

Mineração de Dados

Análise dos votos dos senadores por meio de mineração de dados

Felipe Ken Hirano, Iuri Batista Beserra

Orientador

Prof. Dr. Jan Correa

Defesa de Trabalho de Conclusão de Curso

Departamento de Ciência da Computação
Universidade de Brasília - UnB

04 de Dezembro de 2018

Sumário

- 1 Introdução
- 2 Referencial Teórico
- 3 Metodologia
- 4 Resultados
- 5 Conclusão

O problema

O Senado Federal disponibiliza dados à respeito das votações que ocorreram em plenário. Porém, estes dados estão em um formato de difícil compreensão o que dificulta a análise dos mesmos.

Hipótese

É possível à descoberta de padrões úteis nos dados do Senado Federal, utilizando corretamente algoritmos de mineração de dados para aperfeiçoar a análise.

Objetivo Geral

Esse trabalho acadêmico tem como objetivo geral à análise dos resultados obtidos através da mineração dos dados brutos disponibilizados pelo Senado Federal utilizando ferramentas como o *MySQL*, *Pentaho*, *WEKA* e *Orange* para descobrir padrões concretos sobre as votações que ocorreram nas sessões plenárias nos anos de 2015 à 2017.

Objetivos Específicos

- Analisar os dados gerais de votações de proposições feitas em plenário no período de 2015 à 2017 no Senado Federal;
- Tentar identificar relações entre os votos dos senadores e os resultados das proposições;
- Identificar se existem variações de votos dentro de um mesmo partido;
- Identificar padrões nas votações dos senadores.

O que é

- É a câmara alta do Congresso Nacional;
- Encarregado pela elaboração e fiscalização das leis, apresentando os valores e a vontade da população em nível federal;
- Representa os Estados Federados.

Proposições Legislativas

- É toda matéria submetida à deliberação da Casa Legislativa;
- Compreende a elaboração de emendas à Constituição, das várias modalidades de lei.

Proposta de Emenda à Constituição Federal

- Objetiva alterar partes do texto constitucional, sem a necessidade de se convocar nova assembleia constituinte;
- É discutida e votada em dois turnos, em cada Casa do Congresso, e será aprovada se obtiver êxito, na Câmara e no Senado.

Modalidades de Votações

Votações abertas

- As manifestações pelas aprovações das matérias ocorrem pelo ato dos senadores permanecerem sentados, enquanto aqueles que optam pela rejeição se levantam;
- Caso haja revisão dos votos, ela se dá por um processo nominal que é o registro eletrônico dos votos através do painel que está instalado no Plenário.

Votações fechadas

- Ocorrem no caso de indicação de autoridades, vetos do presidente, na cassação de parlamentares, entre outros motivos;
- Essa modalidade também utiliza o sistema eletrônico, mas só apresenta os resultados da deliberação a respeito da proposição que foi votada.

Data Mining

- Utiliza métodos tradicionais de análises de dados com algoritmos sofisticados para o processamento desse grande volumes de dados [2];
- Campo interdisciplinar que junta técnicas de máquinas de conhecimentos, reconhecimento de padrões, estatísticas, banco de dados e visualização, para conseguir extrair informações de grandes bases de dados [3].

KDD

- A mineração de dados é apenas uma etapa deste processo;
- O KDD é um processo manual de todo um conjunto de dados que se baseia na conversão dos dados para gerar uma informação e consequentemente um conhecimento [1].

Knowledge-Discovery in Databases

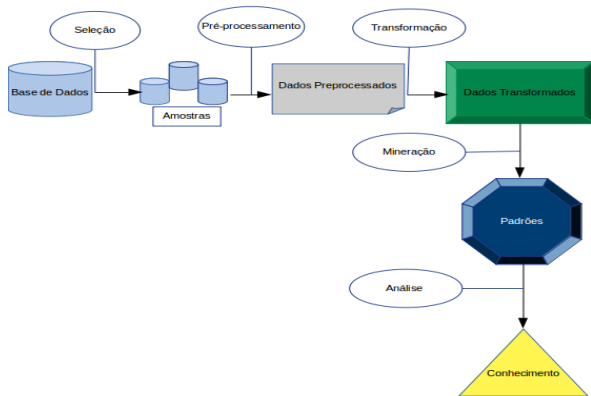


Figura: Processo de KDD proposto por Fayyad et al.[1].

Pentaho Data Integration

- Solução em código aberto que possui funcionalidades para desenvolvimento de mineração de dados;
- Foi estruturada para permitir que seus componentes extraiam, transformem e deem carga nos dados de diversas fontes, para diversos bancos e arquivos.

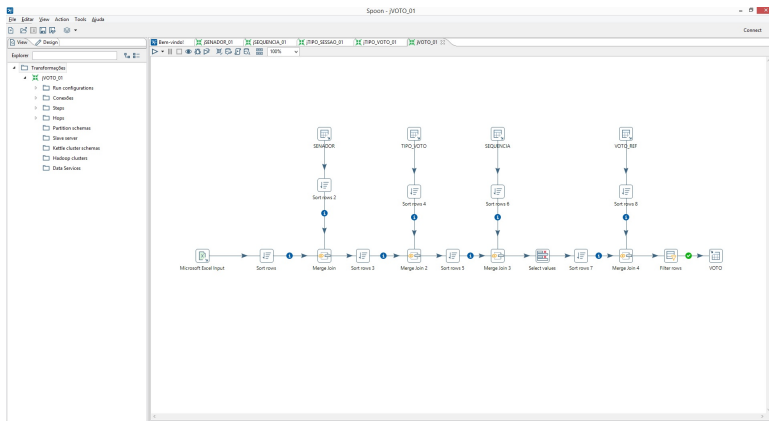


Figura: Interface *Pentaho Data Integration*.

- Desenvolvido em Java, pela Universidade de Waikato, situada na Nova Zelândia;
- É uma ferramenta de código aberto e de multiplataforma que apresenta uma coleção de algoritmos para a realização de mineração de dados através da aprendizagem de máquina.

