo, cerca de 33% precisaram de mais de três meses para adicionar uma nova fonte de dados a um sistema de BI existente na organização, além de apontar que um painel ou relatório complexo de BI demoraria, em média, sete semanas para ser desenvolvido.

Segundo Muntean e Surcel (2013) [46], a forma de resolver esses problemas é por meio da aplicação de soluções de BI ágil que viabilizem a entrega rápida e íntegra de informações aos usuários certos, na hora certa. Para os autores, os componentes chave que, juntamente, promovem uma solução de BI ágil, conforme a Figura 4.2 são: um desenvolvimento ágil, uma análise de negócio (BA) ágil e, por fim, uma infraestrutura de informações ágil.

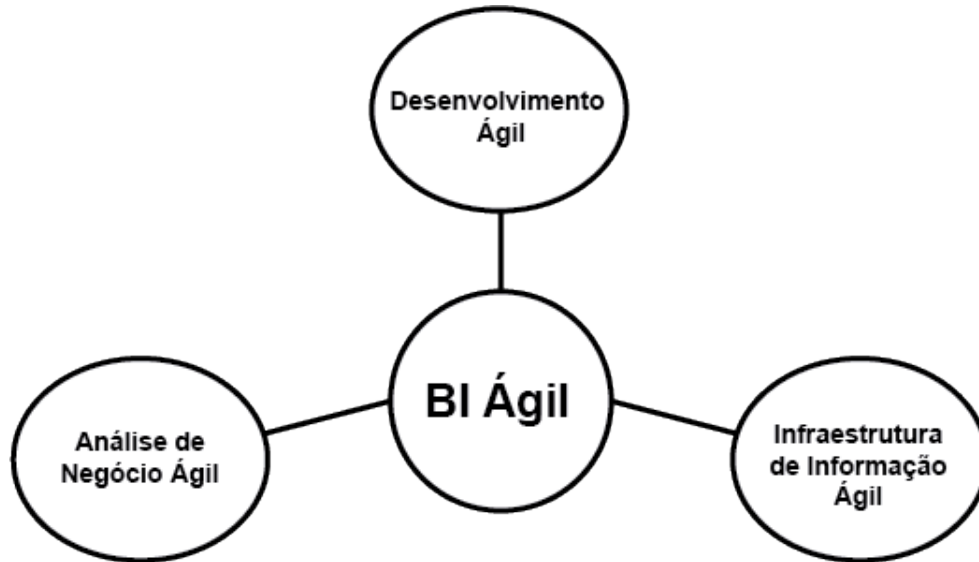


Figura 4.2: Componentes chave que promovem uma solução de BI ágil [46, p.115]

O desenvolvimento ágil é garantido pelo uso de metodologias de desenvolvimento ágil baseadas na colaboração entre equipes multifuncionais, desenvolvimento iterativo e tolerância à mudanças. Existem diferentes metodologias utilizadas no mercado atualmente, podendo ser aplicadas nos projetos de BI desde a estruturação dos times, gerenciamento dos projetos, *design* do sistema de BI, desenvolvimento das aplicações até auxiliar na aplicação das técnicas de análise.

Para garantir uma análise do negócio ágil é necessário, além de metodologias ágeis, ferramentas que permitam o desenvolvimento ágil das aplicações, permitindo que os usuários tenham maior independência da TI. Sendo assim, as ferramentas devem ser de fácil utilização por todos os usuários, possuindo características avançadas de visualização, como painéis interativos que permitam uma exploração em diferentes níveis de detalhe das informações.

Existem várias tecnologias que dão suporte aos requisitos supracitados, dentre elas destacam-se o BI como serviço, conhecido como *SaaS (Software-as-a-Service)* e as tecnologias de BI *in-memory*. O primeiro tem como vantagem a utilização de aplicações pré-feitas e sua disponibilização em nuvem. Já as tecnologias *in-memory*, trazem ganho no tempo de desenvolvimento, tendo como características a facilidade de utilização, processamento em memória, rapidez na resposta e baixo custo. Entre as ferramentas *in-memory* disponíveis no mercado, destacam-se o IBM Cognos, Microstrategy e o QlikView.

Para Muntean e Surcel (2013) [46], a agilidade só é alcançada fazendo com que todas as partes do sistema de BI sejam ágeis. Desta forma, a infraestrutura da informação possui um papel importante, uma vez que, entrega uma arquitetura de dados e uma infraestrutura de integração que garante a agilidade para reagir às mudanças nos requisitos de negócios. Desta forma, para ser ágil necessita que haja 'virtualização' dos dados, ou *data virtualization*. Segundo van der Lans (2012) [64, 65, tradução minha], *data virtualization* é

a tecnologia que oferece aos consumidores de dados uma visão unificada, resumida e encapsulada para consultar e manipular dados armazenados em um conjunto heterogêneo de armazenamentos de dados. A virtualização de dados significa a transformação de dados *on-demand*, integração de dados *on-demand* e limpeza de dados *on-demand*.

Essa virtualização pode ocorrer de diversas formas, tais como: pela utilização de um servidor de virtualização de dados ou pela disponibilização de fontes de dados em nuvem.

Sendo assim, as soluções de BI ágeis permitem que as organizações se adaptem às novas condições de mercado que, atualmente, é caracterizado pelas constantes mudanças de negócio, além de exigir cada vez mais informações estratégicas e concisas com o intuito de auxiliar os usuários em suas tomadas de decisões.

### 4.1.3 Ferramentas de BI

O processo para desenvolvimento de um BI tende a ser custoso e complexo, uma vez que, depende de várias circunstâncias que incluem desde a expertise dos profissionais, até a organização dos processos de negócio da empresa. Entretanto, o sucesso da solução de BI dependerá de uma boa ferramenta.

Atualmente, existem diversas ferramentas de BI no mercado, sendo elas as responsáveis pela interface que o usuário final terá com informações armazenadas na estrutura de BI, tornando-as de grande importância dentro da solução como um todo.

A Gartner Group<sup>1</sup>, além de ter criado o termo BI, conforme dito na Seção 4.1, periodicamente divulga relatórios de análise das principais tecnologias do mercado de BI, através de um gráfico conhecido como 'Quadrante Mágico do Gartner'<sup>2</sup>. Este relatório analisa cada fornecedor de BI e *Analytics* e mostra as posições relativas de cada um no mercado. O quadrante possui dois eixos [34]:

- No eixo X (horizontal) temos a abrangência da visão da empresa em relação a tecnologia;
- No eixo Y (vertical) temos a capacidade de executar o que se propõem.

Esses dois eixos geram quatro quadrantes, onde se encaixam as empresas que:

- *Leaders*: Entregam o que prometem e são inovadores;
- *Visionaries*: Possuem visão, mas ainda não possuem capacidade para entrega;
- *Niche Players*: São de nichos de mercado. Não possuem grande expressão no mercado geral como um todo e comumente possuem produtos específicos;
- *Challengers*: Possuem boa capacidade de execução mas que não agregam tanto em inovação.

Quanto mais a direita e acima, melhor posicionada está a empresa no quadrante. A Figura 4.3 apresenta o 'Quadrante Mágico do Gartner' para BI e Plataformas de Análise de 2016.

---

<sup>1</sup><http://www.gartner.com/technology/about.jsp>

<sup>2</sup>[http://www.gartner.com/technology/research/methodologies/research\\_mq.jsp](http://www.gartner.com/technology/research/methodologies/research_mq.jsp)



Figura 4.3: 'Quadrante Mágico do Gartner' para BI e Plataformas de Análise - Fevereiro de 2016 [34]

O quadrante de *Leaders* possui a presença de três grandes empresas que dominam o mercado de BI e *Analytics*, são elas: a *Tableau*, a *Qlik* e a *Microsoft*. Como foco neste trabalho, analisaremos umas das ferramentas fornecidas pela *Qlik*, o **QlikView**.

#### 4.1.3.1 QlikView

Desenvolvido pela empresa sueca Qlik em 1996, o QlikView é uma solução de acesso a dados que permite analisar e usar informações de diferentes fontes de dados [1]. É uma plataforma de BI que combina a interação do usuário com análises de informações gráficas de fácil compreensão através de *dashboards* dinâmicas.

O QlikView utiliza a tecnologia '*Associative In-Memory*'. A ferramenta cria um grande arquivo com todas as associações entre os dados, formando o *Data Cloud* ou nuvem dados. Desta forma, todos os dados relevantes em todas as dimensões são carregados na memó-



ria RAM do computador, tendo como resultado um ganho significativo em desempenho [2].

Essa modelagem associativa é feita através da tecnologia AQL (acrônimo de *Associative Query Logic*), patenteada pela Qlik e utilizada no QlikView. O AQL promove análises não hierárquicas de dados, garantindo assim análises multilaterais, ou seja, qualquer valor da estrutura de dados pode ser o ponto de partida para a análise. A Figura 4.4 abaixo lista alguns diferenciais do AQL.

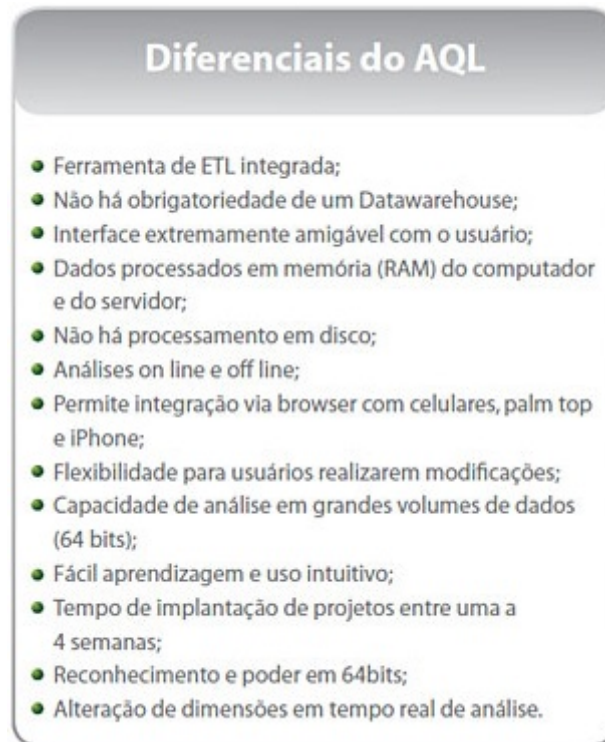


Figura 4.4: Diferenciais do AQL [22]

Para a construção de *dashboards* no QlikView, primeiramente é necessária uma conexão com o banco de dados OLTP onde os dados estão armazenados. A ferramenta possui conectores nativos com bases de dados conhecidos como *Oracle*, e *SQL Server*, além de possibilitar acesso via ODBC (acrônimo de *Open Data Base Connectivity*) e OLEDB (acrônimo de *Object Linking and Embedding Database*), acesso a arquivos de dados como arquivos de texto, planilhas, arquivos de dados como arquivos de textos, planilhas, arquivos *web*, arquivos XML (acrônimo de *eXtensible Markup Language*), entre outros.

Feita a conexão com a fonte dos dados, relacionam-se todas as tabelas necessárias para a análise e efetua-se a carga dos dados para um banco de dados proprietário da ferramenta, chamado QVD (acrônimo de *QlikView Data*). Tendo os dados carregados em memória, a plataforma disponibiliza uma interface amigável que possibilita a criação de painéis analíticos dinâmicos e intuitivos.



## 4.2 Mineração de Dados

Conforme analisado na Seção 4.1, o grande aumento no volume de dados evidenciado nos últimos tempos, exige um olhar tecnológico apropriado no que tange a disponibilização de ferramentas e metodologias computacionais apropriadas para lidar com tamanha vazão. Para Cardoso e Machado (2008) [13], "uma metodologia emergente, que tenta solucionar o problema da análise de grandes quantidades de dados e ultrapassa a habilidade e a capacidade humanas, é a *descoberta de conhecimento em banco de dados*."

Nesse sentido, a mineração de dados (*data mining*), técnica que faz parte de uma das etapas da descoberta de conhecimento em banco de dados, surge como uma metodologia de pesquisa e avaliação, que une conhecimentos e ferramentas da áreas de Computação e Estatística no tratamento de dados e informações e descobertas de padrões de comportamento. Para Azevedo e Santos (2015) [6, p. 32 apud Dantas e Couto (2014) [17]], os recursos computacionais disponibilizam algoritmos e softwares que buscam otimizar o processo de captura, organização, filtragem, processamento e avaliação dos dados, o que possibilita a aprendizagem de máquina, elevando a participação dos sistemas computacionais de simples entidades passivas de processamento a entidades com poder de decisão. Segundo os autores, a Estatística, por sua vez, "provê subsídios para a interpretação de dados e seus possíveis resultados".

O termo *mineração de dados* é conceituado na literatura por vários autores. Dantas e Couto (2014) [17, p. 6], definem o termo como "o processo de descoberta de vários modelos, resumos e valores derivados de uma determinada coleção de dados". Já Galvão e Marin (2009) [33], entendem mineração de dados como:

"uma das alternativas mais eficazes para extrair conhecimento a partir de grandes volumes de dados, descobrindo relações ocultas, padrões e gerando regras para prever e correlacionar dados, que podem ajudar as instituições nas tomadas de decisões mais rápidas ou, até mesmo, a atingir um maior grau de confiança".

Azevedo e Santos (2015) [6, apud Côrtes *et al.* (2002) [18]], enfatizam que embora utilize ferramentas de banco de dados e algoritmos de análise, a mineração de dados não é por si só um processo automático, que independe da intervenção humana. Para Côrtes *et al.* (2002) [18, p.1-2], a mineração de dados consiste em:

"um processo altamente cooperativo entre homens e máquinas, que visa a exploração de grandes bancos de dados, com o objetivo de extrair conhecimentos através do reconhecimento de padrões e relacionamento entre variáveis, conhecimentos esses que possam ser obtidos por técnicas comprovadamente confiáveis e validados pela sua expressividade estatística".

Um erro comumente cometido, conforme destacado por Palmeira e Santos (2014) [51] e por Cardoso e Machado (2008) [13], é a relação de similaridade que se estabelece entre os conceitos de mineração de dados e extração de conhecimento (*KDD - Knowledge Discovery in Databases*). Para Fayyad *et al.* (1996) [27], KDD refere-se ao processo global de descoberta de conhecimento útil a partir de dados. Segundo os autores, é "um processo,

de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados". A mineração de dados, por sua vez, é a aplicação de algoritmos específicos para extrair padrões a partir de dados.

Fayyad *et al.* (1996) [27] representam o processo operacional de KDD em cinco estágios, conforme a Figura 4.5, consistindo das seguintes etapas: seleção, pré-processamento, transformação, mineração dos dados e interpretação dos dados.

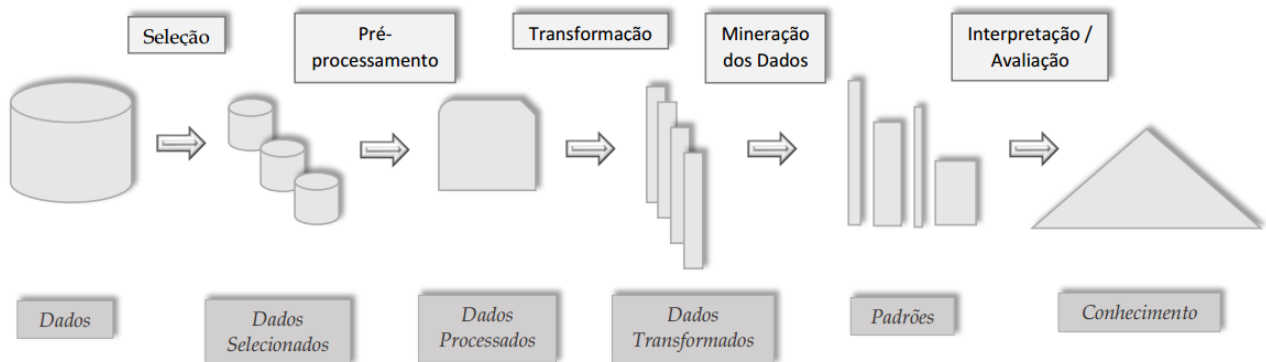


Figura 4.5: Fases do Processo de Extração de Conhecimento (KDD) [6, apud [27]]

Azevedo e Santos (2015) [6] sintetizam as Fases do KDD da seguinte forma:

- **Seleção:** A fase de seleção diz respeito à análise da disponibilidade e relevância dos dados. É nessa etapa que ocorre a seleção do conjunto ou subconjunto de variáveis ou amostras de dados, onde o processo de descoberta será executado;
- **Pré-processamento:** A fase de pré-processamento consiste na filtragem e limpeza dos dados. Os processos de filtragem e limpeza visam a remoção de dados inconsistentes, redundantes, valores faltantes ou extremos, que possam interferir nos resultados ou que sejam irrelevantes durante a mineração;
- **Transformação:** A fase de transformação consiste na formatação adequada dos dados, para que os algoritmos possam ser aplicados corretamente. Segundo Han and Kamber (2006) [38], a fase de transformação pode envolver as atividades de suavização (remoção de ruídos nos dados), agregação (sumarização dos dados), generalização (referência por conceitos de fácil entendimento (alto nível)), normalização (dimensionamento dos dados em uma faixa de valores) e criação de novos atributos;
- **Mineração de Dados:** A fase de mineração de dados consiste na exploração e análise dos dados. Nessa fase, são aplicadas as técnicas de mineração de dados, como classificação, sumarização, regressão, clusterização, dentre outras aos atributos, de modo detectar-se padrões de comportamento dos dados e gerar novas descobertas;
- **Interpretação e Avaliação:** A fase de interpretação e avaliação consiste em alcançar propriamente as informações desejadas. Após a finalização da mineração de dados, os resultados obtidos são analisados, sintetizados, avaliados e organizados para apresentação e publicação.

## 4.2.1 Tarefas e Métodos de Mineração de Dados

Como a mineração de dados vem sendo desenvolvida para os mais diferentes domínios, suas tarefas também vem se diversificando cada vez mais. Essas tarefas podem extrair diferentes tipos de conhecimento. Entretanto, é importante distinguir o que é uma tarefa do que é uma técnica de mineração. Para Amo (2004) [4], as principais diferenças entre os dois termos são:

- **Tarefa de mineração:** consiste na especificação do que estamos querendo buscar nos dados, que tipo de regularidades ou categoria de padrões temos interesse em encontrar, ou que tipo de padrões poderiam nos surpreender (por exemplo, um gasto exagerado de um cliente de cartão de crédito, fora dos padrões usuais de seus gastos);
- **Técnica de mineração:** consiste na especificação de métodos que nos garantam como descobrir os padrões que nos interessam.

A seguir, serão descritas as principais tarefas e técnicas de mineração de dados.

### 4.2.1.1 Tarefas de Mineração de Dados

Como visto anteriormente, tarefa é um tipo de problema de KDD a ser solucionado. Camilo e Silva (2009) [11, p.9 apud LAROSE (2005) [41]] afirmam que a mineração de dados é comumente classificada pela sua capacidade em realizar determinadas tarefas.

As tarefas mais comuns são:

#### Classificação e Predição

A tarefa de classificação consiste em construir um modelo de algum tipo que possa ser aplicado a dados não classificados visando categorizá-los em classes. Segundo Camilo e Silva (2009) [11], "nesta tarefa, o modelo analisa o conjunto de registros fornecidos, com cada registro já contendo a indicação à qual classe pertence, a fim de 'aprender' como classificar um novo registro (aprendizado supervisionado)". Cardoso e Machado (2008) [13, p. 506] definem a classificação na mineração de dados como "o processo de criar modelos (funções) que descrevem e distinguem classes ou conceitos, baseados em dados conhecidos, com o propósito de utilizar esse modelo para predizer a classe de objetos que ainda não foram classificados".

Dantas e Couto (2014) [17, p. 11 apud Gorunescu (2011) [36]] definem o processo de classificação em quatro componentes fundamentais, sendo eles:

- **Classe:** é a variável categórica que representa o rótulo colocado em objetos logo após sua classificação, por exemplo, lealdade de um cliente, classificação das estrelas, classe de um terremoto, etc.
- **Preditores:** são as variáveis independentes do modelo, representadas pelas características dos dados a serem classificados e baseadas em qual classificação será feita, por exemplo, consumo de álcool e cigarro, pressão sanguínea,

frequência da compra, estado civil, características das imagens de um satélite, registros geológicos específicos, direção e velocidade do vento, temporada, localização da ocorrência de um fenômeno, etc.

- **Conjunto de dados de treinamento:** consiste no conjunto de dados contendo valores dos dois componentes anteriores, e é utilizado para “treinar” o modelo, para que reconheça a classe apropriada, baseado nos preditores disponíveis. Exemplos deste componente são: grupos de pacientes que tiveram infartos, grupos de clientes de um supermercado, bancos de dados contendo imagens de um monitoramento telescópico e acompanhamento de objetos astronômicos, bancos de dados em pesquisas de terremotos e bancos de dados em pesquisas de furacões.
- **Conjunto de dados de teste:** contendo os novos dados que serão classificados por um modelo e a precisão da classificação que, desta maneira, pode ser avaliada.

Em algumas aplicações, o usuário está mais interessado em prever alguns valores ausentes em seus dados, em vez de descobrir classes de objetos. Isto ocorre sobretudo quando os valores que faltam são numéricos. Neste caso, a tarefa de mineração é denominada Predição.

### Regressão (ou Estimativa)

Segundo Harrison (1998) [39], estimativa (ou regressão) é usada para definir um valor para alguma variável contínua desconhecida. Ela lida com resultados contínuos, enquanto que a classificação lida com resultados discretos, podendo ser usada para executar uma tarefa de classificação, convencendo-se que diferentes faixas (intervalos) de valores contínuos correspondem a diferentes classes. Para Fayyad *et al.* (1996) [27], "regressão é aprender uma função que mapeia um item de dado para uma variável de predição real estimada".

### Clusterização (ou Segmentação)

De acordo com Azevedo e Santos (2015) [6, apud Han e Kamber (2006) [38]], clusterização consiste em agrupar um conjunto de objetos, sejam estes físicos ou abstratos, em um grupo de objetos de acordo com sua semelhança. Segundo os autores, "um *cluster* é definido como um conjunto de objetos de dados que possuem semelhanças entre si e que possuem características distintas em relação a outros conjuntos de objetos, não possuindo classes previamente definidas".

Dantas e Couto (2014) [17] enfatizam "que é importante não confundir o processo de Classificação com o processo de Clusterização, pois no primeiro estamos lidando com um objeto que recebe um rótulo, pertencendo a uma classe em particular, enquanto no segundo um conjunto de objetos inteiro é particionado em subgrupos bem definidos".

## Análise de *Outliers*

Para Amo (2004) [4], um banco de dados pode conter dados que não apresentam o comportamento geral da maioria. "Estes dados são denominados *outliers* (exceções). Muitos métodos de mineração descartam estes *outliers* como sendo ruído indesejado. Entretanto, em algumas aplicações, tais como detecção de fraudes, estes eventos raros podem ser mais interessantes do que eventos que ocorrem regularmente".

Azevedo e Santos (2015) [6, apud Han e Kamber (2006) [38]], descrevem que

"no processo de detecção de anomalias, são constituídos modelos de comportamento normal para a rede, que são usados para comparar e identificar novos padrões que se afastam dos perfis pré-definidos. Como vantagem, a detecção de anomalias permite identificar mais claramente ruídos que até então não foram observados, porém apresenta como desvantagem a alta porcentagem de detecção de falsos positivos (por exemplo, uma pessoa saudável é classificada como doente)."

### 4.2.1.2 Técnicas (ou Métodos) de Mineração de Dados

Diferentes métodos servem para diferentes propósitos, cada método oferece suas vantagens e suas desvantagens.

Tradicionalmente, de acordo com Han e Kamber (2006) [38] e Fayyad *et al.* (1996) [27], os métodos de mineração de dados são divididos em aprendizado supervisionado (preditivo) e não-supervisionado (descritivo). Camilo e Silva (2009) [11] apontam que a "diferença entre os métodos de aprendizado supervisionados e não-supervisionados reside no fato de que os métodos não-supervisionados não precisam de uma pré-categorização para os registros, ou seja, não é necessário um atributo alvo".

Côrtes *et al.* (2002) [18] diferencia os dois tipos de métodos da seguinte forma:

- A **Análise Descritiva** representa a área da investigação nos dados que busca tanto descrever fatos relevantes, não-triviais e desconhecidos dos usuários, como analisar a base de dados, principalmente pelo seu aspecto de qualidade, para validar todo o processo da mineração e seus resultados, ou seja, o conhecimento encontrado. Podemos subdividi-la em *Análise Prévia* e *Descobrimto*.
  - *Análise Prévia* é o processo de analisar uma base de dados com o objetivo de identificar anomalias ou resultados raros que possam influenciar os resultados da mineração de dados.
  - *Descobrimto* é o processo de examinar uma base de dados com o objetivo de encontrar padrões escondidos, sem que necessariamente exista uma ideia ou hipótese clara previamente estabelecida.

Para facilitar a aplicabilidade dos processos de mineração de dados, podemos especializar tanto a *Análise Prévia* quanto o *Descobrimto* em outras tarefas, conforme a Figura 4.6 a seguir.

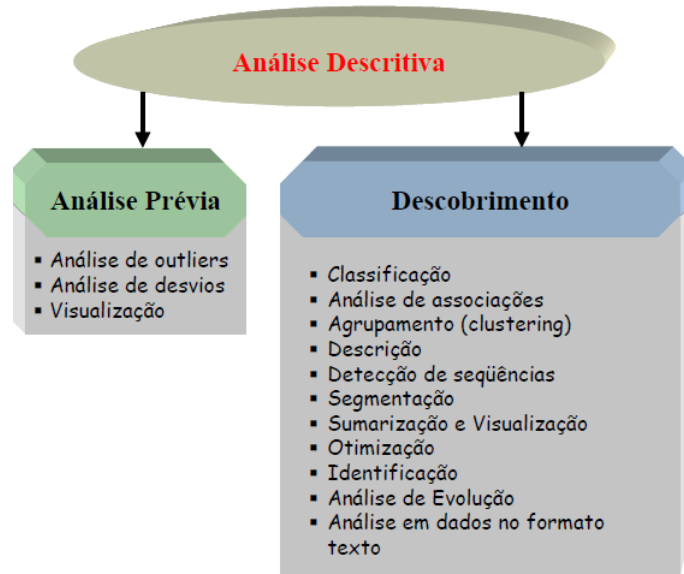


Figura 4.6: Tarefas da Análise Prévia e Descobrimto [18]

- A **Análise de Prognóstico** representa a área da investigação nos dados que busca inferir resultados a partir dos padrões encontrados na análise descritiva, ou seja prognosticar o comportamento de um novo conjunto de dados. Podemos subdividi-la em *Classificação*, *Estimação* e *Predição*.
  - *Estimação* é o processo de predizer algum valor, baseado num padrão já conhecido.
  - *Predição* é o processo de predizer um comportamento futuro, baseado em vários valores.
  - *Classificação* consiste em examinar uma certa característica nos dados e atribuir uma classe previamente definida.

Segundo Córtes *et al.* (2002) [18], a **Análise de Prognóstico** não é dividida em tarefas conforme a **Análise Descritiva**, uma vez que "ao nível de funcionalidade as definições esgotam com bastante clareza a sua aplicação".

#### 4.2.2 Ferramentas para Mineração de Dados

De acordo com Camilo e Silva (2009) [11], "diversas ferramentas foram desenvolvidas no intuito de tornar a aplicação da Mineração de Dados uma tarefa menos técnica, e com isto possibilitar que profissionais de outras áreas possam fazer usa dela". Neste sentido, o mercado de ferramentas, proprietárias e *Open Source*<sup>3</sup>, de mineração de dados tem se tornado bastante atraente.

As ferramentas *Open Source* vem apresentando diversos recursos e funcionalidades, superando, muitas vezes, seus concorrentes pagos com maior flexibilidade devido a possibilidade total de customização em alguns casos, robustez de soluções e bibliotecas, e,

<sup>3</sup>*Open Source* é o termo utilizado para designar softwares de código aberto.

claro, o preço de implementação.

A seguir, serão brevemente descritos alguns desses softwares abertos.

- **R:** O Projeto R (ou somente “R”) é um *software* de estatística que contém diversos *packages* (pacotes com diversas funções estatísticas, matemáticas e econométricas) e vem em um ritmo de crescimento muito grande, em especial nos últimos 3 anos. Tem uma variedade muito grande de recursos estatístico-computacionais e tem como vantagem a sua difusão, e por ter uma linguagem de programação que tem uma curva de aprendizado média, o que permite que estatísticos, economistas, médicos, e cientistas da computação façam interações sob um mesmo código, além de possuir uma bibliografia invejável em termos de livros e materiais disponíveis na *web*. A desvantagem é que não há informações sobre as operações de *background* e como o *software* utiliza os recursos de máquina disponíveis, como o funcionamento em relação a termos computacionais e até mesmo formas de customização para rodar em ambientes com alto poder computacional, além de ter uma interface de integração e pré-processamento de dados pobre em recursos de transformação.<sup>4</sup>
- **RapidMiner:** O RapidMiner é uma suíte de mineração de dados que vem se consolidando há algum tempo, e, antes do R se tornar o *software Open Source* mais popular em análise de dados, era o *software* de mineração de dados mais utilizado. O RapidMiner tem como principais diferenciais a parte de pré-processamento de dados, a baixa curva de aprendizado para início dos trabalhos em mineração de dados, além de produzir gráficos de ótima qualidade. O RapidMiner tem como desvantagem o “travamento” de diversos recursos que são liberados somente com a versão paga, não ter uma comunidade tão ativa e possuir pouca documentação (livros) dirigida para o seu uso prático.<sup>5</sup>
- **Rattle:** O Rattle<sup>6</sup> é uma nova suíte de mineração de dados que tem como característica principal incorporar a robustez das bibliotecas e soluções de mineração de dados do R, conta com uma interface gráfica muito agradável que possui uma excelente usabilidade, além de possuir uma comunidade que vem crescendo muito forte. Por ser uma ferramenta nova, ainda conta com uma literatura escassa.

Além das ferramentas supracitadas, outra amplamente utilizada no processo de mineração de dados é o **Weka**, a ser abordada em mais detalhes na próxima seção. O **Weka** será a ferramenta base para o processo de mineração de dados a ser desenvolvido no presente trabalho.

---

<sup>4</sup><https://www.r-project.org/about.html>

<sup>5</sup><https://rapidminer.com>

<sup>6</sup><http://rattle.togaware.com/>



### 4.2.3 Software Weka

O Weka, uma das ferramentas *Open Source* mais difundidas atualmente, é uma suíte de mineração de dados desenvolvida pela Universidade de Waikato, localizada na Nova Zelândia. Weka é uma abreviação para *Waikato Environment for Knowledge Analysis*, (em tradução livre, Ambiente Waikato para Análise do Conhecimento). O software é composto por algoritmos de aprendizagem de máquina e uma coleção de recursos que realizam o pré-processamento, regressão, classificação, clusterização, aplicação de regras de visualização dos dados e apresentação de resultados. Esses algoritmos podem ser aplicados diretamente da ferramenta, ou utilizados por programas Java, pois a ferramenta é desenvolvida utilizando a linguagem de programação Java e sua distribuição segue os termos da GNU (*General Public Licence* ou Licença Pública Geral) [49].

Segundo Azevedo e Santos (2015) [6], o Weka apresenta uma variedade de recursos e ferramentas com amplo suporte a todas as etapas do processo experimental de mineração de dados, que vai desde a preparação dos dados de entrada, análise estatística de esquemas de aprendizagem até a visualização dos dados e apresentação dos resultados. A ferramenta possui uma interface amigável ao usuário, permitindo que o mesmo tenha facilidade em trabalhar com a grande variedade de algoritmos de treinamento e ferramentas de pré-processamento disponibilizadas pelo *software*, além de possibilitar a integração direta com bancos de dados, o que permite ao usuário obter os dados direto da base e salvá-los em formato adequado para uso posterior no Weka. Outra vantagem é que a ferramenta é de distribuição livre e multiplataforma, funcionando em diferentes sistemas operacionais, como Windows, Linux e Mac OS. A Figura 4.7 apresenta a tela inicial do Weka.

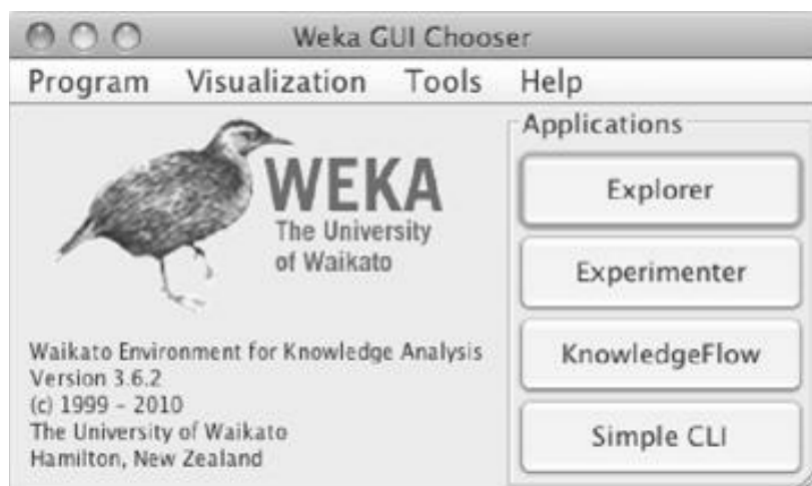


Figura 4.7: Captura de tela inicial do Weka [70]

Witten *et al.* (2005) [70] descreve o uso do Weka a partir do uso de suas três interfaces gráficas: o *Explorer*, o *Experimenter* e o *KnowledgeFlow*.

A interface *Explorer*, mostrada na Figura 4.8, oferece ao usuário a possibilidade de acesso às opções existentes na barra de menu. Azevedo e Santos (2015) [6, apud Witten *et al.* (2005) [70]] apontam que a interface "possibilita ao usuário carregar os dados a serem utilizados e a verificar os resultados gerados pelos algoritmos de mineração. Po-



rém, uma das desvantagens do modo *Explorer* é que todo o conjunto de dados utilizado é mantido em memória, limitando-se a problemas de pequeno e médio porte".

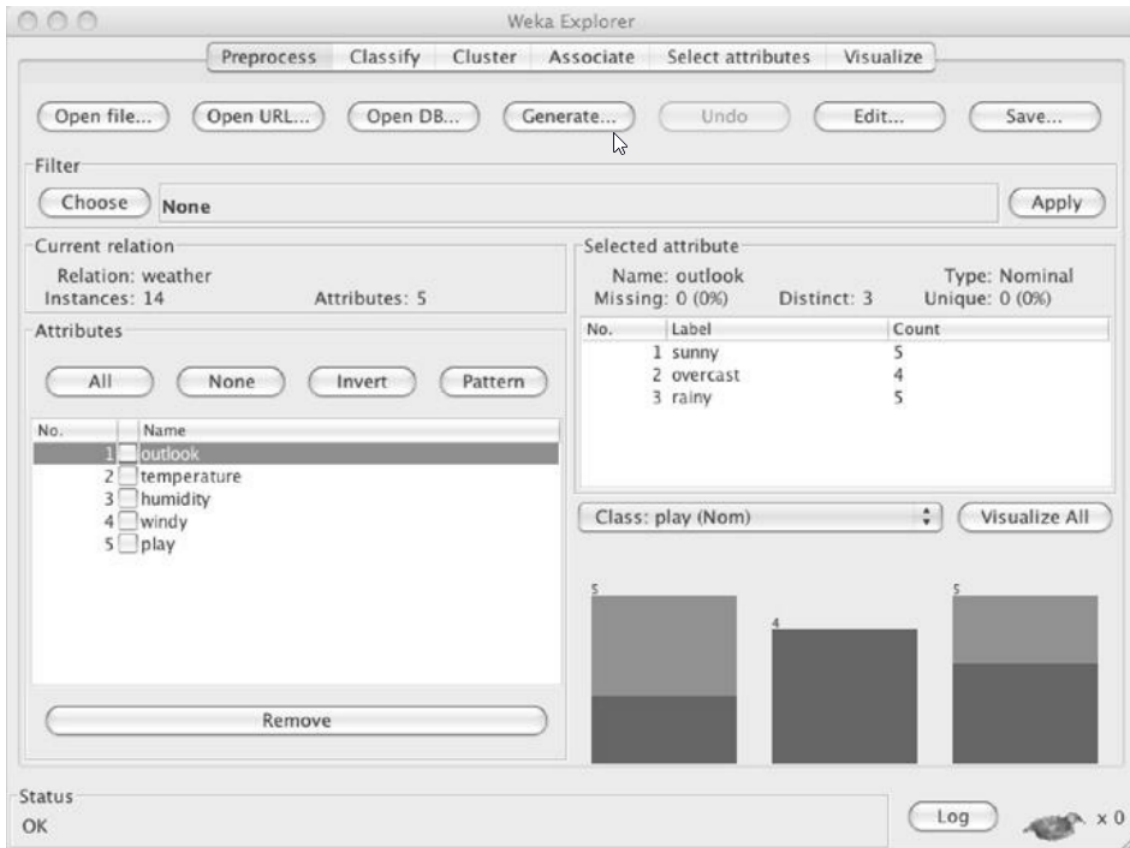


Figura 4.8: Captura de tela da interface *Explorer* do Weka [70]

A interface *KnowledgeFlow*, mostrada na Figura 4.9, permite a criação de configurações para o processamento de fluxo de dados. Pela interface gráfica do *KnowledgeFlow*, os usuários selecionam os componentes Weka a partir de uma barra de ferramentas, colocando-os em uma tela *layout* e conectando-os em um gráfico direcionado que processa e analisa os dados. Através da interface *KnowledgeFlow*, também é possível a criação e execução de configurações para o processamento de dados de *streaming*, que o *Explorer* não faz. O *KnowledgeFlow* fornece uma alternativa para o *Explorer* para aqueles que gostam de pensar em termos de como os dados fluem através do sistema [70].

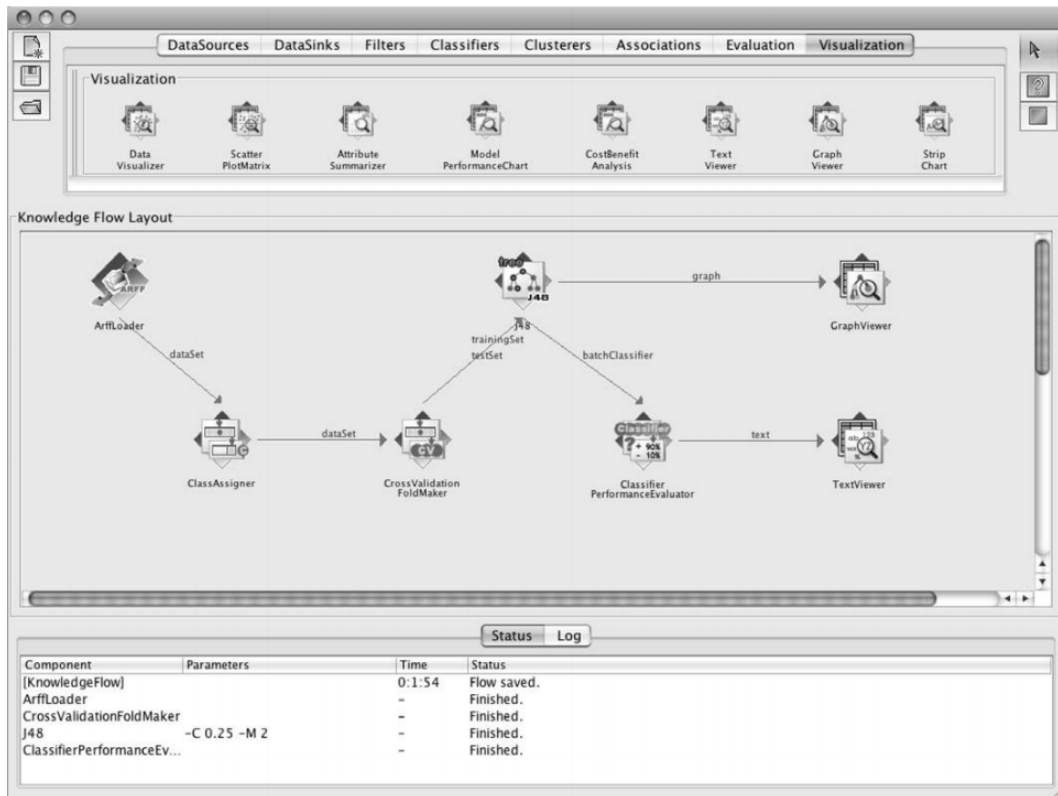


Figura 4.9: Captura de tela da interface *KnowledgeFlow* do Weka [70]

Os ambientes *Explorer* e *KnowledgeFlow* ajudam o usuário a determinar o quão bem sistemas de aprendizagem máquina executam certos conjuntos de dados. Entretanto, trabalhos com cunho investigativo envolvem experimentos mais substanciais - normalmente executando vários sistemas de aprendizagem em diferentes conjuntos de dados, muitas vezes com vários parâmetros de configurações - e essas interfaces não são adequadas para tal [70]. A opção *Experimenter*, mostrada na Figura 4.10, segundo Azevedo e Santos (2015) [6], tem por objetivo facilitar a identificação de quais métodos e parâmetros nas técnicas de classificação e regressão são mais adequados para determinado problema. A interface foi desenvolvida com o intuito de facilitar a comparação de várias técnicas de aprendizado, tornando mais fácil a execução de classificadores e filtros com diferentes definições de parâmetros sobre um conjunto de dados, a coleta de estatísticas de desempenho e a execução e automatização de testes significativos.