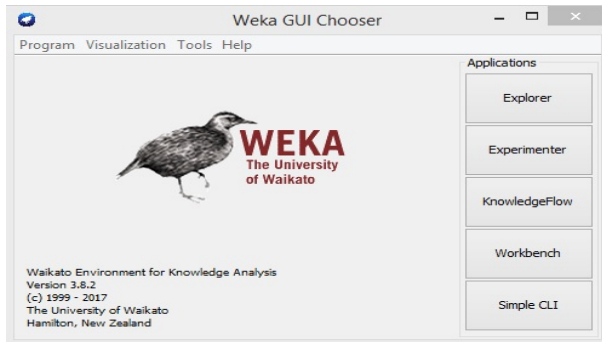


Figura: Interface inicial do Weka.

- Ferramenta de *Machine Learning* de código aberto que permite a análise de dados através de um fluxo vasto de funcionalidades;
- Estruturado para minerar dados, executar processamento de linguagem natural e mineração de texto;
- Utilizado para rodar o algoritmo *Hierarchical Clustering*.



Figura: Fluxo utilizado no *Orange*.

- Seleção dos dados;
- Pré-processamento;
- Tratamento dos dados;
- Mineração de dados;
- Descrição dos resultados.

Base de dados escolhida

- Os dados para estudo deste trabalho foram obtidos através do *WebService* do Senado Federal.

Coleta de dados

- Legislatura atual escolhida 55;
- Duração de 4 anos.

- Os resultados das buscas obtidos retornavam o período selecionado.

The screenshot shows the official website of the Brazilian Senate (Senado Federal). The header includes the logo and the text 'Fale com o Senado'. The main section is titled 'Atividade Legislativa' and contains a search bar. Below this, there are navigation tabs: 'Plenário', 'Projetos e Matérias', 'Comissões', 'Relatórios Legislativos', 'Legislação', 'Órgãos do Parlamento', and 'Autoridades'. A sub-section for 'Atividade Legislativa' includes links for 'Diários e Anais', 'Senado Mult mídia', 'RSS', 'Dados Abertos', and 'Perguntas Frequentes'. A yellow banner with a megaphone icon says 'Pesquisa - Ajude-nos a melhorar nosso portal.' Below this, the 'Votações Nominais' section is displayed, with a description: 'Consulte o voto de cada Parlamentar nas votações ostensivas, aquelas em que os votos foram cadastrados individualmente.' There are tabs for 'Matéria', 'Parlamentar', and 'Período'. The 'Período' tab is active, showing a date range selector with two date pickers and a 'Buscar' button. The footer contains various links and social media icons.

Figura: Página do Web Service do Senado.

- Os resultados das buscas retornavam dados apenas em formato *.pdf*;
- Com isso, foi necessário a conversão dos mesmos para um formato em que estes dados pudessem ficar dispostos em planilhas;
- Essa conversão foi realizada para uma melhor organização e manipulação dos dados em linhas e colunas.

Pré-Processamento

Arquivo Editar Layout Inserir Formatar Planilha Dados Separamentos Janela Ajuda

Sessão 004.xlsx - LibreOffice Calc

Tela: Nova (Remova)

12 100.00 % 0.0

SENAHOR

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	1	SENADOR	UF	PARTIDO	VOTO	DATA	MATERIA	EMENTA	DESCRIÇÃO	SEQUENCIA	SESSÃO	TIPO SESSÃO	RESULTADO		
2	1	Assi Vargas	RO	PP	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
3	2	Assi Vargas	RO	PSDB	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
4	3	Alcides Neves Feres	SP	PSDB	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
5	4	Alcides Dias	PR	PSDB	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
6	5	Assi Amêlia	RS	PP	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
7	6	Angela Pereira	RR	PT	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
8	7	Antonio Santana	MG	PSDB	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
9	8	Antonio Carlos Valadares	SE	PSB	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
10	9	Assis Oliveira	TO	PSDB	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
11	10	Benedicto de Lins	AL	PP	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
12	11	Bilac Magalhães	MT	PR	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
13	12	Caetano Cunha Lima	PI	PSDB	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
14	13	Ciro Nogueira	PI	PP	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
15	14	Conceição Brandão	DF	PPD	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
16	15	Dani Berges	SC	PSDB	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
17	16	Dei Abranches	AC	DEM	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
18	17	Delcídia de Azeite	MS	PT	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
19	18	Dorival Nogueira	TO	PT	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
20	19	Douglas Cunha	PE	PSB	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
21	20	Eduardo Leão	MA	PMDB	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
22	21	Eduardo Antonio	SE	PSB	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
23	22	Elmano Figue	PE	PSB	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
24	23	Enrico Oliveira	AC	PSB	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
25	24	Fátima Brenes	RN	PT	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
26	25	Fernando Bezerra Coelho	PE	PSB	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
27	26	Fernando Collor	AL	PSB	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
28	27	Fernando Ribeiro	PA	PMDB	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
29	28	Filipe Ribeiro	PA	PSDB	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
30	29	Genivaldo Alves Filho	RN	PSB	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
31	30	Gláucia Cunha	AC	PP	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
32	31	Glória Hoffmann	PR	PT	AP	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
33	32	Ilma José	DF	PSD	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
34	33	Ilvaneito Costa	PE	PT	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
35	34	Ivo Gago	RO	PP	Sim	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
36	35	José Alberto Souza	MA	PMDB	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		
37	36	João Capobianco	AP	PSB	ESNEX	10/03/2015	PEC 00032/2010	Altema co art. 92 + 111-A da	Votação da Proposta de	1	26		Aprovado		

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Planilha 1 de 1

PageStyle: Table 1

Módulo : Sessão 0

100%

Figura: Exemplo dos dados dispostos em planilha.

Transformação dos dados

- Modelar banco de dados MySQL;
- Criar *jobs* ETL carga nas tabelas;
- Conexão do MySQL com o Weka;
- Gerar .csv.