Figura 4.10: Captura de tela da interface *Experimenter* do Weka [70]

Além das três interfaces apresentadas anteriormente, a interface *Simple CLI*, é uma outra possibilidade de utilização do Weka. A opção *Simple CLI* apresenta dicas de como utilizar o Weka por linha de comando (via Terminal no Linux ou *Prompt* de Comando no Windows) e permite ao usuário informar os comandos a serem utilizados na mesma janela.

O Weka foi escolhido como ferramenta para esse trabalho por ser uma ferramenta mais completa e por ter uma curva de aprendizagem maior.

# Capítulo 5

## Metodologia

Para atender os objetivos propostos no presente trabalho, fez-se necessário seguir um conjunto de passos e adotar alguns procedimentos metodológicos. Este capítulo tem como objetivo mostrar cada um desses itens. A Seção 5.1 apresenta uma visão geral da metodologia aplicada de forma a descrever os passos e atividades utilizadas para alcançar os objetivos desse trabalho.

### 5.1 Visão Geral

A metodologia de pesquisa realizada neste trabalho está representada, de forma macro, no fluxograma ilustrado na Figura 5.1

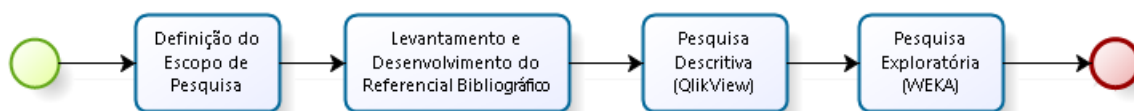


Figura 5.1: Fluxograma da Metodologia de Pesquisa

As seções subsequentes explicam com maior grandeza de detalhe como foi a execução das pesquisas descritiva e exploratória.

#### 5.1.1 Pesquisa Descritiva

Conforme demonstrado no fluxograma da Figura 5.1, uma das atividades deste trabalho constitui-se da condução de uma pesquisa descritiva baseada nos dados de votação de proposições da Câmara dos Deputados.

##### 5.1.1.1 Objetivo

O objetivo da presente fase constitui-se da análise de dados oficiais de votação de proposições da Câmara dos Deputados, com apoio da ferramenta de *Business Intelligence* QlikView. A Pesquisa Descritiva, no âmbito deste trabalho, tem o propósito de identificar

e analisar métricas e indicadores que auxiliem na descoberta de padrões nos dados de votações e orientação do partido, obtidos através do *Web Service* da Câmara.

### 5.1.1.2 Definição do Escopo de Pesquisa

Para definição do escopo da pesquisa faz-se necessária a definição do escopo dos dados utilizados, bem como a definição do escopo das ferramentas escolhidas para realização da pesquisa descritiva.

#### 5.1.1.2.1 Escopo de Dados

Como descrito na Seção 2.2 do Capítulo 2, o Poder Legislativo é exercido pelo Congresso Nacional que, por sua vez, adota um sistema bicameral, de tal modo que a função legislativa em âmbito federal é exercida por duas câmaras distintas: o Senado Federal e a Câmara dos Deputados, sendo a última composta pelos representantes do povo. Desta forma, tendo em vista o objetivo de disponibilizar acesso à informação útil para a população, definiu-se como escopo de pesquisa os dados no âmbito da Câmara dos Deputados.

Para que os deputados exerçam sua função legisladora, os mesmos necessitam de instrumentos pelos quais possam apresentar projetos de novas leis ou até mesmo projetos de alteração das leis existentes. Nesse contexto, surgem as Proposições Legislativas descritas na Seção 2.3.1 do Capítulo 2.

Com foco na relevância e no interesse da população, foram escolhidos os seguintes tipos de proposição que estão ligadas diretamente à alteração das normas federais:

- Projeto de Lei Ordinária (PL)
- Proposta de Emenda à Constituição (PEC)
- Medida Provisória (MPV)

As demais proposições foram excluídas do escopo por terem caráter simbólico e por não despertar o devido interesse da população.

Um vez apresentados os projetos pelos deputados, faz-se necessária a aprovação dos mesmos por meio de votação. Essa aprovação pode ser feita de diversas maneiras dependendo do tipo de proposição, tendo em vista que cada tipo de proposição percorre um caminho diferente de acordo com o que se propõe, conforme explicitado na Seção 2.3 do Capítulo 2. Sendo assim, para as proposições citadas acima, escolheu-se como escopo as votações ocorridas em *plenário* da Câmara dos Deputados, uma vez que a votação em plenário possui caráter finalístico (dentro da Câmara), ou seja, a partir da votação em plenário os projetos de proposição são encaminhados para apreciação do Senado Federal ou do Poder Executivo.

Com o intuito de trazer dados de votações atuais, restringiu-se o escopo de proposições votadas durante o período da atual legislatura<sup>1</sup>, ou seja, compreende o período de 1º de

---

<sup>1</sup>Compreende o período de quatro anos de execução das atividades pelo Congresso Nacional

Fevereiro de 2015 a 31 de Janeiro de 2018. Entretanto, como a atual legislatura ainda não foi encerrada, têm-se o escopo de dados até o dia da extração dos mesmos, dia 26/07/2016.

Desta forma, temos o escopo de dados da pesquisa restrito às votações, em plenário da Câmara, de *PLs*, *PECs* e *MPVs* que ocorreram no período da atual legislatura até a data de 26/07/2016.

#### **5.1.1.2.2 Escopo de Ferramentas**

Conforme explicitado na Seção 4.1.2 do Capítulo 4, existem várias ferramentas de *BI* e de análise de dados disponíveis no mercado, dentre elas as que se destacam no Quadrante Mágico do Gartner como *Leaders*. Para que uma ferramenta se encaixe no quadrante de *Leaders* a mesma precisa ter uma boa capacidade de execução e abrangência de visão. Desta forma, as ferramentas que se enquadram nesse perfil são encaradas como maduras e com grande potencial de inovação.

Entretanto, não basta apenas ter grande potencial de execução, precisa também ser ágil. Desta forma, conforme a Seção 4.1.2 do Capítulo 4 destaca, as plataformas de *BI in-memory* trazem um enorme ganho de tempo no desenvolvimento das aplicações, além de permitir uma análise em 'tempo real' dos dados.

Nesse contexto, avaliando o curto espaço de tempo para o desenvolvimento da aplicação, aliado às demais vantagens especificadas na Seção 4.1.2 e também por estar no quadrante de *Leaders* do Quadrante Mágico do Gartner, definiu-se a ferramenta *QlikView* como a ferramenta a ser utilizada para a Pesquisa Descritiva.

#### **5.1.1.3 Planejamento do Estudo**

O planejamento de estudo da Pesquisa Descritiva foi feito com base no fluxograma representado abaixo na Figura 5.2.

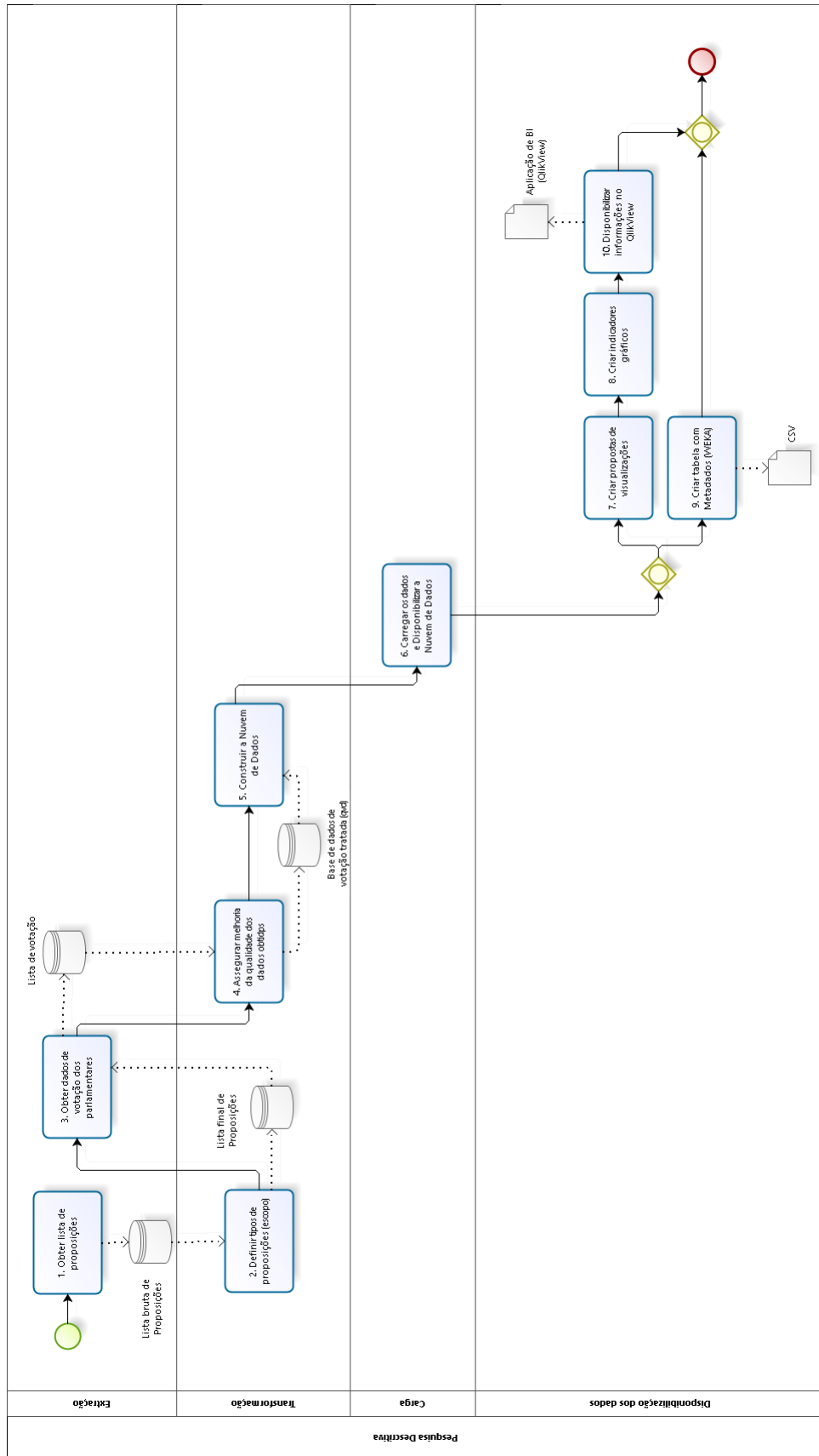


Figura 5.2: Fluxograma do planejamento de estudo da Pesquisa Descritiva

Conforme fluxograma acima, o planejamento do projeto foi inspirado nos processos de ETL (Extração, Transformação e Carga) e é composto por quatro fases, descritas mais detalhadamente abaixo.

#### 5.1.1.3.1 Fase de Extração

A presente fase tem como objetivo a extração dos dados do *Web Service* da Câmara dos Deputados para que os mesmos sejam utilizados na Pesquisa Descritiva.

##### Atividades

- **Obter lista de proposições**

Através do link de *Web Service* [ListarProposicoesVotadasEmPlenario](#)<sup>2</sup> têm-se a listagem de **todas** as proposições que foram votadas em plenário da Câmara para os anos de 2015 e 2016. Para que a extração dos dados ocorra, necessita-se da inserção de parâmetros, tais como: o ano da votação da proposição e o tipo da proposição, sendo o segundo parâmetro optativo. O *script* de busca dos dados no *Web Service* da Câmara encontra-se disponível no Anexo I

**Entrada:** -

**Saída:** Lista bruta das proposições votadas

- **Obter dados de votação**

O link de *Web Service* [ObterVotacaoProposicao](#)<sup>3</sup> retorna os votos dos deputados e a orientação dos partidos para uma determinada proposição em votações ocorridas no plenário da Câmara dos Deputados. Para que a extração dos dados ocorra, necessita-se da inserção de parâmetros, tais como: tipo, o número e o ano da proposição. Nesse contexto, o *Web Service* retorna as votações de **apenas uma** proposição por vez. Desta forma, para automatizar a busca pelos dados, fez-se necessária a criação de um *script* iterativo dentro da ferramenta QlikView (vide Anexo II) para percorrer a lista final de proposições votadas, a fim de obter os dados de votação.

Os dados obtidos através do *Web Service* se dividem em:

- Resumo da votação: contém as informações da votação, como: código da sessão, objeto de votação, resumo dos votos, etc.;
- Orientação partidária: orientação de voto do partido;

---

<sup>2</sup>disponível em: <http://www2.camara.leg.br/transparencia/dados-abertos/dados-abertos-legislativo/webservices/proposicoes-1/ProposicoesVotadasEmPlenario>

<sup>3</sup>disponível em: <http://www2.camara.leg.br/transparencia/dados-abertos/dados-abertos-legislativo/webservices/proposicoes-1/obtervotacao proposicao>



- Votação nominal dos parlamentares: voto de cada parlamentar presente na sessão.

**Entrada:** Lista final de proposições

**Saída:** Lista de votações

### 5.1.1.3.2 Fase de Transformação

A presente fase tem como objetivo o tratamento dos dados obtidos na Fase de Extração. Todas as atividades feitas nessa fase são realizadas pela ferramenta QlikView.

#### Atividades

- **Definir tipo de proposições (escopo)**

Esta atividade visa aplicar o escopo definido na Seção 5.1.2.3 por meio de transformações nos dados, feitas no módulo de ETL da ferramenta QlikView. Desta forma, a partir da lista bruta de proposições serão aplicados os filtros determinados no escopo através de alterações feitas no *script* de carga dos dados.

**Entrada:** Lista bruta das proposições votadas

**Saída:** Lista final de proposições

- **Assegurar melhoria da qualidade dos dados**

Nesta fase, são feitos os tratamentos de forma a garantir a qualidade e a confiabilidade dos dados importados. Alguns desses tratamentos visam corrigir os erros nos nomes dos deputados, a retirada dos caracteres especiais e o tratamento das chaves de ligação entre as tabelas de forma a garantir a construção de um modelo de dados associativo íntegro.

Tratamentos específicos foram realizados no campo **Partido**, que encontra-se inserido no escopo de dados de Orientação Partidária. O campo supracitado continha em seus dados blocos partidários e não apenas partidos, como esperado. Um bloco partidário consiste na união de dois ou mais partidos, equivale a um só comando, como se fosse um único partido. Desta forma, para efeito de pesquisa, os blocos partidários foram separados em partidos atribuindo-se a mesma orientação do bloco para os partidos que o compunham. Sendo assim, caso um bloco fosse composto pelo PT, PMDB e PSOL e a orientação do bloco fosse 'sim', após a separação do bloco, os partidos, separadamente, receberiam a mesma orientação, ou seja, 'sim'. A tabela 'DE-PARA' utilizada para realizar as alterações supracitadas, encontra-se

disponível no Anexo **III**.

Além do tratamento feito no campo **Partido**, foi necessário tratar os caracteres especiais existentes nos campos **Orientação**, **Voto Deputado** e **Nome do Deputado** no intuito de viabilizar o desenvolvimento da pesquisa exploratória através da ferramenta de Data Mining, WEKA. O script utilizado para fazer os referidos tratamentos encontra-se disponível no Anexo **IV**.

**Entrada:** Lista bruta das proposições votadas, Tabela 'DE-PARA' de blocos partidários

**Saída:** Base de dados de votação tratada (QVD<sup>4</sup>)

- **Construir a nuvem de dados**

Realizadas as transformações, constrói-se um modelo de dados conceitual, baseado na hierarquia das informações disponibilizadas e nas regras de negócio.

**Entrada:** Base de dados de votação tratada (QVD)

**Saída:** Modelo de dados

### 5.1.1.3.3 Fase de Carga

A presente fase tem como intuito realizar a carga dos dados na aplicação responsável pela viabilização das análises e na exportação dos dados necessários.

#### Atividades

- **Carregar os dados e disponibilizar a Nuvem de Dados**

Esta atividade realiza a carga dos dados de votação, de orientação dos partidos e do resumo da seção no QlikView, disponibilizando o modelo de dados associados abaixo, para construção das métricas e indicadores necessárias para responder as hipóteses levantadas para o trabalho.

---

<sup>4</sup>Extensão de dados proprietário do QlikView

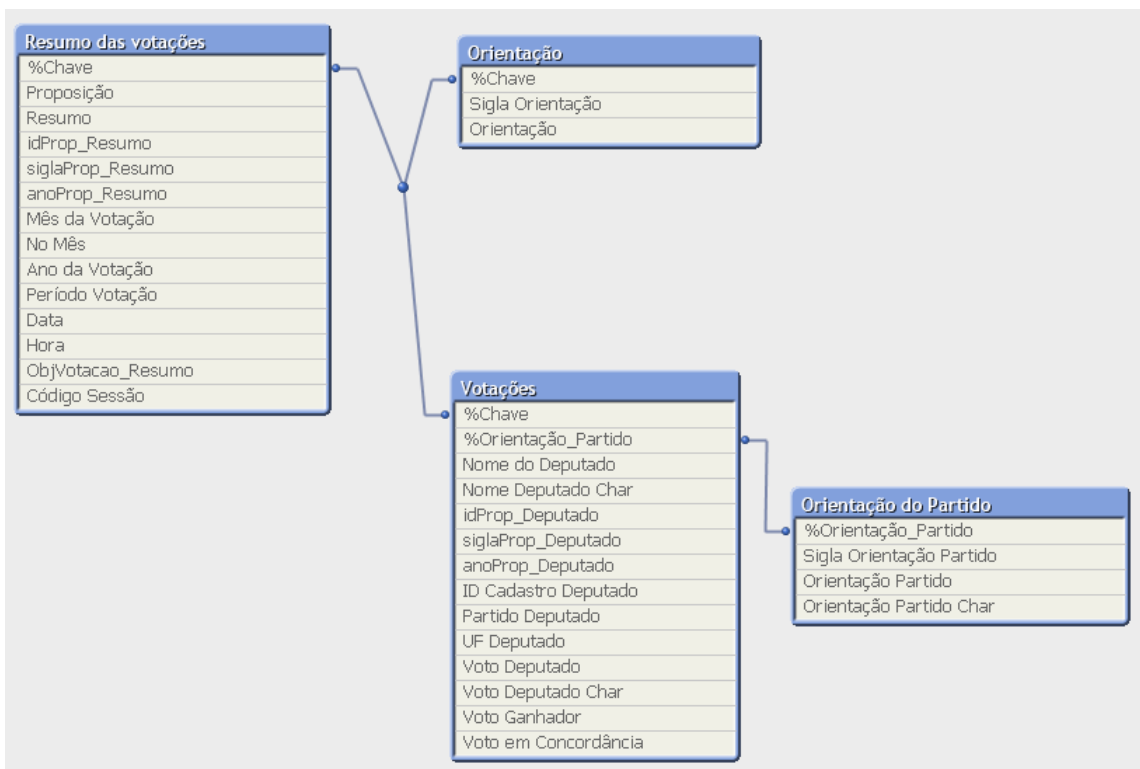


Figura 5.3: Modelo de dados associativo gerado pelo QlikView

Conforme a Figura 5.3 demonstra, o modelo de dados é composto por quatro tabelas. A presença de duas tabelas para Orientação se deve ao fato da necessidade da apresentação da orientação do partido em granularidades diferentes. A tabela 'Orientação' foi criada para estabelecer o vínculo da proposição à orientação partidária e a tabela 'Orientação do Partido' foi criada para estabelecer o vínculo entre a orientação nominal do partido e a votação da proposição.

Para auxiliar no entendimento dos campos, foi criado um Dicionário de Dados presente no Anexo V.

**Entrada:** Modelo de Dados

**Saída:** Aplicação OLAP

#### 5.1.1.3.4 Fase de Disponibilização dos dados

Esta fase tem como objetivo disponibilizar, de forma visual e intuitiva, a informação obtida através da análise dos dados.