Tratamento dos Dados

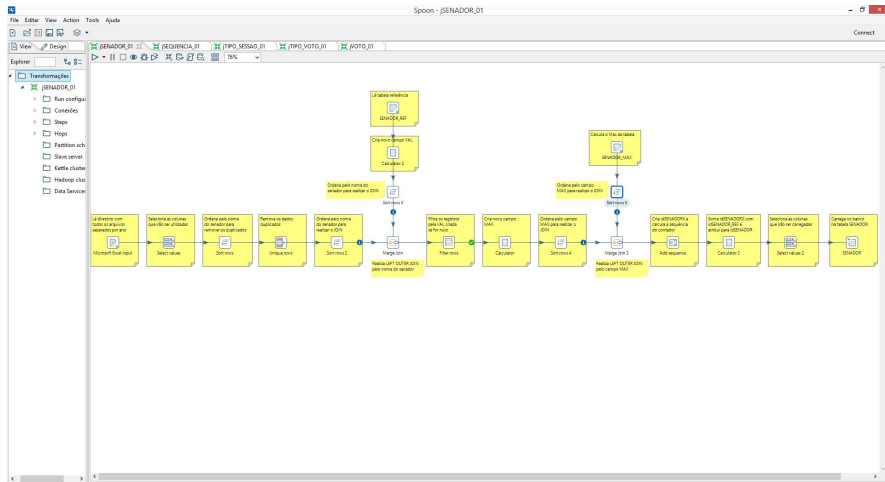


Figura: Job para carga no *Data Warehouse*.

Processo

- Foi necessário realizar uma divisão em duas etapas devido ao impeachment que ocorreu em agosto de 2016 pois o impeachment mudou o governo e suas alianças políticas;
- 58 proposições analisadas no pré impeachment e 37 pós impeachment;
- Aplicação dos algoritmos de mineração utilizando o software Weka e Orange.

Análise Estatística

- Foi gerado uma planilha em que as linhas eram os senadores e as as colunas eram as proposições;
- Foram utilizadas apenas as proposições consideradas acirradas;
- Removeu-se os partidos que tinham apenas um único senador;
- Removeu-se os senadores que tiveram participação menor de 75%;
- Agrupou-se os senadores por seus respectivos partidos.

Cálculo da coincidência de votos

- Os votos foram substituídos da seguinte forma: 1 para “Sim”, -1 para “Não” e 0 para a não emissão de voto;
- Definiu-se uma matriz de coincidência de votos;
- Quando o valor foi de 0, 1 ou -1 para os dois senadores que estão sendo comparados somou-se 1 em ambos os valores.

Resultados

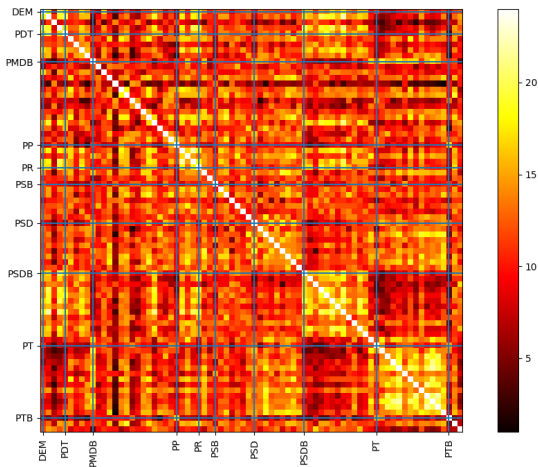


Figura: Gráfico de calor que ilustra as coincidências dos votos dos senadores de cada partido no período pré-impeachment.

Análise estatísticas dos senadores no período Pré-Impeachment

- Tivemos a participação de 75 senadores analisados em um total de 24 proposições;
- PSD, DEM, PT, PSDB para as proposições analisadas foram os partidos que apresentaram maior coincidência de votos internamente;
- O DEM e o PT foram os partidos que mais votaram de forma diferente;
- Em relação aos outros partidos, o PMDB é o partido com menor coincidência de votos dentro do próprio partido.

Resultados

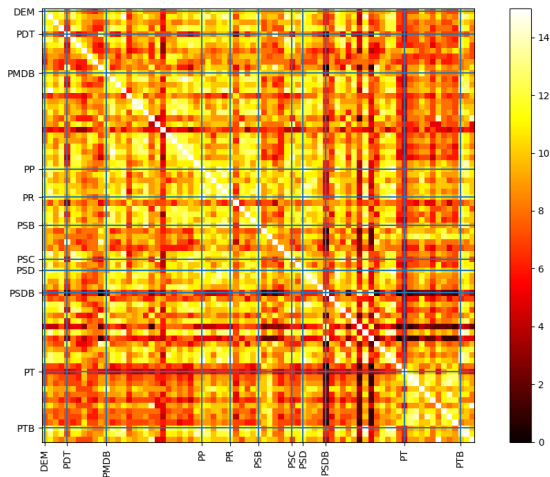


Figura: Gráfico de calor que ilustra as coincidências dos votos dos senadores de cada partido no período pós-impeachment.

Análise estatísticas dos senadores no período Pós-Impeachment

- Tivemos a participação de 77 senadores analisados em um total de 15 proposições;
- O PSDB é o partido que apresenta menor valor de coincidência de votos;
- O PTB, PP e o DEM são os que apresentam maior coincidência de votos;
- PSDB e PT foram os partidos de maior divergência de votos.

- Se constitui através da implantação de um modelo de classes;
- Categorização através de dados pré-estabelecidos;
- Utiliza este modelo para predizer as classes de objetos que não foram classificados.

J48

- Utiliza a abordagem dividir para conquistar de forma recursiva;
- Divide os atributos em subespaços e os associa a uma classe;
- Gera uma árvore de decisão fundamentada a partir um conjunto de dados.

Análise por classificação dos partidos no período Pré e Pós-Impeachment

- Filtramos apenas os votos 'Sim', 'Não', 'P-NRV' (presente não registrou voto) e 'Ncom' (não compareceu);
- Utilizamos o atributo "Resultado" como classe, para poder verificar pela maioria dos votos, os partidos que aprovam ou rejeitam as matérias;
- O voto do partido é definido pela maioria, em caso de empate o MySQL definiu algumas prioridades, nesta ordem: 'Sim', 'P-NRV', 'Não', 'Ncom';
- Tenta-se descobrir se uma proposição é aprovada ou rejeitada dependendo do voto do partido para as proposições.