##### Atividades

- **Criar propostas de visualizações**

Esta atividade visa a criação de propostas de visualização das informações através da criação de uma lista de métricas e indicadores que irão compor a aplicação.

**Entrada:** Aplicação OLAP

**Saída:** Lista de métricas e indicadores

- **Criar indicadores gráficos**

Com base na lista de métricas e indicadores criada na atividade anterior, geram-se, graficamente, os indicadores na ferramenta.

**Entrada:** Lista de métricas e indicadores

**Saída:** Objetos gráficos

- **Criar tabela com metadados (WEKA)**

Esta atividade visa a exportação de uma tabela de metadados, separados por vírgula (csv), com foco na orientação dos partidos, para servir de entrada no processo de Pesquisa Exploratória.

**Entrada:** Dados carregados

**Saída:** Tabela excel (csv)

- **Disponibilizar as informações no QlikView**

O objetivo desta atividade é disponibilizar as métricas e indicadores criados em uma *dashboard*, viabilizando a análise conjunta dos dados.

**Entrada:** Objetos gráficos

**Saída:** *Dashboard*

## 5.1.2 Pesquisa Exploratória

Em continuidade aos processos dispostos no fluxograma da Figura 7.1, a última atividade deste trabalho constitui na condução de uma Pesquisa Exploratória em cima dos dados tratados e gerados pela Pesquisa Descritiva.

### 5.1.2.1 Objetivo

A presente fase tem como objetivo familiarizar-se com fenômenos atrelados à votação de proposições da Câmara a partir do uso de algoritmos de *Data Mining*.

### 5.1.2.2 Hipóteses

A pesquisa exploratória foi realizada em cima das seguintes hipóteses:

- **Hipótese 1:** Tendo em vista o grande número de parlamentares eleitos do Partido do Movimento Democrático Brasileiro (PMDB), formando a maior bancada da Casa e a recorrência desse fato de maneira recorrente ao longo das legislaturas, é sugerida a hipótese de que o PMDB exerce influência sobre os demais partidos e sobre a decisão dos votos na Câmara dos Deputados.
- **Hipótese 2:** Uma vez que a Câmara dos deputados representa o povo, conforme aponta a Seção 2.2.1, os deputados devem representar os interesses do povo nas votações. Entretanto, é sabido que a bancada dos estados interfere, de alguma forma, nas votações ocorridas na Casa. Desta forma, é sugerida a hipótese de que a influência das bancadas estaduais podem ser comparadas às influências do partidos.
- **Hipótese 3:** As bancadas dos partidos na Câmara são formadas pelos deputados eleitos pelos mesmos partidos. Nesse contexto, espera-se que os deputados de um mesmo partido votem de forma semelhante. Sendo assim, sugere-se como uma terceira hipótese, a possibilidade da descoberta do partido do deputado a partir de seu voto nas sessões de plenário.

### 5.1.2.3 Definição do Escopo de Pesquisa

Para definição do escopo da pesquisa exploratória faz-se necessária a definição do escopo dos dados utilizados, bem como a definição do escopo das ferramentas escolhidas para realização da mesma.

#### 5.1.2.3.1 Escopo de Dados

O escopo de dados definido para essa pesquisa coincide com o escopo de dados definido para a Pesquisa Descritiva, conforme descrito na Seção 5.1.1.2.1.

### 5.1.2.3.2 Escopo de Ferramentas

Tendo em vista a realização de uma pesquisa exploratória, têm-se a necessidade da utilização de uma ferramenta que viabilize a análise e a detecção de padrões desconhecidos nos dados através de hipóteses previamente levantadas conforme estudo anterior do negócio em questão. Nesse contexto se destacam as ferramentas de *Data Mining*.

No mercado, atualmente, existem diversas ferramentas de mineração de dados, mas o Weka se destaca por ser um *software* completo e *Open Source*, além de possuir os algoritmos necessários para fazer a mineração de dados relacionadas as hipóteses, como algoritmos de classificação, clusterização e associação, conforme detalha a Seção 4.2.3. Desta forma, por atender aos requisitos de análise para uma pesquisa exploratória, definiu-se o Weka como sendo a ferramenta encarregada por gerar os resultados dessa etapa.

### 5.1.2.4 Planejamento do Estudo

O planejamento de estudo da Pesquisa Exploratória foi feito com base no fluxograma representado abaixo na Figura 5.4.

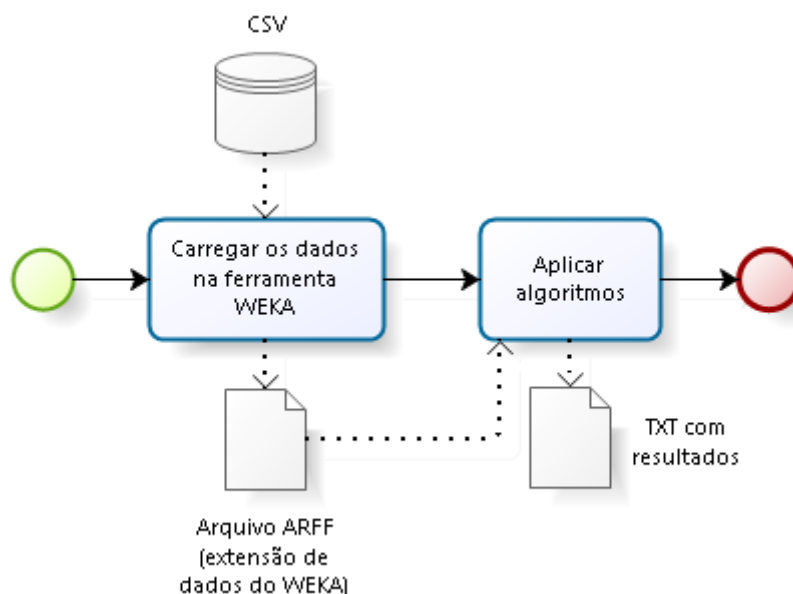


Figura 5.4: Fluxograma do planejamento de estudo da Pesquisa Exploratória

Conforme fluxograma demonstrado acima, o planejamento da pesquisa exploratória é composto por duas atividades descritas abaixo com maior riqueza de detalhes.

#### 5.1.2.4.1 Carregar os dados na ferramenta Weka

A presente atividade tem como objetivo a importação dos dados na ferramenta de mineração de dados, Weka. Os dados utilizados para a pesquisa exploratória foram tratados durante a fase de Pesquisa Descritiva para que, os mesmos, tivessem a estrutura exigida

pelo Weka para que fossem feitas as análises. A atividade dentro da fase de Pesquisa Descritiva que gera os dados para essa fase, encontra-se na Seção 5.1.1.3.4.

**Entrada:** CSV gerado pela etapa de Pesquisa Descritiva

**Saída:** Arquivo ARFF

#### 5.1.2.4.2 Aplicar algoritmos

Conforme o nome indica, o objetivo desta fase é aplicar, nos dados, os seguintes algoritmos:

- **Algoritmo de clusterização *Simple K-means***

Tendo em vista a primeira hipótese levantada para a pesquisa, foi escolhido um algoritmo de clusterização simples, rápido e capaz de separar os partidos através de suas respectivas orientações de votos para as votações presentes nos dados. Segundo Witten et. al [70] o *Simple K-means* é um algoritmo de clusterização iterativo baseado em distância. O algoritmo, através do número de *clusters* determinado pelo número  $K$ , determina pontos na região dos dados, chamados de centroides, de forma a separar grupos de deputados e partidos que votam parecido entre si e diferente dos outros grupos de deputados.

Espera-se que ao final da execução do algoritmo, na análise dos grupos formados, os partidos estejam, em sua maioria, ao redor do partido PMDB, além de indicar quais partidos tem votações semelhantes entre si.

- **Parâmetro utilizado:**

- \*  $K=3$  (descoberto empiricamente)  
Gera 3 *clusters*

- **Algoritmo de classificação *k-NN***

Para responder à indagação levantada na hipótese dois, identificou-se o algoritmo de classificação  $k$ -NN (acrônimo de *K-Nearest Neighbor*) como viabilizador do resultado esperado através da classificação dos deputados em relação às orientações do partidos e em relação às bancadas estaduais, de forma a verificar se é possível descobrir o partido e o estado de um deputado baseado em suas votações.

- **Parâmetros utilizados:**

- \*  $N=5$  (descoberto empiricamente)  
Tenta classificar o deputado baseado nos 5 deputados mais parecidos
    - \* Curva ROC  
Melhor métrica para um problema com várias classes

- **Algoritmo de classificação *JRip***

O algoritmo JRip, por ser capaz de criar um conjunto de regras e entregar os resultados que possuíram maior taxa de acerto, espera-se que tenha a possibilidade de prever o partido dos deputados através de seus votos, como propõe a hipótese três.

– **Parâmetro utilizado:**

\* *default*

A presente fase tem as seguintes entradas e saídas em comum para todos os algoritmos a serem aplicados:

**Entrada:** Arquivo ARFF

**Saída:** Arquivos TXT com os resultados obtidos



# Capítulo 6

## Resultados

Este Capítulo apresenta os resultados obtidos por meio das pesquisas descritiva e exploratória, seguindo a metodologia apresentada no Capítulo 5 para os dados de votações de PL, PEC e MPV acontecidas em plenário da Câmara nos anos de 2015 e 2016. A Seção 6.1 apresenta os resultados observados para votações dentro da pesquisa descritiva através da ferramenta QlikView e a Seção 6.2 apresenta os resultados obtidos na pesquisa exploratória por meio do processo de mineração de dados realizado no Weka.

### 6.1 Pesquisa Descritiva

Nas Seções 6.1.1 e 6.1.2 são apresentados os resultados descobertos pela análise dos dados feita no QlikView, através do acompanhamento de métricas e indicadores levantados para essa pesquisa. A Seção 6.2.1 traz uma visão geral dos resultados obtidos nessa fase do trabalho.

#### 6.1.1 Visão Geral

Os dados de votação extraídos do *Web Service* da Câmara, conforme escopo levantado, geraram uma massa de 74 proposições distintas votadas, sendo 43% das proposições do tipo MPV, 33% do tipo PL e 24% do tipo PEC, conforme a Figura 6.1 demonstra.

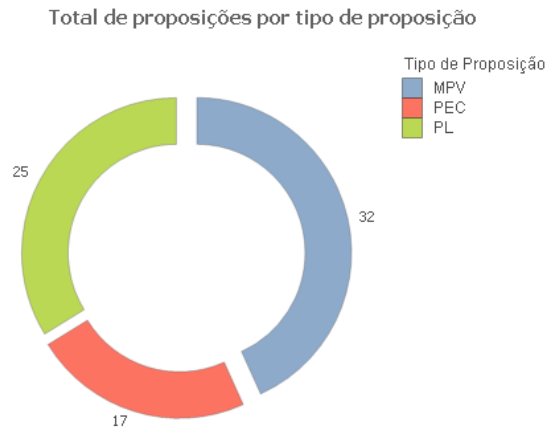


Figura 6.1: Distribuição das proposições por tipo de proposição

A tempestividade dos votos para as proposições varia muito. Por mais que as proposições do tipo MPV tenham um saldo maior de número de proposições votadas, conforme apresenta a Figura 6.1 acima, as MPVs só começaram a ser votadas em Abril de 2015. Já as proposições do tipo PL pararam de ser votadas em novembro de 2015, mas mesmo assim, ainda são mais presentes nas votações que as PECs, que obtiveram votações em todo o escopo de tempo.

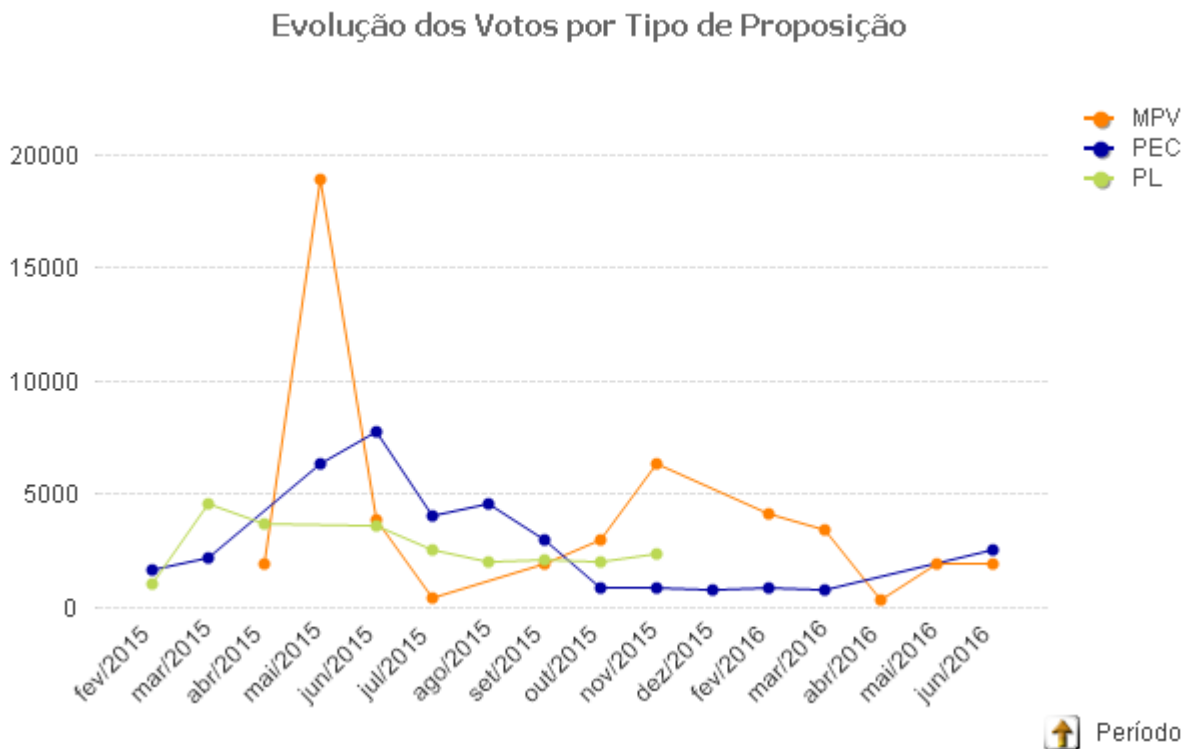


Figura 6.2: Evolução dos votos por proposição

O pico observado no mês de maio de 2015, se deu por conta das votações da MPV

665/2014, que propõe que o desempregado comprove mais tempo de trabalho para ter direito ao seguro-desemprego. Por ser um tema polêmico e de interesse social, econômico e político, observou-se maior número de parlamentares presentes, resultando em um maior número de votos.

Dentre as sessões em plenário levantadas, verificou-se a presença de 591 parlamentares votantes, sendo 78 deles, suplentes, uma vez que o número de deputados eleitos totaliza 513. Desses parlamentares, a maior bancada presente em votações foi a bancada do PMDB com 88 representantes presentes nas votações, seguidos pelo PT (73) e o PSDB (61), conforme demonstra a tabela 6.1.

Tabela 6.1: Lista de partidos presentes nas votações

Partidos	Bancada
PMDB	88
PT	73
PSDB	61
PP	55
PSD	51
PR	48
PSB	41
DEM	33
PDT	32
PTB	27
PMB	25
PRB	24
SD	21
PROS	19
PSC	19
PCDOB	15
PTN	15
PPS	13
PV	12
PHS	9
PSOL	7
PTDOB	6
PRP	5
REDE	5
PEN	4
PSL	4
PMN	3
PRTB	2
PSDC	2
PTC	2

Já a distribuição dos votos foi caracterizada, em sua maioria, por votos a favor da causa em discussão (54%), seguidos pelos votos negativos (44%), conforme aponta a Figura 6.3. Nota-se a presença do voto Art. 17 como o último no ranking, uma vez que esse tipo de voto é restrito apenas ao Presidente da Câmara, possibilitando que, o mesmo, vote sem ser identificado.

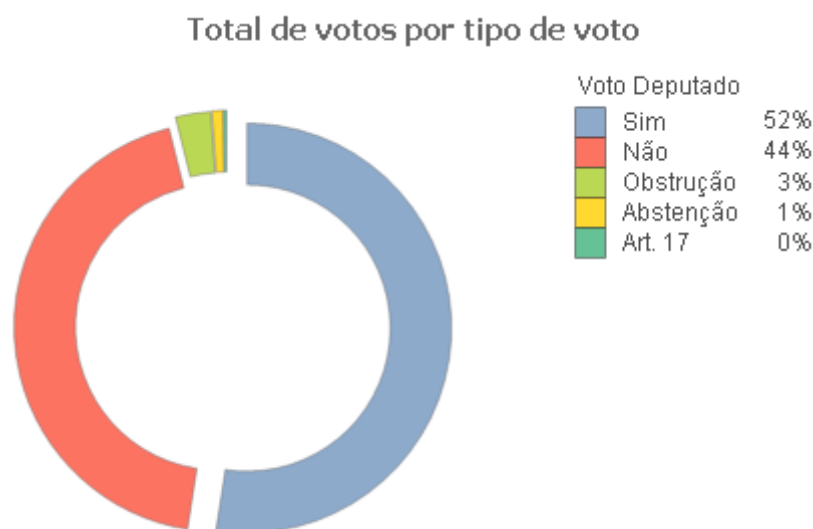


Figura 6.3: Distribuição dos votos

Dentre os parlamentares votantes, o deputado que mais esteve presente nas votações foi o deputado do Solidariedade de Espírito Santo, Carlos Manato, presente em 266 sessões e logo após o deputado Tiririca do PR de São Paulo.

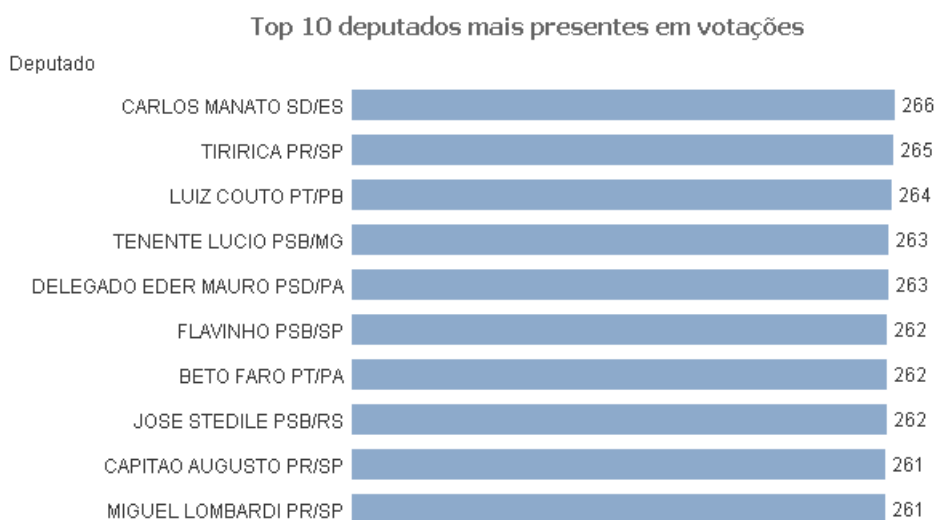


Figura 6.4: Os 10 parlamentares com maior presença em votações

Conforme observado na figura 6.3 acima, apenas 1% dos parlamentares se abstiveram da votação. Desse 1%, o parlamentar que mais se absteve da votação foi o deputado Weliton Prado do PT de Minas Gerais, seguido pelo Julião Amin do PDT do Maranhão,

apontado pela Figura 6.5. Concomitantemente, o partido que mais se absteve da votação foi o PT com quase 2% dos seus votos e o que menos se absteve foi o PSOL com 0.09%, como demonstra a Figura 6.6.