Resultados

=== Run information ===

```
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: 1.4.0-weka.filters.unsupervised.attribute.Remove-RI
Instances: 94
Attributes: 16
          DEM
          PCdoB
          PDT
          PMDB
          PP
          PPS
          PR
          PSB
          PSC
          PSD
          PSDB
          PSOL
          PT
          PTR
          RESULTADO
```

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

```
PP = Sim
|   PSB = Sim
|   |   PSDB = Sim: Aprovado (48.0)
|   |   PSDB = P-NRV: Rejeitado (6.0)
|   PSB = P-NRV: Rejeitado (21.0/2.0)
|   PSB = Não: Aprovado (1.0)
PP = P-NRV: Rejeitado (12.0)
PP = Não: Rejeitado (6.0)
```

Number of Leaves : 6

Size of the tree : 9

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	90	95.7447 %
Incorrectly Classified Instances	4	4.2553 %
Kappa statistic	0.9149	
Mean absolute error	0.0666	
Root mean squared error	0.2045	
Relative absolute error	13.4083 %	
Root relative squared error	41.0322 %	
Total Number of Instances	94	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,922	0,000	1,000	0,922	0,959	0,918	0,963	0,975	Aprovado
	1,000	0,078	0,915	1,000	0,956	0,918	0,963	0,934	Rejeitado
Weighted Avg.	0,957	0,036	0,961	0,957	0,958	0,918	0,963	0,956	

=== Confusion Matrix ===

```
a b <-- classified as
47 4 | a = Aprovado
 0 43 | b = Rejeitado
```

Figura: Resultado do algoritmo J48 pré-impeachment por partido.

Análise por classificação dos partidos no período Pré-Impeachment

- 94 instâncias (proposições) e 16 atributos (partidos) analisados;
- Destas instâncias 51 eram aprovadas e 43 rejeitadas;
- 95,74% instâncias foram analisadas corretamente e apenas 4,25% incorretamente.

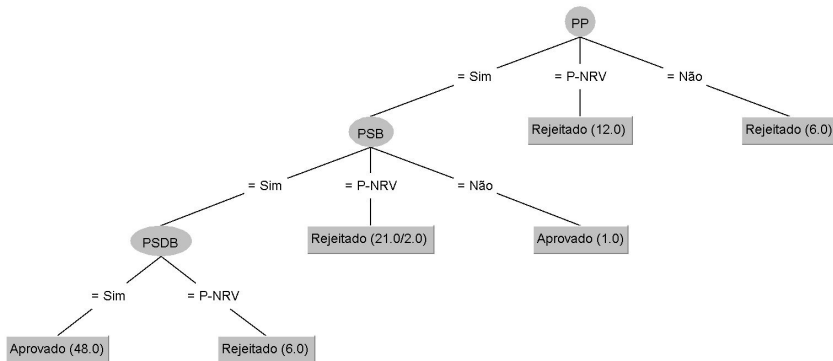


Figura: Árvore de decisão pré-impeachment por partido.

Análise da árvore de decisão dos partidos no período Pré-Impeachment

- Podemos observar na árvore que o partido PP é a raiz. O algoritmo o definiu como base classificar as proposições como aprovadas ou rejeitadas;
- Se os partidos PP, PSB e PSDB votaram sim, 51,06% foram classificadas como aprovadas;
- Quando o PP votou sim e o PSB esteve presente mas não registrou voto (P-NRV), 22,34% foram classificadas como rejeitadas.

Resultados

```
=== Run information ===
Time taken to build model: 0 seconds

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    1.4.0-weka.filters.unsupervised.attribute.Remove-R1
Instances:   62
Attributes:  16
             DEM
             PCDoB
             PDT
             PMDS
             PP
             PPS
             PR
             PRB
             PSB
             PSC
             PSD
             PSDB
             PSOL
             PT
             PTB
             RESULTADO

Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

DEM = Sim
|   PSC = Sim: Aprovado (31.0)
|   PSC = Não: Rejeitado (8.0/1.0)
|   PSC = P-NRV: Aprovado (1.0)
DEM = Não: Rejeitado (8.0)
DEM = P-NRV: Rejeitado (14.0)

Number of Leaves :    5
Size of the tree :    7

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      60          96.7742 %
Incorrectly Classified Instances    2           3.2258 %
Kappa statistic                    0.9355
Mean absolute error                 0.0579
Root mean squared error             0.1899
Relative absolute error             11.6063 %
Root relative squared error         38.0029 %
Total Number of Instances          62

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.   0.968    0.028    0.970     0.968    0.968      0.937    0.972     0.945     Aprovado
                0.968    0.028    0.970     0.968    0.968      0.937    0.972     0.945     Rejeitado

=== Confusion Matrix ===
      a  b  <-- classified as
31  2  |  a = Aprovado
 0 29  |  b = Rejeitado
```

Figura: Resultado do algoritmo J48 pós-impeachment por partido.

Análise por classificação dos partidos no período Pós-Impeachment

- 62 instâncias (proposições) e 16 atributos (partidos) analisados;
- Destas instâncias 33 eram aprovadas e 29 rejeitadas;
- 96,77% instâncias foram analisadas corretamente e apenas 3,22% incorretamente.

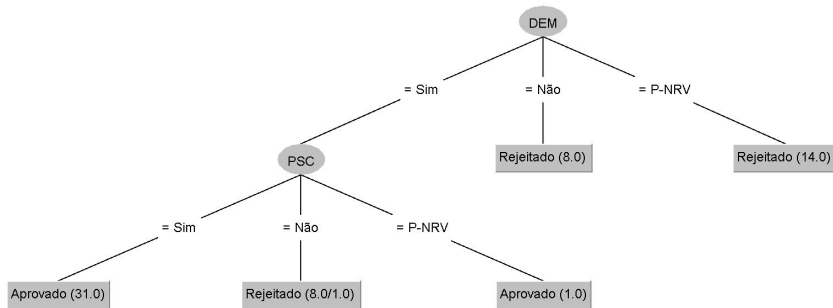


Figura: Árvore de decisão pós-impeachment por partido.

Análise da árvore de decisão dos partidos no período Pós-Impeachment

- Podemos observar na árvore que o partido DEM é a raiz. O algoritmo o definiu como base classificar as proposições como aprovadas ou rejeitadas;
- Se os partidos DEM e PSC votaram sim, 50,00% foram classificadas como aprovadas;
- Quando o DEM esteve presente mas não registrou voto (P-NRV), 22,58% foram classificadas como rejeitadas.

Regras de Associação

- Encontram as relações entre os diferentes itens dentro de um banco de dados;
- Foi utilizado o algoritmo *Apriori*.