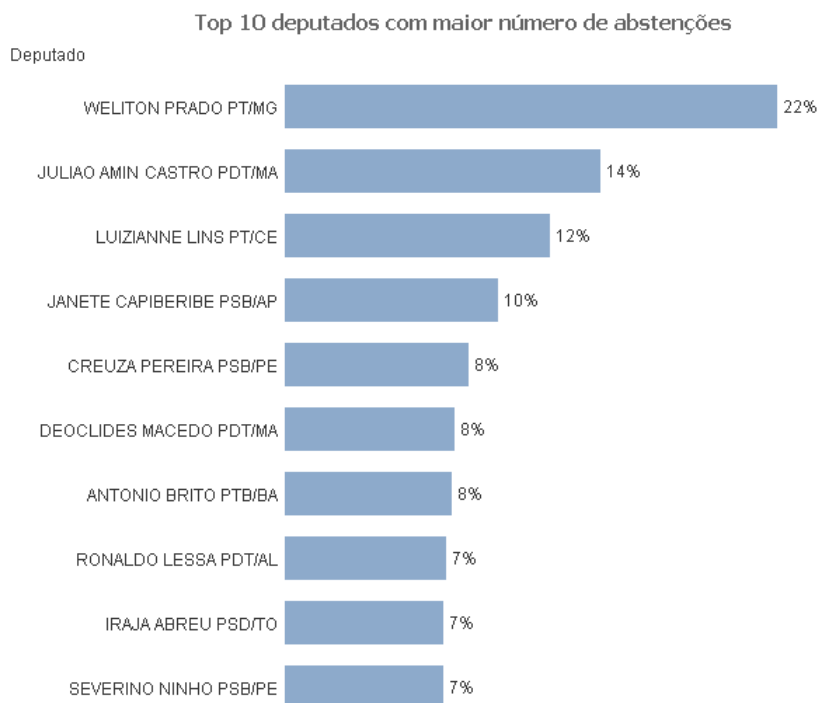


Figura 6.5: Os 10 parlamentares com maior número de abstenções

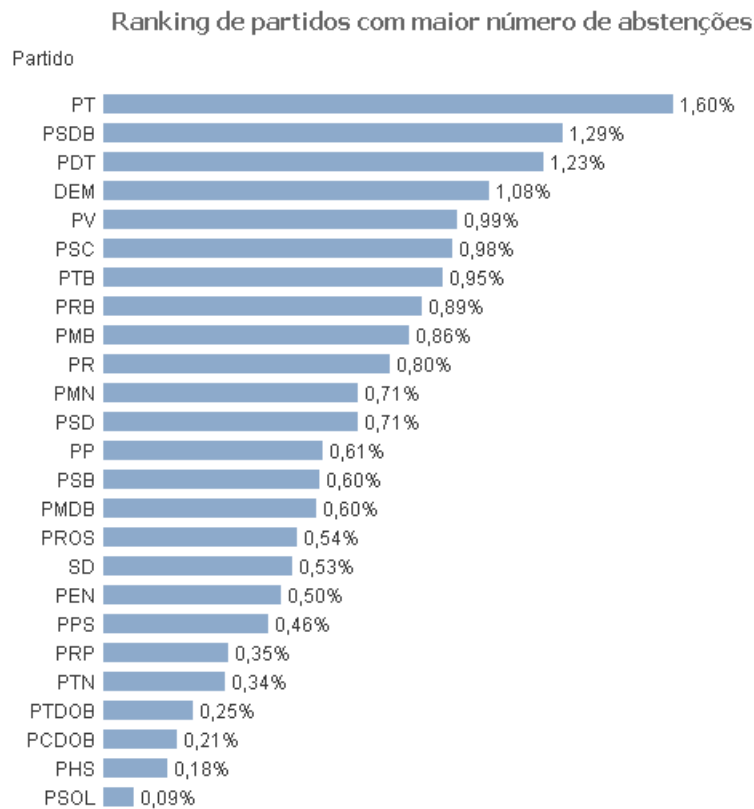


Figura 6.6: Os 10 partidos com maior número de abstenções

### 6.1.2 Análise dos Votos sob a Orientação dos Partidos

Os partidos orientam seus parlamentares por meio de orientações partidárias, não sendo elas obrigatórias, ou seja, um partido pode ou não determinar uma orientação específica para cada votação. Tal fato pode ser evidenciado na tabela 6.2, onde observa-se que a soma das diferentes orientações diferem por partido, podendo ser justificadas pela ausência do partido em uma votação específica.

Um partido pode orientar seus parlamentares a votarem a favor, contra ou até para que se abstenham ou obstruam a votação para cada sessão específica em plenário. A abstenção e a obstrução indicam, respectivamente, a liberação do voto pelo partido e impedimento do prosseguimento dos trabalhos com intuito de ganhar tempo dentro de uma ação política.

Dentre as orientações dadas, destaca-se o PSDB por ter o maior número de obstruções como orientação, conforme ilustra a tabela 6.2.

Tabela 6.2: Distribuição das orientações por partido

Partidos	Sim	Não	Abstenção	Liberado	Obstrução
DEM	133	120	1	17	24
PCDOB	124	125	0	0	21
PDT	127	116	0	16	11
PEN	632	505	0	98	2
PHS	696	561	0	112	5
PMB	7	10	0	2	0
PMDB	640	535	0	98	2
PMN	934	878	0	144	9
PP	712	585	0	112	7
PPS	143	122	0	1	23
PR	181	161	0	8	9
PRB	985	927	0	149	10
PROS	181	138	1	17	13
PRP	895	838	0	135	9
PRTB	657	705	0	135	0
PSB	138	119	0	15	17
PSC	712	581	0	112	7
PSD	182	146	0	22	8
<b>PSDB</b>	<b>141</b>	<b>114</b>	<b>1</b>	<b>8</b>	<b>25</b>
PSDC	888	805	0	144	9
PSL	717	632	0	149	12
PSOL	68	74	0	0	24
PT	124	119	1	5	21
PTB	712	585	0	112	7
PTC	898	827	0	140	9
PTDOB	969	845	0	149	12
PTN	1001	945	0	149	12
PV	124	97	0	51	4
REDE	26	26	0	0	12
SD	146	111	1	8	18

Nesse contexto, pôde ser avaliado o grau de fidelidade dos deputados aos seus partidos, ou seja, quão alinhados os parlamentares estão com a orientação de seus respectivos partidos. O gráfico presente na Figura 6.7 ranqueia os 10 deputados que possuem maior índice de fidelidade, classificados pelo número de votos de forma descendente.

Tal classificação se fez necessária tendo em vista que suplentes podem votar, e através das análises observou-se que suplentes votam menos que os deputados. Desta forma, caso o suplente tenha ido votar apenas uma vez e seu voto tenha sido de acordo com a orientação do partido, ele obteria uma taxa de 100% de fidelidade. Desse modo, esse resultado não é expressivo para a pesquisa.

Sendo assim, a partir da análise do gráfico abaixo representado na Figura 6.7, podemos observar que o deputado Carlos Manato do SD do Espírito Santo possui uma taxa de fidelidade menor, comparado ao deputado Tiririca do PR de São Paulo, pois o deputado do Solidariedade teve um número maior de votos.

Entretanto, é importante ressaltar o índice de fidelidade do deputado Tiririca. O percentual apresentado de 99,62% de fidelidade ao partido é bem expressivo pelo fato de o deputado ser o segundo parlamentar mais presente nas votações.

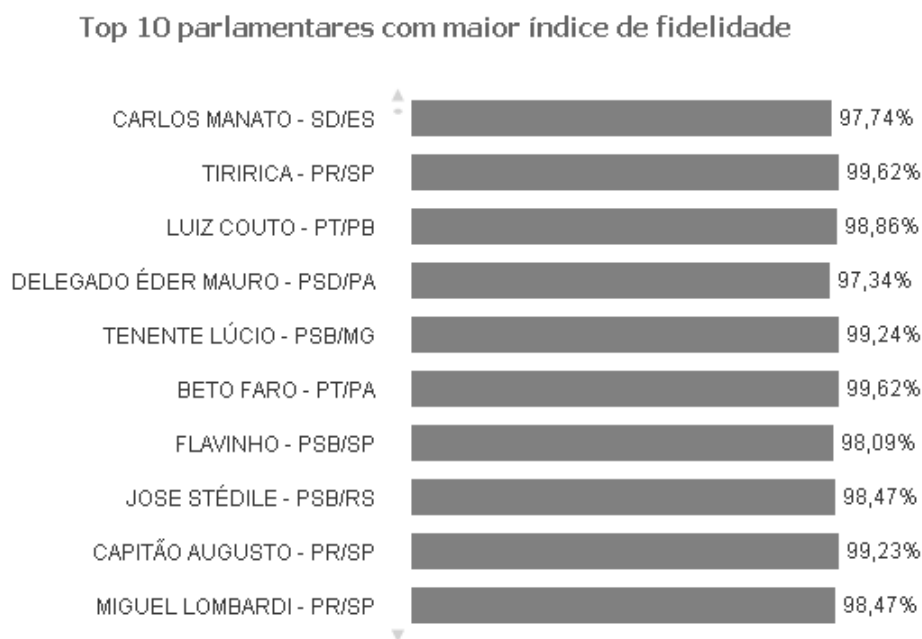


Figura 6.7: Ranking de deputados fiéis à orientação do partido

Assim como os deputados, foram analisados os índices de fidelidade por partido e por Unidade da Federação (UF). O cálculo utilizado apresenta a porcentagem de fidelidade dos partidos, calculada através da razão entre os total de votos em acordo com a orientação do partido e o total de votos obtidos em sessões de votações em plenário.



A Figura 6.8 apresenta uma comparação entre a porcentagem de fidelidade dos partidos e a quantidade de votos em acordo com a orientação do partido.

Nesse sentido, no ranking demonstrado pela figura abaixo, a bancada do PEN se apresentou como mais fiel ao partido com 99,26% dos votos dos parlamentares de acordo com a orientação.

Entretanto, é importante analisar que não basta ter um número grande de votos em acordo com a orientação, é preciso que a maioria de seus votos estejam de acordo com a orientação. Os partidos PMDB e PT se destacam nesse sentido. Desta forma, é possível identificar que o percentual de fidelidade comparado ao número de votos em concordância, mostrou que partidos com uma bancada grande podem ter maior dificuldade em manter a coesão dos votos dos parlamentares em relação à orientação do partido. As duas maiores bancadas partidárias da Casa (PT e PMDB), apresentaram um índice médio de 2% a menos que o índice do PEN, líder do ranking. Entretanto, essa diferença, por ser muito pequena, não pode afirmar que o PT e PMDB seja partidos 'inféis'.

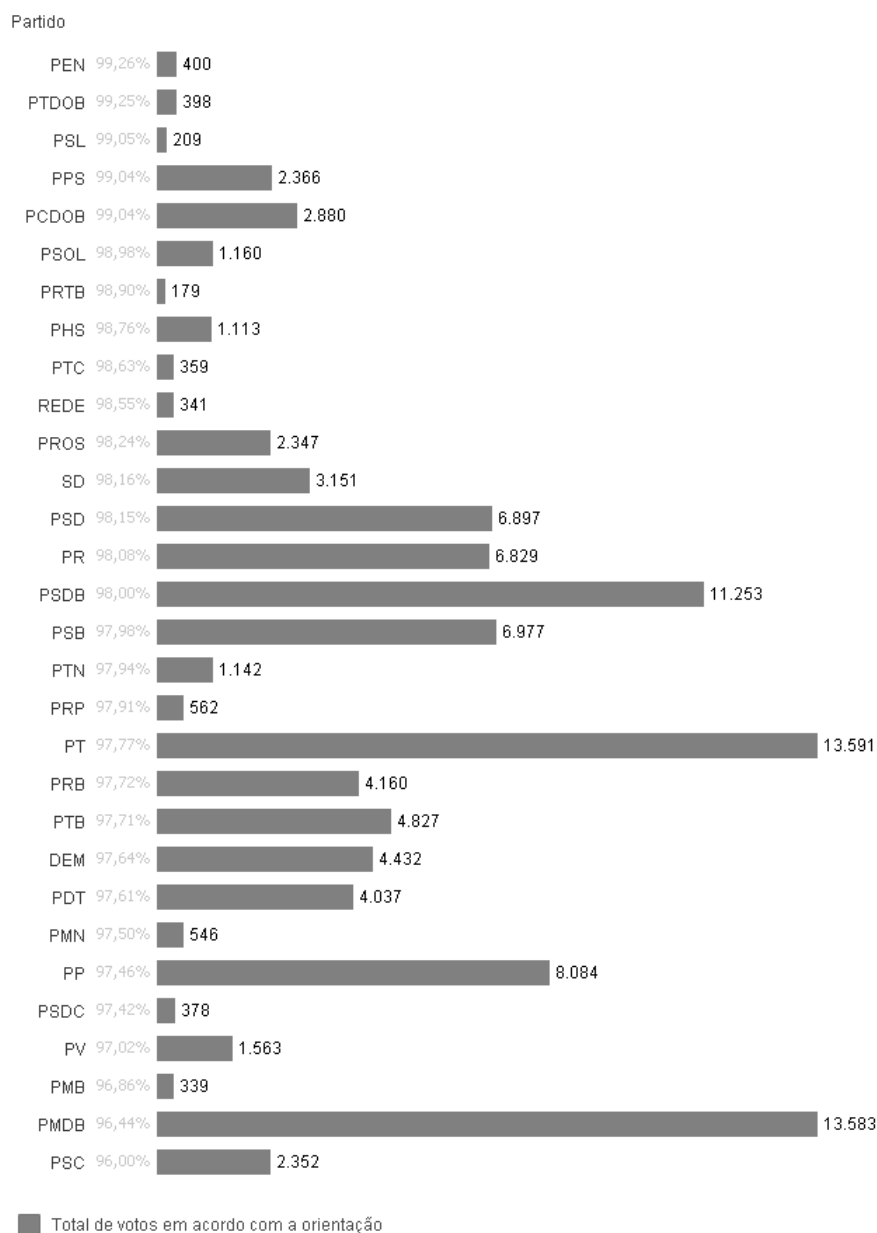


Figura 6.8: Ranking de bancadas partidárias fiéis à orientação do partido

A partir da análise de fidelidade dos parlamentares em relação às bancadas estaduais, percebe-se que o RJ encontra-se em último lugar. Tal fato ocorre em decorrência dos votos do, até então Presidente da Câmara, Eduardo Cunha do PMDB do Rio de Janeiro. Tendo em vista que 100% das vezes, o Presidente invocou o Art. 17 não permitindo que seu voto fosse divulgado, todos os votos de Eduardo Cunha irão contra a orientação do partido, levando o RJ para a última posição do ranking.

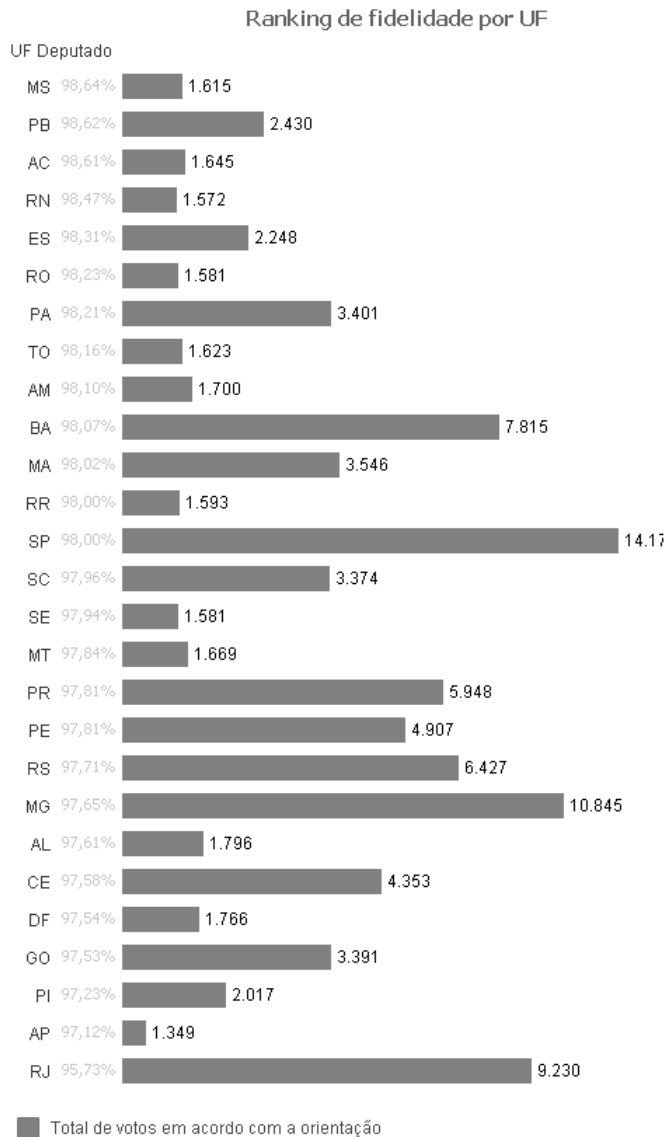


Figura 6.9: Ranking de UFs com parlamentares fiéis à orientação do partido

### 6.1.3 Análise dos Votos em Comparação com a Maioria

A Seção 6.1.2 apresentou os resultados obtidos por meio da análise da concordância dos votos dos parlamentares com a orientação dos partidos. Entretanto, também foi analisado a correlação entre os votos de acordo com a orientação e os votos com a maioria.

Os resultados mostraram que apenas 4 deputados votaram mais vezes de acordo com a votação da maioria do que de acordo com a orientação do partido. Tal fato leva à conclusão que a orientação do partido tem peso maior nas decisões dos parlamentares na hora da votação. A Figura 6.10 mostra o ranking dos 10 deputados com maior fidelidade ao partido em comparação com os votos que teve com a maioria.

### Top 10 Deputados em Concordância com a Maioria

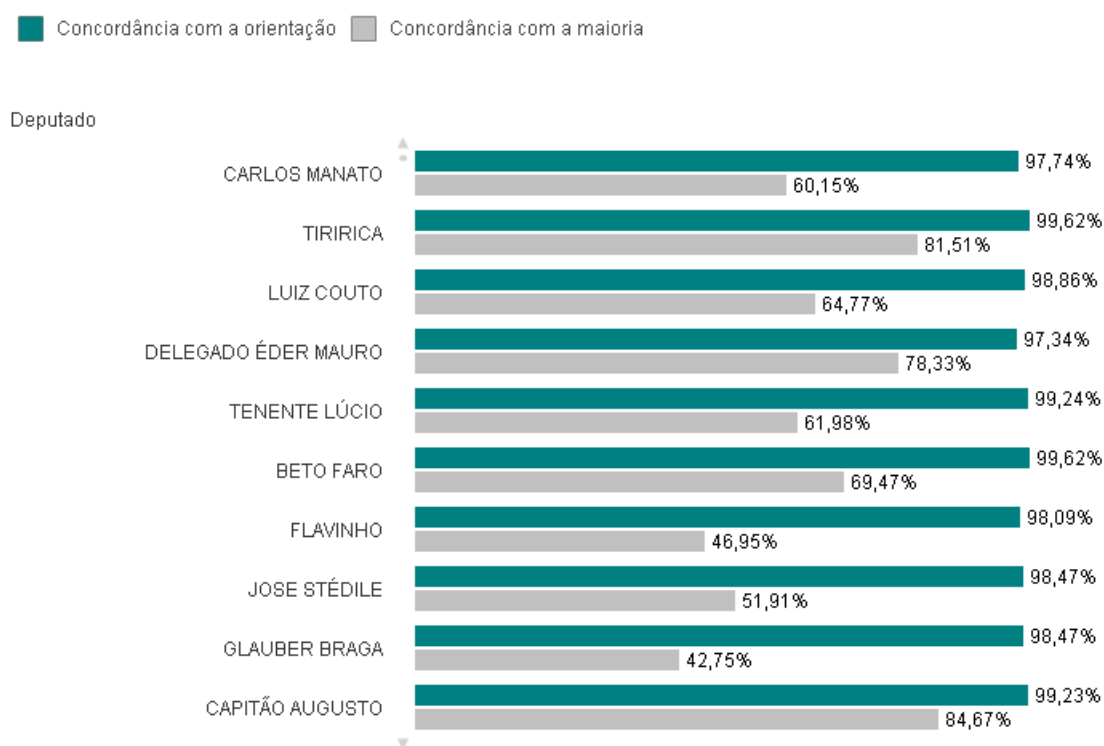


Figura 6.10: Os 10 Deputados em Concordância com a Maioria

Entretanto, observa-se que o número de deputados que votaram mais vezes de acordo com o maioria do que de acordo com a orientação do partido foi irrisório na comparação entre partidos. Como a Figura 6.11 mostra, nenhum partido obteve o número de votações em concordância com a maioria, maior que a quantidade de votos em concordância com a orientação de seus partidos.

### Ranking de votos em concordância por Partido

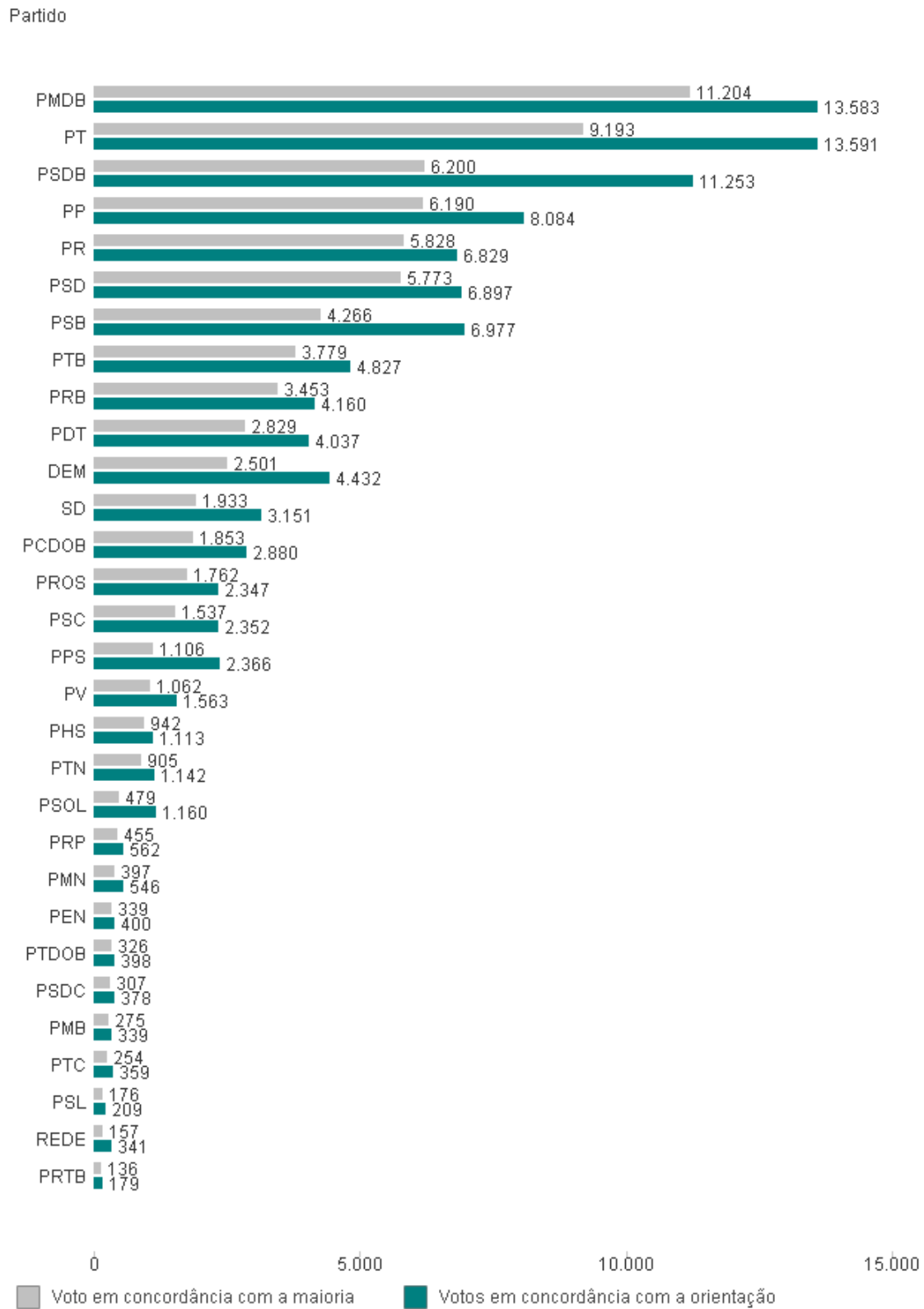


Figura 6.11: Os 10 Partidos em Concordância com a Maioria

### 6.1.4 Análise dos Votos em Apoio às Maiores Bancadas

Tendo em vista o momento político vivenciado no Brasil, é importante visualizar a predominância de influências nas votações ocorridas em plenário da Câmara dos Deputados. As duas maiores bancadas partidárias são PT e PMDB, sendo o primeiro, líder do governo.

A Figura 6.12 abaixo demonstra a evolução dos votos em apoio ao PT e ao PMDB. O cálculo utilizado se baseou na comparação dos votos de todos os deputados com a orientação do PT e do PMDB. Desta forma, a linha do PMDB mostra, na evolução do meses, quantos deputados votaram conforme a orientação do PMDB, ocorrendo da mesma forma para o PT. Os números correspondem a varias votações, podendo ocorrer no mesmo dia. O mesmo voto pode ser contabilizado para os dois partidos, caso as orientações sejam iguais.

Nesse contexto, o gráfico mostra que com a aprovação do processo de *impeachment* da Presidente Dilma Rousseff ocorrido em abril de 2016, houve uma divergência no apoio entre o PMDB e PT, que até então, possuíam comportamento semelhante.

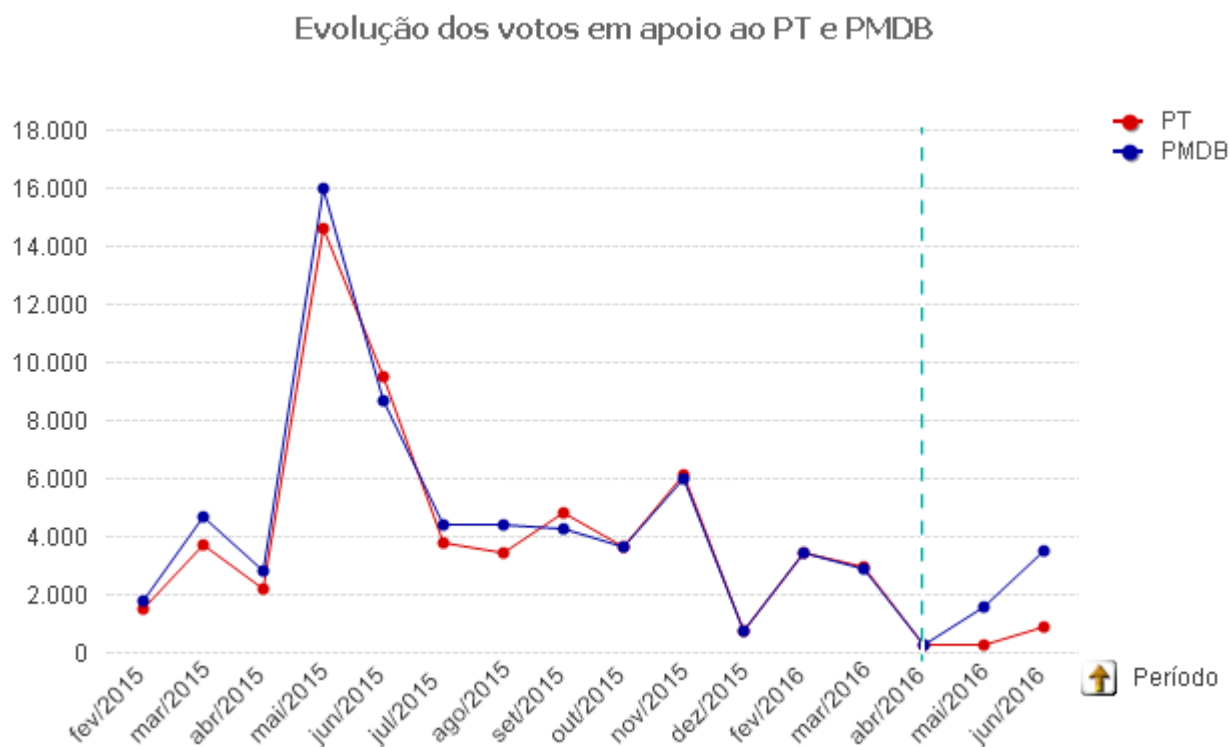


Figura 6.12: Evolução dos votos em apoio ao PT e ao PMDB

### 6.1.5 Painéis de Análise

Para viabilizar a análise heurística dos dados, sugeriu-se a criação de painéis analíticos em QlikView. Com o intuito de fornecer uma navegação intuitiva, as informações foram dispostas de forma a permitir que filtros específicos sejam selecionados, como sugere a Figura 6.13.

A Figura 6.13 mostra um resumo da votação da proposição MPV 660/2014 (que aparece selecionado em verde), na qual possui um total de 455 votos, sendo 350 deles do tipo 'sim', 103 do tipo 'não' e apenas 1 uma abstenção para o objeto de votação 'EMENDA Nº 1 DO SENADO FEDERAL AO PLV Nº 1/2015'.

A Figura 6.14 apresenta uma segunda sugestão de painel analítico com a disponibilização de indicadores que permitem que seleções sejam feitas nos filtros presentes na porção esquerda do painel ou até mesmo que seleções sejam feitas por meio de cliques nos próprios gráficos.



Análise heurística do padrão de votação dos deputados em proposições na Câmara dos Deputados entre 2015 e 2016

### Orientação Partidária

**Partido**

Abstenção  
Não  
Sim

### Votação Parlamentar

**Nome do Deputado**

UF

**Data**

**Partido Deput...**

**Proposição**

**Ano da Propo...**

**Código Sessão**

**Objeto da Votação**

ABREVIAR O REGIME JURÍDICO DO FULCRO DE 35 DO ART. 10 DA CF CONSTATADO DO ART. 10 DA PEC ADMISSIBILIDADE BY GLOBO DOS DESTAQUES SIMPES

ART. 1.º DO SUBSTITUTIVO

ART. 1.º, ART. 3.º, ART. 4.º (EXCETO 4.º), ART. 9.º, ART. 10, ART. 11, ART. 13, ...

ART. 7.º DO SUBSTITUTIVO - POSSE DO PRESIDENTE E GOVERNADORES

**Resumo**

Aprovada a Emenda nº 1 do Senado Federal. - Sim: 350; não: 103; abstenção: 1; total: 454.

Aprovada a Emenda nº 15. - Sim: 231; não: 143; total: 374.

Aprovada a Emenda nº 65. - Sim: 174; não: 166; abstenção: 1; total: 341.

Aprovada a Emenda Adjudicativa nº 1. - Sim: 193; não: 71; abstenção: 1; total: 271.

Aprovada a Emenda Adjudicativa nº 1. - Sim: 201; não: 200; abstenção: 1; total: 402.

**TOTAL DE VOTOS**

**QTD. VOTOS SIM** 350

**QTD. VOTOS NÃO** 103

**QTD. VOTOS ABSTENÇÃO** 1

Figura 6.13: TELA1



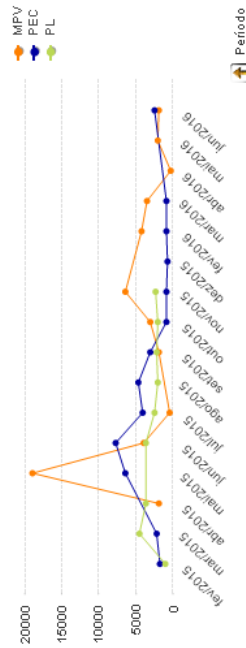


## Análise heurística do padrão de votação dos deputados em proposições na Câmara dos Deputados entre 2015 e 2016

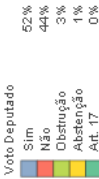
### Filtros de Pesquisa

**Partido**  **UF**   
**Proposição**   
**Ano da Proposição**

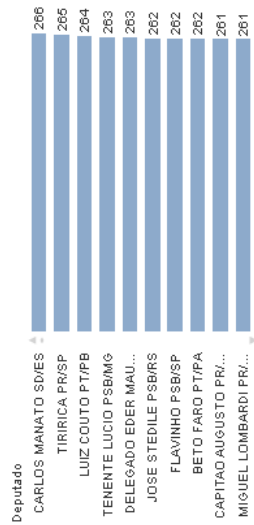
### Evolução dos Votos por Tipo de Proposição



### Total de votos por tipo de voto



### Ranking dos deputados mais presentes em votações



### Top 10 parlamentares com maior índice de fidelidade

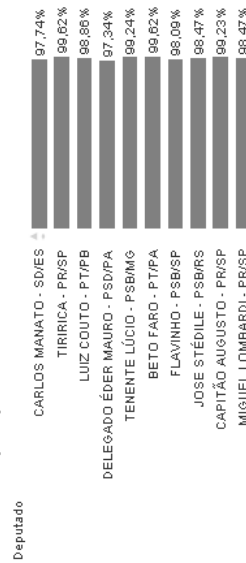


Figura 6.14: TELA2

## 6.2 Pesquisa Exploratória

A Seção 6.2.1 apresenta as regras e padrões descobertos pelo processo de mineração de dados. Foram realizados processos de mineração sobre os 266 registros de votação dos parlamentares em PECs, PLs e MPVs no plenário da Câmara para a atual legislatura.

### 6.2.1 Análise dos votos dos parlamentares

Para a análise dos votos obtidos em sessões no plenário da Câmara, foi utilizado o algoritmo de clusterização *Simple K-means* e os algoritmos de classificação *k-NN* e JRip, explicados anteriormente na seção 5.1.2.4.2.

Foi utilizando o algoritmo *Simple K-means* com três *clusters* para os votos dos parlamentares analisados. Os *clusters* descobertos foram: PT, PMDB e PSDB. Isto significa, por exemplo, que os deputados do *cluster* do PT tem votos mais semelhantes do que os deputados do *cluster* do PSDB. Esses partidos possuem um agrupamento de deputados que votam semelhantes à orientação do partido, sendo esse o motivo de o algoritmo ter gerado esses *clusters*. Nesse contexto, observou-se que o PT, mesmo sendo a bancada do governo, possui apenas 8% dos deputados agrupados no seu *cluster*. Em contrapartida, observa-se que o PMDB, além de ter a maior bancada na Câmara, possui 78% dos deputados agrupados no seu *cluster*, conforme demonstra a Tabela 6.3.

Além da análise macro dos resultados, percebemos que, dos partidos identificados nos *clusters* (PT, PSDB e PMDB), o PT se destaca por ser o único partido que não tem a maioria dos deputados no próprio *cluster*. Conforme a Tabela 6.3 demonstra, o parlamentares do PT se dividem entre o próprio PT e o PMDB, com 62 e 61 parlamentares respectivamente. Esses números, por diferirem apenas de 1 (um) parlamentar, podendo indicar uma divisão do partido. Embora o PT tenha um índice de fidelidade grande, um *cluster* pode ser descoberto se as inconsistências nos votos podem ser agrupadas.

Outra análise interessante se faz sobre a análise dos parlamentares apoiantes do PSDB. Dentre os partidos dispostos, o PPS se apresentou como o único partido com parlamentares que ficam mais próximos do *cluster* do PSDB que as outras bancadas. Sendo assim, dentre os partidos identificados nos *clusters*, o PSDB se apresenta como o que tem menos deputados no mesmo *cluster*.

Utilizando-se o mesmo algoritmo para fazer análise dos partidos, os *clusters* escolhidos foram o PEN, PCDOB e DEM. A grosso modo, os partidos escolhidos como *clusters*, pelo algoritmo, não possuem muitos deputados, pois o *Simple K-means* não tem condições de determinar a representatividade de cada *cluster*, ele apenas agrupa os dados por proximidade. Desta forma, a análise deve ser feita em cima dos partidos relevantes para o cenário. Nesse sentido, observou-se que o PT detém o maior numero de partidos com votos similares, mesmo não tendo o maior numero de parlamentares, como visto acima. Tal fato é possível uma vez que os partidos aliados ao PT são partidos de pequeno porte, ou seja, com bancadas menores, contrapondo com o PMDB que possui a segunda maior

Tabela 6.3: Resultado do algoritmo *Simple K-means*

Partidos analisados			
PT	PSDB	PMDB	Partidos apoiadores
0	18	29	DEM
4	0	43	PDT
0	0	25	PMB
1	24	35	PSB
0	0	30	PRB
0	0	10	PHS
0	11	57	PP
0	3	28	PTB
0	0	8	PTDOB
62	1	61	PT
0	0	64	PR
0	0	27	PROS
0	2	81	PSD
2	0	13	PTN
1	2	128	PMDB
0	50	43	PSDB
0	0	5	REDE
0	11	7	PPS
0	0	6	PRP
0	0	5	PSL
13	0	10	PCDOB
0	0	3	PSDC
0	0	5	PEN
0	6	23	PSC
0	3	15	PV
0	0	6	PMN
0	9	24	SD
0	0	2	PTC
3	1	3	PSOL
0	0	4	S.PART.
0	0	2	PRTB
86 (8%)	141 (14%)	802 (78%)	TOTAL

base aliada, composta por um número menor de partidos, mas com bancadas maiores, vide Tabela 6.4

Tabela 6.4: Distribuição de partidos por base aliada

Partidos analisados			
PEN	PCDOB	DEM	Partidos que compõe a base aliada
0	0	1	DEM
0	1	0	PCDOB
0	1	0	PDT
1	0	0	PEN
1	0	0	PHS
0	1	0	PMB
1	0	0	PMDB
0	1	0	PMN
1	0	0	PP
0	0	1	PPS
1	0	0	PR
0	1	0	PRB
0	1	0	PROS
0	1	0	PRP
0	1	0	PRTB
0	0	1	PSB
1	0	0	PSC
1	0	0	PSD
0	0	1	PSDB
0	1	0	PSDC
0	1	0	PSL
0	0	1	PSOL
0	1	0	PT
1	0	0	PTB
0	1	0	PTC
0	1	0	PTDOB
0	1	0	PTN
0	0	1	PV
0	1	0	REDE
0	0	1	SD
8	15	7	TOTAL

Entretanto, por mais que o PT não tenha tanta influência sobre os votos, o mesmo possui maior consistência entre seus parlamentares, como demonstra o resultado do algoritmo  $k$ -NN disposto na tabela 6.5. Nesse contexto, o PT se apresenta como o partido com o maior número de parlamentares que seguem a orientação do partido nas votações, com um total de 110 votantes em concordância. A seguir, no ranking, aparecem os partidos PMDB e PSDB com 76 e 72 votantes, respectivamente.

Tabela 6.5: Os 5 partidos com maior consistência entre os parlamentares da bancada

Partido	Nº de votantes em concordância
PT	110
PMDB	76
PSDB	72
PR	35
PSB e PP	31

A partir da análise feita em cima dos mesmos dados, mas com foco no estado onde o parlamentar foi eleito, percebeu-se que a análise dos dados foi inconsistente, uma vez que a taxa de classificações incorretas ficou em torno dos 90%. Desta forma, comparando os dados da análise anterior, que tinha como foco a determinação da consistência do partido a partir dos votos dos parlamentares, e a análise feita em cima da determinação da consistência do voto dos parlamentares mediante comparação com os estados, parece indicar que a influência do partido se sobrepõe à influência das bancadas estaduais para as proposições determinadas no escopo. Entretanto, pode ser que bancadas estaduais fortes influenciem a orientação do partido, como é o caso das bancadas do RJ e de SP.

Uma vez que os parlamentares votam de acordo com a orientação do partido, é possível determinar o parlamentar pelo voto através da aplicação do algoritmo JRip no âmbito de dados importados para o Weka. O resultado obtido aponta o PT e a REDE como os partidos mais fáceis de serem identificados através dos votos dos parlamentares de sua bancada, ambos representados pela maior área da curva ROC (sendo 1 o número ideal e 0.5 um número totalmente ao acaso [70]), com uma taxa de aproximadamente 85% de acerto. A Tabela 6.6 representa o ranking dos 5 partidos mais fáceis de serem identificados através dos votos de seus parlamentares. Este resultado consolida o resultado obtido através do algoritmo *Simple K-means* que apontou o PT como o partido mais consistente dentre os partidos atuantes na Câmara dos Deputados, por ser um *cluster* facilmente identificável.

Tabela 6.6: Os 5 partidos com maior facilidade de serem identificados através dos votos de seus parlamentares

Partido	Curva ROC de votantes em concordância
PT	0.849
PPS	0.790
PCDOB	0.753
PR	0.723
PSD	0.716

# Capítulo 7

## Considerações Finais

O presente trabalho teve como objetivo geral a análise heurística dos resultados obtidos por meio das atividades e procedimentos realizados em cima dos dados de votações de proposições em plenário da Câmara dos Deputados ocorridos em 2015 e 2016. Para tal, foram realizadas pesquisas em cima dos dados extraídos dos *Web Services* da Câmara dos Deputados. Nesse sentido, foi definida uma metodologia de pesquisa visando apoio à proposição do trabalho. A metodologia consiste em quatro fases, sendo elas: Definição do Escopo de Pesquisa, Levantamento e Desenvolvimento do Referencial Bibliográfico, Pesquisa Descritiva e Pesquisa Exploratória, conforme ilustra a Figura 7.1

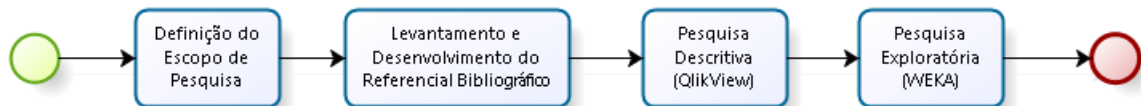


Figura 7.1: Fluxograma da Metodologia de Pesquisa

A Pesquisa Descritiva teve como objetivo o descobrimento de padrões por meio da análise heurística dos documentos oficiais da Câmara dos Deputados. Os resultados dessa etapa foram obtidos em cima da análise de métricas e indicadores, por meio da ferramenta de *Data Discovery* QlikView. A Pesquisa Exploratória tem como objetivo familiarizar-se com fenômenos atrelados à votação de proposições da Câmara a partir do uso de algoritmos de *Data Mining*.