Suporte

- O *support* é uma indicação da frequência com que um conjunto de itens aparece em um conjunto de dados;
- O *support* pode ser calculado conforme a equação 1:

$$\frac{F(X, Y)}{\text{TotaldeTransacoes}} \quad (1)$$

Confiança

- É uma indicação de frequência dado que uma regra já foi considerada verdadeira, descrevendo a chance de satisfazer apenas um item, já que satisfaz o segundo item;
- A confiança (*confidence*) pode ser calculado coforme a equação 2:

$$\frac{F(X, Y)}{F(X)} \quad (2)$$

Lift

- Mede a indenpendência entre os itens;
- Valores próximos de 1 indica que são indepedentes;
- Valores distantes e maiores que 1 podem indicar dependência entre os itens;

$$\frac{\text{confidence}(A \rightarrow C)}{\text{support}(C)} \quad (3)$$

Resultados

=== Run information ===

```
Scheme:      weka.associations.Apriori -N 100 -T 1 -C 1.1 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    1.3.2-weka.filters.unsupervised.attribute.Remove-R4,23,31,40,68,73,89,94,98,103-weka.filters.unsupervised.attribute.Remove-R1
Instances:   58
Attributes:  94
```

Figura: Cabeçalho dos resultados da associação tendo como base as matérias e os senadores no período pré-impeachment.

Análise por associação dos senadores no período Pré-Impeachment

Foram utilizadas 58 instâncias (votações que ocorreram em plenário) com a participação de 94 senadores como ilustra a figura 15 e essa análise trouxe fortes relações de associação entre os seguintes senadores conforme a figura 16:

- Raimundo Lira – PMDB/PB com Simone Tebet – PMDB/MS;
- Valdir Raupp – PMDB/RO com Waldemir Moka – PMDB/MS;
- Simone Tebet – PMDB/MS com Valdir Raupp – PMDB/RO;
- Raimundo Lira – PMDB/PB com Valdir Raupp – PMDB/RO.

Resultados

Apriori

Minimum support: 0.7 (41 instances)
Minimum metric: < lift: 1.1
Number of cycles performed: 6

Generated sets of large itemsets:

Size of set of large itemsets L(1): 36

Size of set of large itemsets L(2): 74

Size of set of large itemsets L(3): 26

Size of set of large itemsets L(4): 2

Best rules found:

```
1. Raimundo Lira - PMDB/PB-Sim Simone Tebet - PMDB/MS-Sim 44 ==> Valdir Raupp - PMDB/RO-Sim Waldemir Moka - PMDB/MS-Sim 41 conf:(0.93) < lift:(1.29)> lev:(0.16) [9] conv:(3.03)
2. Valdir Raupp - PMDB/RO-Sim Waldemir Moka - PMDB/MS-Sim 42 ==> Raimundo Lira - PMDB/PB-Sim Simone Tebet - PMDB/MS-Sim 41 conf:(0.98) < lift:(1.29)> lev:(0.16) [9] conv:(5.07)
3. Simone Tebet - PMDB/MS-Sim Valdir Raupp - PMDB/RO-Sim 43 ==> Raimundo Lira - PMDB/PB-Sim Waldemir Moka - PMDB/MS-Sim 41 conf:(0.95) < lift:(1.26)> lev:(0.14) [8] conv:(3.46)
4. Raimundo Lira - PMDB/PB-Sim Waldemir Moka - PMDB/MS-Sim 44 ==> Simone Tebet - PMDB/MS-Sim Valdir Raupp - PMDB/RO-Sim 41 conf:(0.93) < lift:(1.26)> lev:(0.14) [8] conv:(2.84)
5. Raimundo Lira - PMDB/PB-Sim Simone Tebet - PMDB/MS-Sim 44 ==> Waldemir Moka - PMDB/MS-Sim RESULTADO-Aprovado 41 conf:(0.93) < lift:(1.23)> lev:(0.13) [7] conv:(2.66)
6. Raimundo Lira - PMDB/PB-Sim Waldemir Moka - PMDB/MS-Sim 44 ==> Simone Tebet - PMDB/MS-Sim RESULTADO-Aprovado 41 conf:(0.93) < lift:(1.23)> lev:(0.13) [7] conv:(2.66)
7. Simone Tebet - PMDB/MS-Sim RESULTADO-Aprovado 44 ==> Raimundo Lira - PMDB/PB-Sim Waldemir Moka - PMDB/MS-Sim 41 conf:(0.93) < lift:(1.23)> lev:(0.13) [7] conv:(2.66)
8. Waldemir Moka - PMDB/MS-Sim RESULTADO-Aprovado 44 ==> Raimundo Lira - PMDB/PB-Sim Simone Tebet - PMDB/MS-Sim 41 conf:(0.93) < lift:(1.23)> lev:(0.13) [7] conv:(2.66)
9. Simone Tebet - PMDB/MS-Sim 48 ==> Valdir Raupp - PMDB/RO-Sim Waldemir Moka - PMDB/MS-Sim 42 conf:(0.88) < lift:(1.21)> lev:(0.12) [7] conv:(1.89)
10. Valdir Raupp - PMDB/RO-Sim Waldemir Moka - PMDB/MS-Sim 42 ==> Simone Tebet - PMDB/MS-Sim 42 conf:(1) < lift:(1.21)> lev:(0.12) [7] conv:(7.24)
11. Simone Tebet - PMDB/MS-Sim 48 ==> Raimundo Lira - PMDB/PB-Sim Valdir Raupp - PMDB/RO-Sim Waldemir Moka - PMDB/MS-Sim 41 conf:(0.85) < lift:(1.21)> lev:(0.12) [7] conv:(1.76)
12. Raimundo Lira - PMDB/PB-Sim Valdir Raupp - PMDB/RO-Sim Waldemir Moka - PMDB/MS-Sim 41 ==> Simone Tebet - PMDB/MS-Sim 41 conf:(1) < lift:(1.21)> lev:(0.12) [7] conv:(7.07)
13. Raimundo Lira - PMDB/PB-Sim 47 ==> Simone Tebet - PMDB/MS-Sim Valdir Raupp - PMDB/RO-Sim 42 conf:(0.89) < lift:(1.21)> lev:(0.12) [7] conv:(2.83)
14. Simone Tebet - PMDB/MS-Sim Valdir Raupp - PMDB/RO-Sim 43 ==> Raimundo Lira - PMDB/PB-Sim 42 conf:(0.98) < lift:(1.21)> lev:(0.12) [7] conv:(4.08)
15. Raimundo Lira - PMDB/PB-Sim 47 ==> Valdir Raupp - PMDB/RO-Sim Waldemir Moka - PMDB/MS-Sim 41 conf:(0.87) < lift:(1.2)> lev:(0.12) [6] conv:(1.85)
16. Valdir Raupp - PMDB/RO-Sim Waldemir Moka - PMDB/MS-Sim 42 ==> Raimundo Lira - PMDB/PB-Sim 41 conf:(0.98) < lift:(1.2)> lev:(0.12) [6] conv:(3.98)
17. Raimundo Lira - PMDB/PB-Sim 47 ==> Simone Tebet - PMDB/MS-Sim Valdir Raupp - PMDB/RO-Sim Waldemir Moka - PMDB/MS-Sim 41 conf:(0.87) < lift:(1.2)> lev:(0.12) [6] conv:(1.85)
18. Simone Tebet - PMDB/MS-Sim Valdir Raupp - PMDB/RO-Sim Waldemir Moka - PMDB/MS-Sim 42 ==> Raimundo Lira - PMDB/PB-Sim 41 conf:(0.98) < lift:(1.2)> lev:(0.12) [6] conv:(3.98)
19. Raimundo Lira - PMDB/PB-Sim Valdir Raupp - PMDB/RO-Sim 43 ==> Simone Tebet - PMDB/MS-Sim Waldemir Moka - PMDB/MS-Sim 41 conf:(0.95) < lift:(1.2)> lev:(0.12) [6] conv:(2.97)
20. Simone Tebet - PMDB/MS-Sim 46 ==> Raimundo Lira - PMDB/PB-Sim Valdir Raupp - PMDB/RO-Sim 41 conf:(0.89) < lift:(1.2)> lev:(0.12) [6] conv:(1.98)
21. Simone Tebet - PMDB/MS-Sim 48 ==> Raimundo Lira - PMDB/PB-Sim Waldemir Moka - PMDB/MS-Sim 43 conf:(0.9) < lift:(1.18)> lev:(0.11) [6] conv:(1.93)
22. Raimundo Lira - PMDB/PB-Sim Simone Tebet - PMDB/MS-Sim 44 ==> Waldemir Moka - PMDB/MS-Sim 43 conf:(0.98) < lift:(1.18)> lev:(0.11) [6] conv:(3.79)
23. Waldemir Moka - PMDB/MS-Sim 48 ==> Raimundo Lira - PMDB/PB-Sim Simone Tebet - PMDB/MS-Sim 43 conf:(0.9) < lift:(1.18)> lev:(0.11) [6] conv:(1.93)
24. Raimundo Lira - PMDB/PB-Sim Waldemir Moka - PMDB/MS-Sim 44 ==> Simone Tebet - PMDB/MS-Sim 43 conf:(0.98) < lift:(1.18)> lev:(0.11) [6] conv:(3.79)
25. Simone Tebet - PMDB/MS-Sim 48 ==> Waldemir Moka - PMDB/MS-Sim RESULTADO-Aprovado 43 conf:(0.9) < lift:(1.18)> lev:(0.11) [6] conv:(1.93)
26. Waldemir Moka - PMDB/MS-Sim 48 ==> Simone Tebet - PMDB/MS-Sim RESULTADO-Aprovado 43 conf:(0.9) < lift:(1.18)> lev:(0.11) [6] conv:(1.93)
27. Simone Tebet - PMDB/MS-Sim RESULTADO-Aprovado 44 ==> Waldemir Moka - PMDB/MS-Sim 43 conf:(0.98) < lift:(1.18)> lev:(0.11) [6] conv:(3.79)
28. Waldemir Moka - PMDB/MS-Sim RESULTADO-Aprovado 44 ==> Simone Tebet - PMDB/MS-Sim 43 conf:(0.98) < lift:(1.18)> lev:(0.11) [6] conv:(3.79)
29. Simone Tebet - PMDB/MS-Sim 48 ==> Raimundo Lira - PMDB/PB-Sim Valdir Raupp - PMDB/RO-Sim 42 conf:(0.88) < lift:(1.18)> lev:(0.11) [6] conv:(1.77)
30. Raimundo Lira - PMDB/PB-Sim Valdir Raupp - PMDB/RO-Sim 43 ==> Simone Tebet - PMDB/MS-Sim 42 conf:(0.98) < lift:(1.18)> lev:(0.11) [6] conv:(3.71)
```

Figura: Resultado da associação tendo como base as matérias e os senadores no período pré-impeachment.

```
=== Run information ===
```

```
Scheme:      weka.associations.Apriori -N 100 -T 1 -C 1.1 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1  
Relation:    1.3.2-weka.filters.unsupervised.attribute.Remove-R1,9,15,24-26,37,40,61,80,88,103  
Instances:   37  
Attributes:  93
```

Figura: Cabeçalho dos resultados da associação tendo como base as matérias e os senadores no período pós-impeachment.

Análise por associação dos senadores no período Pós-Impeachment

Foram utilizadas 37 instâncias (votações que ocorreram em plenário) com a participação de 93 senadores como ilustra a figura e essa análise trouxe fortes relações de associação entre os seguintes senadores conforme a figura 18:

- Benedito de Lira - PP/AL e Ivo Cassol - PP/RO;
- José Agripino - DEM/RN e Benedito de Lira - PP/AL;
- Benedito de Lira - PP/AL e Paulo Bauer - PSDB/SC;
- Benedito de Lira - PP/AL e Flexa Ribeiro - PSDB/PA.

Resultados

Apriori

Minimum support: 0.75 (28 instances)

Minimum metric <lift>: 1.1

Number of cycles performed: 5

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10

Size of set of large itemsets L(2): 16

Size of set of large itemsets L(3): 9

Size of set of large itemsets L(4): 1

Best rules found:

```
1. Benedito de Lira - PP/AL-Sim 30 ==> Ivo Cassol - PP/RO-Sim 29      conf:(0.97) < lift:(1.23)> lev:(0.15) [5] conv:(3.24)
2. Ivo Cassol - PP/RO-Sim 29 ==> Benedito de Lira - PP/AL-Sim 29      conf:(1) < lift:(1.23)> lev:(0.15) [5] conv:(5.49)
3. Benedito de Lira - PP/AL-Sim 30 ==> Flexa Ribeiro - PSDB/PA-Sim Ivo Cassol - PP/RO-Sim 29      conf:(0.97) < lift:(1.23)> lev:(0.15) [5] conv:(3.24)
4. Benedito de Lira - PP/AL-Sim Flexa Ribeiro - PSDB/PA-Sim 30 ==> Ivo Cassol - PP/RO-Sim 29      conf:(0.97) < lift:(1.23)> lev:(0.15) [5] conv:(3.24)
5. Ivo Cassol - PP/RO-Sim 29 ==> Benedito de Lira - PP/AL-Sim Flexa Ribeiro - PSDB/PA-Sim 29      conf:(1) < lift:(1.23)> lev:(0.15) [5] conv:(5.49)
6. Flexa Ribeiro - PSDB/PA-Sim Ivo Cassol - PP/RO-Sim 29 ==> Benedito de Lira - PP/AL-Sim 29      conf:(1) < lift:(1.23)> lev:(0.15) [5] conv:(5.49)
7. Benedito de Lira - PP/AL-Sim 30 ==> Ivo Cassol - PP/RO-Sim RESULTADO-Aprovado 28      conf:(0.93) < lift:(1.23)> lev:(0.14) [5] conv:(2.43)
8. Ivo Cassol - PP/RO-Sim RESULTADO-Aprovado 28 ==> Benedito de Lira - PP/AL-Sim 28      conf:(1) < lift:(1.23)> lev:(0.14) [5] conv:(5.3)
9. Benedito de Lira - PP/AL-Sim 30 ==> Flexa Ribeiro - PSDB/PA-Sim Ivo Cassol - PP/RO-Sim RESULTADO-Aprovado 28      conf:(0.93) < lift:(1.23)> lev:(0.14) [5] conv:(2.43)
10. Benedito de Lira - PP/AL-Sim Flexa Ribeiro - PSDB/PA-Sim 30 ==> Ivo Cassol - PP/RO-Sim RESULTADO-Aprovado 28      conf:(0.93) < lift:(1.23)> lev:(0.14) [5] conv:(2.43)
11. Ivo Cassol - PP/RO-Sim RESULTADO-Aprovado 28 ==> Benedito de Lira - PP/AL-Sim Flexa Ribeiro - PSDB/PA-Sim 28      conf:(1) < lift:(1.23)> lev:(0.14) [5] conv:(5.3)
12. Flexa Ribeiro - PSDB/PA-Sim Ivo Cassol - PP/RO-Sim RESULTADO-Aprovado 28 ==> Benedito de Lira - PP/AL-Sim 28      conf:(1) < lift:(1.23)> lev:(0.14) [5] conv:(5.3)
13. Ivo Cassol - PP/RO-Sim 29 ==> Benedito de Lira - PP/AL-Sim RESULTADO-Aprovado 28      conf:(0.97) < lift:(1.23)> lev:(0.14) [5] conv:(3.14)
14. Benedito de Lira - PP/AL-Sim RESULTADO-Aprovado 29 ==> Ivo Cassol - PP/RO-Sim 28      conf:(0.97) < lift:(1.23)> lev:(0.14) [5] conv:(3.14)
15. Ivo Cassol - PP/RO-Sim 29 ==> Benedito de Lira - PP/AL-Sim Flexa Ribeiro - PSDB/PA-Sim RESULTADO-Aprovado 28      conf:(0.97) < lift:(1.23)> lev:(0.14) [5] conv:(3.14)
16. Flexa Ribeiro - PSDB/PA-Sim Ivo Cassol - PP/RO-Sim 29 ==> Benedito de Lira - PP/AL-Sim RESULTADO-Aprovado 28      conf:(0.97) < lift:(1.23)> lev:(0.14) [5] conv:(3.14)
17. Benedito de Lira - PP/AL-Sim RESULTADO-Aprovado 29 ==> Flexa Ribeiro - PSDB/PA-Sim Ivo Cassol - PP/RO-Sim 28      conf:(0.97) < lift:(1.23)> lev:(0.14) [5] conv:(3.14)
18. Benedito de Lira - PP/AL-Sim Flexa Ribeiro - PSDB/PA-Sim RESULTADO-Aprovado 29 ==> Ivo Cassol - PP/RO-Sim 28      conf:(0.97) < lift:(1.23)> lev:(0.14) [5] conv:(3.14)
19. Benedito de Lira - PP/AL-Sim 30 ==> José Agripino - DEM/RN-Sim 28      conf:(0.93) < lift:(1.19)> lev:(0.12) [4] conv:(2.16)
20. José Agripino - DEM/RN-Sim 29 ==> Benedito de Lira - PP/AL-Sim 28      conf:(0.97) < lift:(1.19)> lev:(0.12) [4] conv:(2.74)
21. Benedito de Lira - PP/AL-Sim 30 ==> Paulo Bauer - PSDB/SC-Sim 28      conf:(0.93) < lift:(1.19)> lev:(0.12) [4] conv:(2.16)
22. Paulo Bauer - PSDB/SC-Sim 29 ==> Benedito de Lira - PP/AL-Sim 28      conf:(0.97) < lift:(1.19)> lev:(0.12) [4] conv:(2.74)
23. Benedito de Lira - PP/AL-Sim 30 ==> Flexa Ribeiro - PSDB/PA-Sim José Agripino - DEM/RN-Sim 28      conf:(0.93) < lift:(1.19)> lev:(0.12) [4] conv:(2.16)
24. Benedito de Lira - PP/AL-Sim Flexa Ribeiro - PSDB/PA-Sim 30 ==> José Agripino - DEM/RN-Sim 28      conf:(0.93) < lift:(1.19)> lev:(0.12) [4] conv:(2.16)
25. José Agripino - DEM/RN-Sim 29 ==> Benedito de Lira - PP/AL-Sim Flexa Ribeiro - PSDB/PA-Sim 28      conf:(0.97) < lift:(1.19)> lev:(0.12) [4] conv:(2.74)
26. Flexa Ribeiro - PSDB/PA-Sim José Agripino - DEM/RN-Sim 29 ==> Benedito de Lira - PP/AL-Sim 28      conf:(0.97) < lift:(1.19)> lev:(0.12) [4] conv:(2.74)
27. Benedito de Lira - PP/AL-Sim 30 ==> Flexa Ribeiro - PSDB/PA-Sim Paulo Bauer - PSDB/SC-Sim 28      conf:(0.93) < lift:(1.19)> lev:(0.12) [4] conv:(2.16)
28. Benedito de Lira - PP/AL-Sim Flexa Ribeiro - PSDB/PA-Sim 30 ==> Paulo Bauer - PSDB/SC-Sim 28      conf:(0.93) < lift:(1.19)> lev:(0.12) [4] conv:(2.16)
29. Paulo Bauer - PSDB/SC-Sim 29 ==> Benedito de Lira - PP/AL-Sim Flexa Ribeiro - PSDB/PA-Sim 28      conf:(0.97) < lift:(1.19)> lev:(0.12) [4] conv:(2.74)
30. Flexa Ribeiro - PSDB/PA-Sim Paulo Bauer - PSDB/SC-Sim 29 ==> Benedito de Lira - PP/AL-Sim 28      conf:(0.97) < lift:(1.19)> lev:(0.12) [4] conv:(2.74)
```