## 7.1 Resultados

Com a execução das atividades do trabalho, diversos resultados foram obtidos. Dentre os mais significativos estão os relacionados à influência do PMDB nas votações. Os resultados obtidos por meio da Pesquisa Descritiva e da Pesquisa Exploratória, indicam a possibilidade do PMDB ser o partido com maior número de deputados na Câmara que votam de forma semelhante à sua orientação.

Os resultados dessa pesquisa indicam que, em número de votos de acordo com a orientação do partido, o PMDB fica à frente do PT que possui uma maior base aliada em termos de número de partidos.

Outro resultado importante encontrado, está relacionado com a influência das bancadas estaduais nas votações. O algoritmo de classificação não conseguiu identificar uma influência clara das bancadas estaduais sobre os partidos. Esse resultado é importante, pois avalia os parlamentares em termos de suas obrigações para com o povo, uma vez que a Câmara dos Deputados é a Casa no Congresso Nacional que defende os interesses do povo. Desta forma, os parlamentares não poderiam sobrepor os interesses do estado aos interesses do partido.

Foi identificado por meio das análises feitas no QlikView, que os parlamentares do PEN são os mais fiéis ao partido respeitando as orientações dadas para as votações.

## 7.2 Limitações e Trabalhos Futuros

Ao longo do desenvolvimento deste trabalho, algumas limitações foram encontradas, tais limitações expandem o leque de oportunidades para o desenvolvimento de trabalhos futuros. Dentre as limitações encontradas podemos citar a complexidade do cenário e a busca pelos dados no *Web Service* da Câmara dos Deputados.

O *Web Service* da Câmara impõe um limite de extração dos dados de votação restritos a apenas uma proposição por vez, tal limitação fez com que fosse criado um script iterativo para busca automatizada desses dados. Além disso, foram necessários vários tratamentos nos dados em razão da má qualidade dos dados obtidos, ocasionando maior tempo de desenvolvimento.

A complexidade do cenário estudado foi um grande fator limitante, uma vez que o conhecimento prévio a respeito do Regimento Interno da Câmara não foi suficiente. Foram necessárias várias idas ao Balcão do Serviço de Informação ao Cidadão (SIC) para o levantamento de informações, além da ida ao setor de TI da Câmara para o levantamento de dúvidas em relação ao funcionamento do *Web Service*.

Como trabalho futuro, é possível incluir todas as proposições para um estudo mais exato do comportamento dos deputados. Outra possibilidade, seria a de estudar a integração da ferramenta Weka com a ferramenta QlikView afim de fazer a análise conjunta dos

dados em uma plataforma única, sem a necessidade de comparar as informações obtidas em ferramentas diferentes.



# Referências

- [1] QlikTech International AB. *Manual de Referência do QlikView 11*. 2011. 28
- [2] QlikTech International AB. *TECHNICAL BRIEF: QLIKVIEW ARCHITECTURE AND SYSTEM RESOURCE USAGE*. 2011. 29
- [3] J. ABDALA and M. R. NASCIMENTO. Lei de acesso à informação: Um instrumento de controle social da administração pública. 18
- [4] Sandra de AMO. Técnicas de mineração de dados. *Universidade Federal de Uberlândia , Faculdade de Computação*, 2004. 32, 34
- [5] REIS E. S. ANGELONI, M. T. Business intelligence como tecnologia de suporte a definição de estratégias para melhoria da qualidade do ensino. *XXX Encontro Nacional de Pós-Graduação em Administração*, 2006. 21
- [6] L. A. AZEVEDO and Y. S. SANTOS. Mineração de dados aplicada ao estudo da evasão e desempenho dos alunos do bacharelado em ciência da computação da universidade de Brasília. *Brasília*, 2015. v, 16, 30, 31, 33, 34, 37, 39
- [7] Carlos BARBIERI. *Business Intelligence: modelagem e tecnologia*. Rio de Janeiro, 2001. v, 21, 22, 23
- [8] B. A. BLUME. *Câmara e Senado: Qual a diferença?*. 2016. Disponível em: <<http://www.politize.com.br/atualidades/camara-e-senado-qual-diferenca/>> Acesso em: 08 jun. 2016. 8
- [9] BRASIL. *Constituição da República Federativa do Brasil*. Brasília, 1988. 292 p. 1, 5, 6, 7, 8, 9, 11, 12, 13, 18
- [10] C. A. A. F. BRITTO. Os sentidos do vocábulo ‘poder’ na Constituição Brasileira. *Revista de Direito Público*, v. 15(n. 61):p. 60–64, Jan/Mar 1982. 4
- [11] Cassio O. CAMILO and João Carlos S. SILVA. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. *universidade Federal de Goiás, Instituto de Informática*, 2009. 32, 34, 35
- [12] Eduardo CANTERGI. Processo legislativo federal da lei complementar e da lei ordinária no Brasil. 2007. Disponível em: <[http://www3.pucrs.br/pucrs/files/uni/poa/direito/graduacao/tcc/tcc2/trabalhos2007\\_1/eduardo\\_cantergi.pdf](http://www3.pucrs.br/pucrs/files/uni/poa/direito/graduacao/tcc/tcc2/trabalhos2007_1/eduardo_cantergi.pdf)> Acesso em: 05 jun. 2016. 5, 6, 7, 8

- [13] O. N. P. CARDOSO and R. T. M. MACHADO. Gestão do conhecimento usando data mining: estudo de caso na universidade federal de lavras. *RAP. Revista Brasileira de Administração Pública, Rio de Janeiro*, v.42(n.3):p. 495–528, maio/jun. 2008. Disponível em: <<http://repositorio.ufla.br/jspui/handle/1/184>> Acesso em: 06 jun. 2016. 15, 16, 20, 30, 32
- [14] A. C. S. CARNEIRO and L. C. A. dos. SANTOS. *Curso de Regimento Interno*. 3 ed edition, 2014. v, 5, 6, 7, 10, 11
- [15] Manuel CASTELLS. *A era da informação: economia, sociedade e cultura*, volume v. 1. São Paulo, 2000. 16
- [16] Controladoria-Geral da União (CGU). Acesso à informação pública: Uma introdução à lei 12.527, de 18 de novembro de 2011. 2011. 1
- [17] C. COUTO, Gustavo and A. N. A. DANTAS, Marília. Utilizando mineração de dados para análise de gênero nos cursos de computação na unb. 2014. 30, 32, 33
- [18] S. S. CÔRTEZ, R. M. PORCARO, and S. LIFSCHITZ. Mineração de dados - funcionalidades, técnicas e abordagens. *PUC-RioInf.MCC10/02*, 2002. v, 30, 34, 35
- [19] Presidência da República. *LEI COMPLEMENTAR Nº 95, DE 26 DE FEVEREIRO DE 1988*. 1988. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/leis/LCP/Lcp95.htm](http://www.planalto.gov.br/ccivil_03/leis/LCP/Lcp95.htm)> Acesso em: 15 jun. 2016. 8
- [20] Presidência da República. *LEI COMPLEMENTAR Nº 78, DE 30 DE DEZEMBRO DE 1993*. 1993. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/leis/LCP/Lcp78.htm](http://www.planalto.gov.br/ccivil_03/leis/LCP/Lcp78.htm)> Acesso em: 15 jun. 2016. 6
- [21] Presidência da República. *LEI COMPLEMENTAR Nº 107, DE 26 DE ABRIL DE 2001*. 2001. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/leis/lcp/lcp107.htm](http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp107.htm)> Acesso em: 15 jun. 2016. 8
- [22] Inteligência de Negócios. *AQL X OLAP*. Disponível em: <<http://www.in1.com.br/tecnologia/aql-x-olap>> Acesso em: 02 jul. 2016. v, 29
- [23] BRASIL. Câmara dos Deputados. *Regimento Interno da Câmara dos Deputados*. 17 ed edition, 2016. Disponível em: <<http://www2.camara.leg.br/documentos-e-pesquisa/edicoes/paginas-individuais-dos-livros/regimento-interno-da-camara-dos-deputados-1>> Acesso em: 05 jun. 2016. 10
- [24] Brasília : Câmara dos Deputados. *Lei de Acesso à Informação: Cartilha de Orientação ao Cidadão*. 2012. 31 p. 19
- [25] Câmara dos Deputados. *O Papel da Câmara dos Deputados*. Disponível em: <<http://www2.camara.leg.br/a-camara/conheca>> Acesso em: 08 jun. 2016. 6
- [26] Câmara dos Deputados. *RESOLUÇÃO DA CÂMARA DOS DEPUTADOS Nº 17, DE 1989*. 1989. Disponível em: <<http://www2.camara.leg.br/legin/fed/rescad/1989/resolucaodacamaradosdeputados-17-21-setembro-1989-320110-norma-pl.html>> Acesso em: 15 jun. 2016. 9

- [27] U. FAYYAD, G. PIATETSKY-SHAPIRO, and P. SMITH. From data mining to knowledge discovery in databases. *American Association for Artificial Intelligence*, v.17(n.3), 1996. Disponível em: <<https://www.aaai.org/ojs/index.php/aimagazine/article/viewFile/1230/1131>> Acesso em: 30 mai. 2016. v, 30, 31, 33, 34
- [28] Governo Federal. *Acesso à Informação: Cumprimento da LAI*. Disponível em: <<http://www.acessoainformacao.gov.br/assuntos/conheca-seu-direito/quem-garante-o-cumprimento-da-LAI>> Acesso em: 15 jun. 2016. 19
- [29] Governo Federal. *Acesso à Informação: Principais Aspectos*. Disponível em: <<http://www.acessoainformacao.gov.br/assuntos/conheca-seu-direito/principais-aspectos>> Acesso em: 15 jun. 2016. 18
- [30] Senado Federal. *Processo Legislativo: Conceitos Básicos*. Disponível em: <<https://www12.senado.leg.br/jovensenador/arquivos/conceitos-processo-legislativo>> Acesso em: 20 jun. 2016. 10
- [31] Senado Federal. *RESOLUÇÃO DO SENADO FEDERAL Nº 93, DE 1970*. 1970. Disponível em: <<http://www2.senado.leg.br/bdsf/item/id/99774>> Acesso em: 15 jun. 2016. 9
- [32] A. B. H. FERREIRA. *Miniaurélio: O minidicionário da língua portuguesa*. 2004. 4
- [33] N. D. GALVÃO and H. F. MARIN. Técnica de mineração de dados: uma revisão da literatura. *Acta paul. enferm;* 22(5): 686-690, 2009. 30
- [34] Gartner. *Gartner Magic Quadrant*. Disponível em: <[http://www.gartner.com/technology/research/methodologies/research\\_mq.jsp](http://www.gartner.com/technology/research/methodologies/research_mq.jsp)> Acesso em: 02 jul. 2016. v, 27, 28
- [35] Ronaldo GOLDSCHMIDT and Emmanuel PASSOS. *Data Mining: Um Guia Prático*. 2005. 20
- [36] Florin GORUNESCU. *Data mining: Concepts, models and techniques*. Springer, v. 12, 2011. 32
- [37] Ernesto HABERKORN. *Gestão Empresarial com ERP*. São Paulo, 2 ed. edition, 2004. 21
- [38] J. HAN and M. KAMBER. *Data Mining: Concepts and Techniques*. Number 800p. Morgan Kaufmann, 2 edition edition, 2006. 31, 33, 34
- [39] Thomas H. HARRISON. *Intranet data warehouse*. São Paulo, Berkeley Brasil, 1998. 33
- [40] W. H. INMON. *Como Construir o Data Warehouse*. 1997. 23
- [41] D. T. LAROSE. *Discovering Knowledge in Data: An Introduction to Data Mining*. 2005. 32

- [42] COELHO I. M. BRANCO P. G. G. MENDES, G. F. *Curso de Direito Constitucional*. São Paulo, 2009. 12
- [43] MONTESQUIEU. *O Espírito das Leis*. São Paulo, 1993. 5
- [44] A. MORAES. *Direito Constitucional*. 29 ed. edition, 2004. 948p. 7, 8
- [45] C. C. P. MOREIRA. Proposições de iniciativa parlamentar: os projetos de lei apresentados no senado federal entre 1987 e 2005. Mathesis, Universidade de Brasília, Brasília, out 2006. 5, 9
- [46] Mihaela MUNTEAN and Traian SURCEL. Agile BI – the future of BI. *Informatica Economică vol. 17, no. 3/2013*, 2013. Disponível em: <http://revistaie.ase.ro/content/67/10%20-%20Muntean,%20Surcel.pdf> Acesso em: 10 ago. 2016. v, vi, 24, 25, 26
- [47] Congresso Nacional. *RESOLUÇÃO DO CONGRESSO NACIONAL Nº 1, DE 1970*. 1970. Disponível em: <http://www2.camara.leg.br/legin/fed/rescon/1970-1979/resolucao-1-11-agosto-1970-497934-republicacaoatualizada-1-pl.html> Acesso em: 15 jun. 2016. 9
- [48] Vanessa NORONHA. *Business Intelligence - Conjunto de Softwares que ajudam em decisões estratégicas*. 2013. 21
- [49] University of Waikato. *Weka 3: Data Mining Software in Java*. 2014. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/index.html> Acesso em: 30 mai. 2016. 37
- [50] L. B. PACHECO and P. R. MENDES. *Questões sobre Processo Legislativo e Regimento Interno*. 3 ed edition, 2015. 11, 12, 13, 14
- [51] L. B. PALMEIRA and M. P. SANTOS. Evasão no bacharelado em ciência da computação da universidade de Brasília: análise e mineração de dados. *Departamento de Ciência da Computação. Universidade de Brasília, Brasília.*, 2014. 30
- [52] F. V. PRIMAK. *Decisões com B.I. - Business Intelligence*. 2008. 21, 22, 23, 24
- [53] TDWI Research. *Tdwi bi benchmark reports: Organizational and performance metrics for business intelligence teams*. 2011. Disponível em: <http://tdwi.org/research/2011/09/2011-tdwi-bi-benchmark> Acesso em: 11 ago. 2016. 24
- [54] L. O. RODRIGUES. *Poder político*. 1999. Disponível em: <http://mundoeducacao.bol.uol.com.br/sociologia/poder-politico.htm> Acesso em: 07 jun. 2016. 4
- [55] J. E. SENE. A sociedade do conhecimento e as reformas educacionais. 2008. Disponível em: <http://www.ub.edu/geocrit/-xcol/91.htm> Acesso em: 15 jun. 2016. 15, 16
- [56] HIGHTOWER R. J. SHARIAT, M. Conceptualizing business intelligence architecture. *Marketing Management Journal*, v. 17(n. 2):p. 40–46, 2007. 21

- [57] José Afonso da SILVA. *Processo Constitucional de Formação de Leis*. 2 ed. edition, 2006. 8
- [58] R. A SILVA, J. L. SILVA, and S. P. RESENDE. O trabalho docente e tutorial no curso de administração da puc minas virtual. 2013. 15
- [59] Imre SIMON. A revolução digital e a sociedade do conhecimento. *São Paulo: IME-USP*, 1999. Disponível em: <http://www.ime.usp.br/~is/ddt/mac333> Acesso em: 15 jun. 2016. 16
- [60] M. C. SIQUEIRA. *Gestão Estratégica da Informação*. Rio de Janeiro, 2005. 158p. 15
- [61] Tadao TAKAHASHI. *Sociedade da informação no Brasil: livro verde*. Brasília, 2000. <http://livroaberto.ibict.br/bitstream/1/434/1/Livro> 16, 17, 18
- [62] Michel TEMER. *Elementos de Direito Constitucional*. São Paulo, 24 ed. edition, 2014. 4, 6, 7, 12, 13
- [63] F. P. TITÃO and N. VIAPIANA. A importância da organização da informação no século XXI: reflexões. *Revista ACB: Biblioteconomia em Santa Catarina*, v. 13(n. 1):p. 26–36, jan/jun 2008. 15
- [64] R. F. van der Lans. Data virtualization for business intelligence agility. 2012. Disponível em: [http://purl.manticoretechnology.com/ImgHost/582/12917/2012/resources/whitepapers/Whitepaper\\_DVAgility\\_RickVDL\\_February%202012.pdf](http://purl.manticoretechnology.com/ImgHost/582/12917/2012/resources/whitepapers/Whitepaper_DVAgility_RickVDL_February%202012.pdf) Acesso em: 11 ago. 2016. 26
- [65] R. F. van der Lans. Data virtualization for business intelligence systems: revolutionizing data integration for data warehouses. 2012. Disponível em: <http://www.irmuk.co.uk/articles/Rick%20van%20der%20Lans%20-%20What%20is%20Data%20Virtualization.pdf> Acesso em: 11 ago. 2016. 26
- [66] A. A. VANTI. Implantación de sistemas de información y la contribución de la cultura organizacional: analisis desde el punto de vista empresarial. *International Conference and Technology in the New Enterprise, Havana*, 2003. 21
- [67] Panos VASSILIADIS and Alkis SIMITSIS. *Extraction, Transformation, and Loading*. Springer US, Boston, MA, 2009. 22
- [68] Guilherme VIEIRA. Implantação de demonstrativos e indicadores de desempenho através das ferramentas de business intelligence do qlik view para análise da produção na fagundes construção e mineração ltda. 2001. 21
- [69] Jorge WERTHEIN. A sociedade da informação e seus desafios. *Ci. Inf., Brasília*, v. 29(n. 2):p. 71–77, mai/ago 2000. Disponível em: <http://www.scielo.br/pdf/%0D/ci/v29n2/a09v29n2.pdf> Acesso em: 17 jun. 2016. 16, 17
- [70] I. H. WITTEN, E. FRANK, and M. C. HALL. *Data Mining: Practical Machine Learning Tools and Techniques*. Number 560p. Morgan Kaufmann, San Francisco, 2 ed edition, 2005. v, 37, 38, 39, 40, 52, 74

[71] World Wide Web Consortium (W3C). Dados abertos governamentais. 2011. 1

# Anexo I

## *Script* de busca dos dados de proposição

```
//-----Primeira Etapa (Extração): Busca das Proposições  
  Votadas no Web Service da Câmara, -----//  
//disponível em: http://www.camara.gov.br/SitCamaraWS/  
  Proposicoes.aspx/ListarProposicoesVotadasEmPlenario  
/*
```

Os parâmetros utilizados são: Ano: 2015 e 2016 (Atual  
legislatura) | Tipo de Proposição: Todas  
\*/

```
//-----Tabela Proposição ano 2016-----//  
proposicao:
```

LOAD

```
//Função DISTINCT para evitar coleta de dados duplicados  
DISTINCT codProposicao,
```

```
//A partir do nomeProposição (ex.: PEC 171/2016), busca a  
  informação do ID da Proposição (ex.: 171)  
LEFT(SubField(nomeProposicao, ' ', 2), LEN(SubField(  
  nomeProposicao, ' ', 2))-5) AS idProp,
```

```
//A partir do nomeProposição (ex.: PEC 171/2016), busca a  
  informação do Tipo de Proposição (ex.: PEC)  
SubField(nomeProposicao, ' ', 1) AS tipoProp,
```

```
//A partir do nomeProposição (ex.: PEC 171/2016), busca a  
  informação do Ano de Proposição (ex.: 2016)  
Right(SubField(nomeProposicao, ' ', 2), 4) AS anoProp,
```

```
nomeProposicao
```

```

FROM [http://www.camara.gov.br/SitCamaraWS/Proposicoes.aspx/
ListarProposicoesVotadasEmPlenario?ano=2016&tipo=]
(XmlSimple, Table is [proposicoes/proposicao]);

//-----Tabela Proposição ano 2015-----//
LOAD
//Função DISTINCT para evitar coleta de dados duplicados
DISTINCT codProposicao,

//A partir do nomeProposição (ex.: PEC 171/2016), busca a
informação do ID da Proposição (ex.: 171)
LEFT(SubField(nomeProposicao, ' ', 2), LEN(SubField(
nomeProposicao, ' ', 2))-5) AS idProp,

//A partir do nomeProposição (ex.: PEC 171/2016), busca a
informação do Tipo de Proposição (ex.: PEC)
SubField(nomeProposicao, ' ', 1) AS tipoProp,

//A partir do nomeProposição (ex.: PEC 171/2016), busca a
informação do Ano de Proposição (ex.: 2016)
Right(SubField(nomeProposicao, ' ', 2), 4) AS anoProp,

nomeProposicao

FROM [http://www.camara.gov.br/SitCamaraWS/Proposicoes.aspx/
ListarProposicoesVotadasEmPlenario?ano=2015&tipo=]
(XmlSimple, Table is [proposicoes/proposicao]);

//-----Função para gerar o arquivo de dados Qlik (.qvd)
-----//
STORE proposicao INTO [C:\Users\laura.solano\Desktop\Laura\TCC
Laura\Dados\proposicao.qvd](qvd);

//-----Função para deletar a tabela da aplicação
-----//
drop table proposicao;

//-----Leitura do arquivo de dados gerado previamente (.qvd
),
filtrado pelo tipo de proposição PEC, MPV e PL-----//
proposicao:

```



```
LOAD codProposicao ,
      idProp ,
      tipoProp ,
      anoProp ,
      nomeProposicao
FROM
[C:\Users\laura.solano\Desktop\Laura\TCC Laura\Dados\proposicao.
  qvd]
(qvd)
WHERE tipoProp = 'PEC' OR tipoProp = 'MPV' OR tipoProp = 'PL';
```

## Anexo II

# Script iterativo para busca dos dados no *Web Service* da Câmara dos Deputados

```
//-----Segunda Etapa (Extração e Transformação): Busca dos  
  Detalhes de Votação das Proposições no Web Service da Câmara  
-----//  
//disponível em: http://www.camara.gov.br/SitCamaraWS/  
  Proposicoes.aspx/ObterVotacaoProposicao  
  
//Definição da variável com o time out necessário para execução  
  da leitura no Web Service da Câmara  
timeoutProposicao = 100;  
  
//Definição da variável com a quantidade de linhas existentes na  
  tabela 'proposicao' lida na aba anterior  
numProposicoes = NoOfRows('proposicao') - 1;  
  
//Votos Deputados  
  
//Criação do loop iterativo de 0 até o número de linhas  
  existentes na tabela proposicao  
for i = 0 to $(numProposicoes)  
  
  idProp = Peek('idProp', i, 'proposicao'); //Busca da  
  informação de idProp da tabela 'proposicao' na linha  
  'i'  
  LET siglaProp = Peek('tipoProp', i, 'proposicao'); //  
  Busca da informação de siglaProp da tabela '  
  proposicao' na linha 'i'
```

```

anoProp = Peek('anoProp', i, 'proposicao'); //Busca da
        informação de anoProp da tabela 'proposicao' na linha
        'i'

//Definição da URL de acesso ao Web Service da Câmara,
        utilizando as variáveis idProp, siglaProp e anoProp
        definidos previamente.
urlProp = 'http://www.camara.gov.br/SitCamaraWS/
        Proposicoes.aspx/ObterVotacaoProposicao?tipo=' &
        siglaProp & '&numero=' &$(idProp) & '&ano=' &$(
        anoProp);

/*
        Tabela que armazena as informações do Voto do
        Deputado
*/
[Voto Deputado]:
LOAD DISTINCT Nome AS [Nome do Deputado],
        $(idProp) AS idProp_Deputado,
        '$(siglaProp)' AS siglaProp_Deputado,
        $(anoProp) AS anoProp_Deputado,
        ideCadastro AS [ID Cadastro Deputado],
        UPPER(Partido) AS [Partido Deputado], //Função UPPER
        que capitaliza todos os caracteres
        UF AS [UF Deputado],
        Voto AS [Voto Deputado],
        %Key_Votacao_F7152F3C01FFE75C AS %Key_Vot_Dep
FROM $(urlProp) (XmlSimple, Table is [proposicao/
        Votacoes/Votacao/votos/Deputado]);

        SLEEP $(timeoutProposicao); //Espera o timeout definido
        na variável no início da aba
NEXT //Passa para a próxima linha i+1

//Orientação Partido
for i = 0 to $(numProposicoes) //Criação do loop iterativo de 0
        até o número de linhas existentes na tabela proposicao

        idProp = Peek('idProp', i, 'proposicao'); //Busca da
        informação de idProp da tabela 'proposicao' na linha
        'i'
        LET siglaProp = Peek('tipoProp', i, 'proposicao'); //
        Busca da informação de siglaProp da tabela '
        proposicao' na linha 'i'

```

```

anoProp = Peek('anoProp', i, 'proposicao'); //Busca da
informação de anoProp da tabela 'proposicao' na linha
'i'

//Definição da URL de acesso ao Web Service da Câmara,
utilizando as variáveis idProp, siglaProp e anoProp
definidos previamente.
urlProp = 'http://www.camara.gov.br/SitCamaraWS/
Proposicoes.aspx/ObterVotacaoProposicao?tipo=' &
siglaProp & '&numero=' &$(idProp) & '&ano=' &$(
anoProp);

/*
Tabela que armazena as informações da Orientação
Partidária
*/
*/
Orientação:
LOAD UPPER(Sigla) AS [Sigla Orientação],
$(idProp) AS idProp_Orientação,
'$(siglaProp)' AS siglaProp_Orientação,
$(anoProp) AS anoProp_Orientação,
orientacao AS [Orientação],
%Key_Votacao_F7152F3C01FFE75C AS %Key_Vot_Ori
FROM $(urlProp) (XmlSimple, Table is [proposicao/
Votacoes/Votacao/orientacaoBancada/bancada]);

SLEEP $(timeoutProposicao); //Espera o timeout definido
na variável no início da aba
NEXT //Passa para a próxima linha i+1

//Resumo
for i = 0 to $(numProposicoes) //Criação do loop iterativo de 0
até o número de linhas existentes na tabela proposicao

idProp = Peek('idProp', i, 'proposicao'); //Busca da
informação de idProp da tabela 'proposicao' na linha
'i'
LET siglaProp = Peek('tipoProp', i, 'proposicao'); //
Busca da informação de siglaProp da tabela '
proposicao' na linha 'i'
anoProp = Peek('anoProp', i, 'proposicao'); //Busca da
informação de anoProp da tabela 'proposicao' na linha
'i'

```

```

//Definição da URL de acesso ao Web Service da Câmara,
//utilizando as variáveis idProp, siglaProp e anoProp
//definidos previamente.
urlProp = 'http://www.camara.gov.br/SitCamaraWS/
Proposicoes.asmx/ObterVotacaoProposicao?tipo=' &
siglaProp & '&numero=' &$(idProp) & '&ano=' &$(
anoProp);

/*
Tabela que armazena as informações do Resumo da
Votação da Proposição
*/
Resumo:
LOAD Resumo,
$(idProp) AS idProp_Resumo,
'$(siglaProp)' AS siglaProp_Resumo,
$(anoProp) AS anoProp_Resumo,
Data,
Hora,
ObjVotacao AS ObjVotacao_Resumo,
codSessao AS [Código Sessão],
%Key_Votacao_F7152F3C01FFE75C AS %Key_Vot_Res //
Key for this table: proposicao/Votacoes/Votacao
FROM $(urlProp) (XmlSimple, Table is [proposicao/
Votacoes/Votacao]);

SLEEP $(timeoutProposicao); //Espera o timeout definido
na variável no início da aba
NEXT //Passa para a próxima linha i+1

//-----Função para gerar o arquivo de dados Qlik (.qvd)
//-----//
STORE Orientação INTO [C:\Users\laura.solano\Desktop\Laura\TCC
Laura\Dados\orientacao.qvd](qvd);
STORE [Voto Deputado] INTO [C:\Users\laura.solano\Desktop\Laura\
TCC Laura\Dados\voteputado.qvd](qvd);
STORE Resumo INTO [C:\Users\laura.solano\Desktop\Laura\TCC Laura
\Dados\resumo.qvd](qvd);

//-----Função para deletar a tabela da aplicação
//-----//
drop tables proposicao, Orientação, Resumo, [Voto Deputado];

```

## Anexo III

# Tabela DE-PARA de tratamento de Blocos Partidários

Tabela III.1: DE-PARA Blocos Partidários

BLOCOS PARTIDÁRIOS	PARTIDOS
PMDBPEN	PMDB;PEN
PMDBPPPPTBDEM...	PMDB;PP;PTB;DEM
PMDBPPPPTBDEMSPSCPHSPEN	PMDB;PP;PTB;DEM;SD;PSC;PHS;PEN
PMDBPPPPTBPSCPHSPEN	PMDB;PP;PTB;PSC;PHS;PEN
PPPROS	PP;PROS
PPPTBPSC	PP;PTB;PSC
PPPTBPSCPHS	PP;PTB;PSC;PHS
PPSPV	PPS;PV
PRBPTNPMNPRPPSDCPRTB	PRB;PTN;PMN;PRP;PSDC;PRTB
PRBPTNPMNPRPPSDCPRTBPTCPSLPTDOB	PRB;PTN;PMN;PRP;PSDC;PRTB;PTC;PSL;PTDOB
PRBPTNPMNPRPPSDCPTCPSLPTDOB	PRB;PTN;PMN;PRP;PSDC;PTC;PSL;PTDOB
PRBPTNPMNPTCPTDOB	PRB;PTN;PMN;PTC;PTDOB
PRBPTNPMNPTCPTDOBPSL	PRB;PTN;PMN;PTC;PTDOB;PSL
PRBPTNPTDOBPSL	PRB;PTN;PTDOB;PSL
PRPSDPROS	PR;PSD;PROS
PRPTDOBPRP	PR;PTDOB;PRP
PRPTDOBPRPPHSPTCPSLPRTB	PR;PTDOB;PRP;PHS;PTC;PSL;PRTB
PSDBPSBPPS	PSDB;PSB;PPS
PSDBPSBPPSPV	PSDB;PSB;PPS;PV
PTBPSDC	PTB;PSDC
PTNPTDOBPSL	PTN;PTDOB;PSL
PTPSDPRPDTPROSPCDOB	PT;PSD;PR;PDT;PROS;PCDOB
PVPPS	PV;PPS
PMDBPPPPTBDEMSPSCPHSPEN	PMDB;PP;PTB;DEM;PSC;PHS;PEN
PRBPTNPTCPTDOBPSL	PRB;PTN;PTC;PTDOB;PSL
PRPROS	PR;PROS

## Anexo IV

### *Script* de transformação de dados

```
//-----Terceira Etapa (Transformação e Carga): Leitura dos
dados salvos no arquivo Qlik (.qvd)-----//

/*
    Tabela Orientação do Partido
*/
[Orientação do Partido]:
LOAD DISTINCT //Remove as linhas de dados duplicadas, caso
    exista

    //Aplica o mapeamento do bloco partidário e unificação
    do partido solidariiedade
    APPLYMAP('MUDA_SOLIDARIEDADE', SUBFIELD(APPLYMAP('
        BLOCO_PARTIDARIO', [Sigla Orientação]), ';')) AS [
        Sigla Orientação Partido],

    //Criação da chave de ligação das tabelas 'Orientação
    Partido' e 'Votações'
    siglaProp_Orientação & ' ' & idProp_Orientação & '/' &
    anoProp_Orientação & '-' & %Key_Vot_Ori & '-' &
    APPLYMAP('MUDA_SOLIDARIEDADE', SUBFIELD(APPLYMAP('
        BLOCO_PARTIDARIO', [Sigla Orientação]), ';')) AS %
    Orientação_Partido,

    Orientação AS [Orientação Partido],

    //Aplica o mapeamento da remoção dos caracteres especiais do
    campo Orientação
    MapSubString('MAP_CHAR_ESP', Orientação) AS [Orientação
    Partido Char]
FROM
```

```

[C:\Users\laura.solano\Desktop\Laura\TCC Laura\Dados\orientacao.
  qvd]
(qvd);

/*
    Tabela Resumo das votações
*/
[Resumo das votações]:
LOAD siglaProp_Resumo & ' ' & idProp_Resumo & '/' &
    anoProp_Resumo & '-' & %Key_Vot_Res AS %Chave, //Criação da
    chave de ligação entre as tabelas 'Resumo das votações' e '
    Votações'
    siglaProp_Resumo & ' ' & idProp_Resumo & '/' &
    anoProp_Resumo AS Proposição, //Gera o campo nome
    completo da proposição
    Resumo,
    idProp_Resumo,
    siglaProp_Resumo,
    anoProp_Resumo,
    MONIH(Data) AS [Mês da Votação], //Coleta o dado do nome do
    mês de votação do campo 'Data'
    SUBFIELD(Data, '/', 2) AS [No Mês], //Coleta o dado do nú
    mero do mês de votação do campo 'Data'
    YEAR(Data) AS [Ano da Votação], //Coleta o dado do ano de
    votação do campo 'Data'
    YEAR(Data) & IF(LEN(SUBFIELD(Data, '/', 2)) = 1, 0&
    SUBFIELD(Data, '/', 2), SUBFIELD(Data, '/', 2)) AS [
    Período Votação], //Coleta o dado do período (Ano/Mês)
    de votação do campo 'Data'
    Data,
    Hora,
    ObjVotacao_Resumo,
    [Código Sessão]
FROM
[C:\Users\laura.solano\Desktop\Laura\TCC Laura\Dados\resumo.qvd]
(qvd)
WHERE YEAR(Data) > 2014;

/*
    Tabela Votações
*/
[Votações]:
LOAD siglaProp_Deputado & ' ' & idProp_Deputado & '/' &
    anoProp_Deputado & '-' & %Key_Vot_Dep AS %Chave, //Criação da
    chave de ligação entre as tabelas 'Resumo das votações' e '
    Votações'

```



```

//Criação da chave de ligação das tabelas 'Orientação
Partido' e 'Votações'
siglaProp_Deputado & ' ' & idProp_Deputado & '/' &
anoProp_Deputado & '-' & %Key_Vot_Dep & '-' &
APPLYMAP('MUDA_SOLIDARIEDADE',[Partido Deputado]) AS
%Orientação_Partido,
UPPER([Nome do Deputado]) AS [Nome do Deputado], //Funç
ão para capitalizar os nomes dos deputados
APPLYMAP('MAP_DEP', UPPER(MapSubString('MAP_CHAR_ESP',
[Nome do Deputado]))) AS [Nome Deputado Char], //
Aplica o mapeamento de unificação dos nomes dos
deputados
idProp_Deputado,
siglaProp_Deputado,
anoProp_Deputado,
[ID Cadastro Deputado],
APPLYMAP('MUDA_SOLIDARIEDADE', [Partido Deputado]) as [
Partido Deputado], //Aplica o mapeamento de unificação
do partido solidariedade
[UF Deputado],
[Voto Deputado],
MapSubString('MAP_CHAR_ESP', [Voto Deputado]) AS [Voto
Deputado Char] //Aplica o mapeamento de remoção dos
caracteres especiais
FROM
[C:\Users\laura.solano\Desktop\Laura\TCC Laura\Dados\
votodeputado.qvd]
(qvd);

left join //Junção da tabela 'Votações' com a tabela agrupada de
votos em concordância com o resultado vencedor

[Votos Agrupados]:
LOAD %Chave,
IF ([Votos Sim]>[Votos Não] AND [Votos Sim]>[Votos
Abstenção], 'Sim',
IF ([Votos Não]>[Votos Sim] AND [Votos Não]>[Votos
Abstenção], 'Não', 'Abstenção')) AS [Voto Ganhador]
//Define qual foi o voto vencedor de cada votação da
proposição
;
LOAD siglaProp_Deputado & ' ' & idProp_Deputado & '/' &
anoProp_Deputado & '-' & %Key_Vot_Dep AS %Chave, //Chave que
faz a ligação da tabela 'Votações' com a tabela 'Votos
Agrupados'

```

```

SUM(IF ([Voto Deputado]='Sim', 1, 0)) AS [Votos Sim], //Cria
ção da flag para os votos 'Sim'
SUM(IF ([Voto Deputado]='Não', 1, 0)) AS [Votos Não], //Cria
ção da flag para os votos 'Não'
SUM(IF ([Voto Deputado]='Abstenção', 1, 0)) AS [Votos Absten
ção] //Criação da flag para os votos 'Abstenção'
FROM
[C:\Users\laura.solano\Desktop\Laura\TCC Laura\Dados\
votodeputado.qvd]
(qvd)
GROUP BY siglaProp_Deputado & ' ' & idProp_Deputado & '/' &
anoProp_Deputado & '-' & %Key_Vot_Dep; //Agrupado por proposi
ção

//-----Função para gerar o arquivo de dados Qlik (.qvd)
-----//
STORE [Votações] INTO [C:\Users\laura.solano\Desktop\Laura\TCC
Laura\Dados\FINAL_VOTO_DEPUTADO.qvd](qvd);

//-----Função para deletar a tabela da aplicação
-----//
DROP TABLE [Votações];

/*
Tabela 'Votações' Final
*/
[Votações]:
LOAD %Chave,
%Orientação_Partido,
[Nome do Deputado],
[Nome Deputado Char],
idProp_Deputado,
siglaProp_Deputado,
anoProp_Deputado,
[ID Cadastro Deputado],
[Partido Deputado],
[UF Deputado],
[Voto Deputado],
[Voto Deputado Char],
[Voto Ganhador],
IF ([Voto Deputado]=[Voto Ganhador], 1, 0) AS [Voto em
Concordância] //Criação da Flag dos votos dos deputados
em concordância com o resultado ganhador da proposição
FROM
[C:\Users\laura.solano\Desktop\Laura\TCC Laura\Dados\
FINAL_VOTO_DEPUTADO.qvd]

```

(qv d);

## Anexo V

# Dicionário de Dados

Tabela V.1: Dicionário de dados da tabela Resumo das Votações

Tabela: Resumo das Votações		
Campo	Descrição	Tipo
%Chave	Chave de ligação entre as tabelas 'Resumo das votações' e 'Votações'	String
Proposição	Nome completo da proposição. Ex.: PEC 457/2005	String
Resumo	Resumo da sessão de votação	String
id_Prop_Resumo	Identificador da proposição	String
siglaProp_Resumo	Sigla da proposição	String
anoProp_Resumo	Ano da proposição	Int
Mês da Votação	Nome do mês em que ocorreram as votações	String
No Mês	Número do mês em que ocorreram as votações	Int
Ano da Votação	Ano em que ocorreram as votações	Int
Período Votação	Ano/Mês das votações	String
Data	Data em que ocorreu a votação	Data
Hora	Hora da sessão	String
ObjVotacao_Resumo	Objeto colocado em votação na sessão	String
Código Sessão	Código da sessão em plenário	String

Tabela V.2: Dicionário de dados da tabela Votações

<b>Tabela: Votações</b>		
<b>Campo</b>	<b>Descrição</b>	<b>Tipo</b>
%Chave	Chave de ligação entre as tabelas 'Resumo das votações' e 'Votações'	String
%Orientacao_Partido	Chave de ligação das tabelas 'Orientação Partido' e 'Votações'	String
Nome do Deputado	Nome do deputado	String
Nome do Deputado Char	Nome do deputado sem caracteres especiais	String
idProp_Deputado	Identificador da proposição votada pelo deputado	Int
siglaProp_Deputado	Sigla da proposição votada pelo deputado	Int
anoProp_Deputado	Ano da proposição votada pelo deputado	Int
ID Cadastro Deputado	Identificador do deputado	Int
Partido Deputado	Partido do deputado	String
UF Deputado	Estado do deputado	String
Voto Deputado	Voto do deputado. Ex.: Sim, Não, Abstenção	String
Voto Deputado Char	Voto do deputado sem caracteres especiais	String
Voto ganhador	Voto vencedor de casa votação	String
Voto em concordância	Flag dos votos dos deputados em concordância com o resultado ganhador da votação	Bool

Tabela V.3: Dicionário de dados da tabela Orientação

<b>Tabela: Orientação</b>		
<b>Campo</b>	<b>Descrição</b>	<b>Tipo</b>
%Chave	Chave de ligação entre as tabelas 'Resumo das votações' e 'Votações'	String
Sigla Orientação	Sigla do Partido	String
Orientação	Orientação do partido. Ex.: Sim, Não, Obstrução, Abstenção.	String

Tabela V.4: Dicionário de dados da tabela Orientação Partido

<b>Tabela: Orientação do Partido</b>		
<b>Campo</b>	<b>Descrição</b>	<b>Tipo</b>
%Orientacao_Partido	Chave de ligação das tabelas 'Orientação Partido' e 'Votações'	String
Sigla Orientação Partido	Sigla do Partido	String
Orientação Partido	Orientação do partido. Ex.: Sim, Não, Obstrução, Abstenção.	String
Orientação Partido Char	Orientação do partido sem caracteres especiais	String