Figura: Resultado da associação tendo como base as matérias e os senadores no período pós-impeachment.

- É um método de divisão de um agrupamento de dados (registros, tuplas, instâncias) em vários outros grupos denominados clusters;
- Permite que um usuário faça grupos de dados para determinar os padrões dos dados;
- Cada atributo no conjunto de dados será usado para analisar os dados.

Hierarchical Clustering

- Gera uma sequência de agrupamento *Top-Bottom*;
- Não é necessário passar a quantidade de clusters como parâmetro;
- Atribuí-se todas as observações a um único cluster e particiona-se o cluster em dois clusters menos semelhantes;
- Recursão até que haja um cluster para cada análise.

Análise por *Clustering* dos partidos dos senadores no período Pré-Impeachment

Os dados em análise estão dispostos nos ramos do dendrograma conforme ilustra a figura 19 em que foram analisadas um total de 58 matérias:

- PDT – PMDB;
- PT - PTB;
- DEM – PSC;
- PCdoB – PSOL.

Resultados

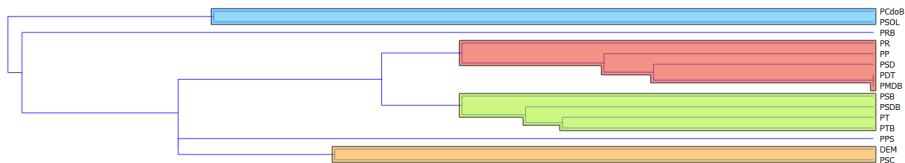


Figura: Dendrograma gerado tendo como referência os partidos dos senadores.

Análise por *Clustering* dos partidos dos senadores no período Pós-Impeachment

Os dados em análise estão dispostos nos ramos do dendrograma conforme ilustra a figura 20 em que foram analisadas um total de 37 matérias:

- PDT – PMDB;
- DEM - PSDB;
- PP - PR;
- PSC - PTB;
- PCdoB - PSOL.

Análise por *Clustering* dos estados dos senadores no período Pré-Impeachment

Os dados em análise estão dispostos nos ramos do dendrograma conforme ilustra a figura 21 em que foram analisadas um total de 58 matérias:

- BA - MA;
- AM - PI;
- CE - TO;
- PA - RN;
- GO - MS.

Resultados

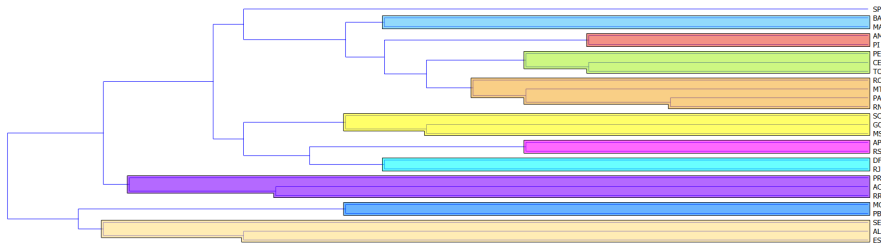


Figura: Dendrograma gerado tendo como referência os estados dos senadores.

Análise por *Clustering* dos estados dos senadores no período Pós-Impeachment

Os dados em análise estão dispostos nos ramos do dendrograma conforme ilustra a figura 22 em que foram analisadas um total de 37 matérias:

- PR - SE;
- PA - RR;
- RJ - RN;
- MT - RO;
- ES - PI.

Resultados

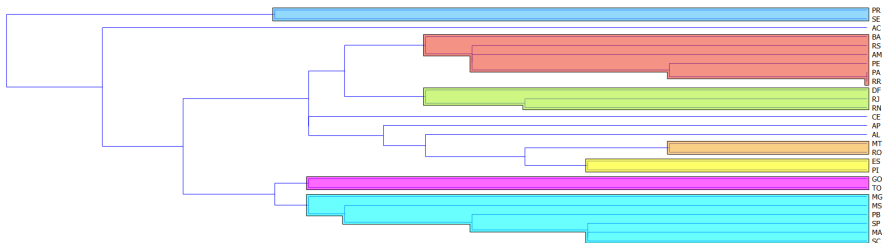


Figura: Dendrograma gerado tendo como referência os estados dos senadores.

Conclusão

- Os senadores dos partidos do DEM, PSD, PP, PTB, PSDB e PT foram os senadores que trouxeram a maior taxa de coincidência de votos;
- O DEM foi o partido que mais divergiu do que convergiu com os votos dos senadores do partido do PT.

Trabalhos Futuros

- Utilizar *scripts* para obter um escopo maior de dados.

Obrigado!

Referências I



J. Han and M. Kamber.

Data mining : concepts and techniques.

AKaufmann, San Francisco, 2005.



T. Pang-Ning e V. Kumar M. Steinbach.

Introduction to data mining.

first edition, 2005.



P; STADLER R; JAAPVERHEES; ZANASI A CABENA, P;
HADJINIAN.

Discovering data mining:

From concept to implementation, 1998.



I. H. Witten, E. Frank and M. A. Hell.

Data Mining: Practical Machine Learning Tools and Technique.

Morgan Kaufmann, 2011.



BRAGA. Everaldo Miranda

Mineração de Dados de Posição Geográfica e Compras

Departamento de Ciência da Computação, Universidade de Brasília,
Brasília, 2012.