



ISSN: 2447-5580

Disponível em: <http://periodicos.ufes.br/BJPE/index>



Campus São Mateus

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

ARTIGO ORIGINAL

OPEN ACCESS

MINERAÇÃO DE DADOS NA BASE DE DADOS ABERTA DA CÂMARA LEGISLATIVA FEDERAL BRASILEIRA: ÊNFASE NA ANÁLISE DOS DADOS DA LEGISLATURA 54 (2011-2013)

DATA MINING IN THE OPEN DATABASE OF THE BRAZILIAN FEDERAL LEGISLATIVE CHAMBER: EMPHASIS IN THE ANALYSIS OF DATA IN LEGISLATURE 54 (2011-2013)

Vanessa Lourenço Mattede Formigoni¹; Icaro Henrique Honorato²; Silvia das Dores Rissino³

^{1,2,3}Departamento de Computação e Eletrônica do Centro Universitário Norte do Espírito Santo da Universidade Federal do Espírito Santo, Rodovia BR 101 Norte, Km. 60, Bairro Litorâneo, CEP 29932-540, São Mateus.

*vanessa.formigoni@hotmail.com

ARTIGO INFO.

Recebido em: 31/10/2018

Aprovado em: 14/11/2018

Disponibilizado em: 15/12/2018

PALAVRAS-CHAVE:

Câmara Federal, Dados Abertos, Mineração de Dados, KDD, WEKA.

KEYWORDS:

Federal Chamber, Open Data, Data Mining, KDD, WEKA

Copyright © 2018, FORMIGONI, HONORATO & RISSINO. Esta obra está sob uma Licença Creative Commons Atribuição-Uso.

*Autor Correspondente: Vanessa Lourenço Mattede Formigoni.

RESUMO

A descoberta de conhecimentos em bases de dados tem por objetivo encontrar padrões e relações ocultas em grandes volumes de dados através das suas etapas operacionais. O objetivo deste trabalho é aplicar as técnicas de mineração de dados na base de dados

aberta da câmara legislativa federal brasileira e encontrar padrões nos gastos gerados pelos partidos políticos na legislatura 54 (2011-2013) usando ferramenta WEKA. Na mineração de dados utiliza-se dois algoritmos: c4.5 (algoritmo supervisionado) e o *a priori* (algoritmo não supervisionado). É possível inferir que o algoritmo não supervisionado *a priori* apresenta os melhores resultados para o problema proposto em relação ao algoritmo supervisionado c4.5.

ABSTRACT

The discovery of databases aims to find patterns and alliances in large volumes of data for the operational stages. The objective of this work is to apply data mining techniques in the database of the Brazilian Federal Legislative Chamber and to find patterns in the expenditures generated by political parties in the 54 (2011-2013) legislature using WEKA tool. In data mining, use algorithms: c4.5 (supervised algorithm) and *a priori* (unsupervised algorithm). It is possible to infer that the *a priori* unsupervised algorithm presents the results of results for the problem with respect to the c4.5 supervised algorithm.

Citação (APA): FORMIGONI, V. L. M., HONORATO, I. H. & RISSINO, S. D. D. (2018). Mineração de dados na base de dados aberta da Câmara legislativa Federal Brasileira: Ênfase na análise dos dados da legislatura 54 (2011-2013). *Brazilian Journal of Production Engineering*, 4(4): 156-170.

Brazilian Journal of Production Engineering, São Mateus, Vol. 4, N.º 4, p. 156-170. (2018). Editora CEUNES/DETEC.

Disponível em: <http://periodicos.ufes.br/BJPE>

1. INTRODUÇÃO

O Poder Legislativo desempenha papel imprescindível perante a sociedade do país, visto que desempenha três funções básicas para a consolidação da democracia: representar o povo brasileiro, legislar sobre os assuntos de interesse nacional e fiscalizar a aplicação dos recursos públicos. A Câmara dos Deputados, 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 (FEDERAL, Acesso em: 23 jun. 2017a). “Um deputado pode propor novas leis, alterações ou revogação de leis existentes, incluindo a própria Constituição. As propostas são votadas pelo plenário quando for o caso. [...]” (FEDERAL, Acesso em: 20 jul. 2017b). Proposição é a denominação genérica de toda matéria submetida à apreciação do Senado, da Câmara ou do Congresso Nacional. São algumas proposições: propostas de emenda à Constituição (PEC); projetos de lei (PL), projeto de lei complementar (PLC), projeto de decreto legislativo (PDL), projeto de resolução (PRC), requerimentos (REC), medidas provisórias (PV), indicações (INC), emendas (EMD) entre outros (NOTÍCIAS, Acesso em: 2 jul. 2017).

No Brasil, atualmente a Câmara Federal é composta por 513 deputados federais (FEDERAL, Acesso em: 2 jul. 2017c), que geram pra União um custo orçamentário de aproximadamente 2 milhões de reais por ano por parlamentar (DE MINAS, Acesso em: 20 jul. 2017). Este valor elevado se deve não apenas aos altos salários, mas aos auxílios denominados Cota para Exercício da Atividade Parlamentar (CEAP), a qual os mesmos têm direitos mensais. Os benefícios incluem: passagens aéreas, telefonia, manutenção de escritórios, serviços postais, fornecimento de alimentação do parlamentar, hospedagem (exceto do parlamentar no Distrito Federal), despesas com locomoção, combustíveis e lubrificantes, serviços de segurança, contratação (para fins de apoio ao exercício do mandato parlamentar), divulgação da atividade parlamentar, participação do parlamentar em cursos, palestras, seminários, simpósios, congressos ou eventos congêneres, complementação do auxílio-moradia.

Todas as informações referentes a Câmara Federal incluindo gastos gerados pelos Deputados Federais, estão disponíveis a toda a população através do sítio eletrônico <http://www2.camara.leg.br/> onde possui uma aba denominada “Transparência” aonde

apresentam-se os custos associados a cada congressista. Tais informações podem ser avaliadas e processadas por qualquer cidadão a fim de fiscalizar a devida utilização dos recursos públicos.

O objetivo deste trabalho consiste em encontrar padrões nos gastos gerados pelos partidos políticos na legislatura 54 (2011-2013) usando a ferramenta de mineração de dados WEKA. O objetivo específico que conduzirá ao objetivo geral é a aplicação do KDD (Descoberta de Conhecimento em Base de Dados) na base de dados aberta da Câmara Legislativa Federal Brasileira.

2. DESCOBERTA DE CONHECIMENTOS EM BASE DE DADOS – KDD

Os constantes avanços na área da Tecnologia da Informação têm viabilizado o armazenamento de grandes e múltiplas bases de dados (BRAGA, 2005). Informações são geradas e arquivadas a todo o instante e é possível aproveitar-se de todo este conhecimento para benefício empresarial ou em prol da sociedade.

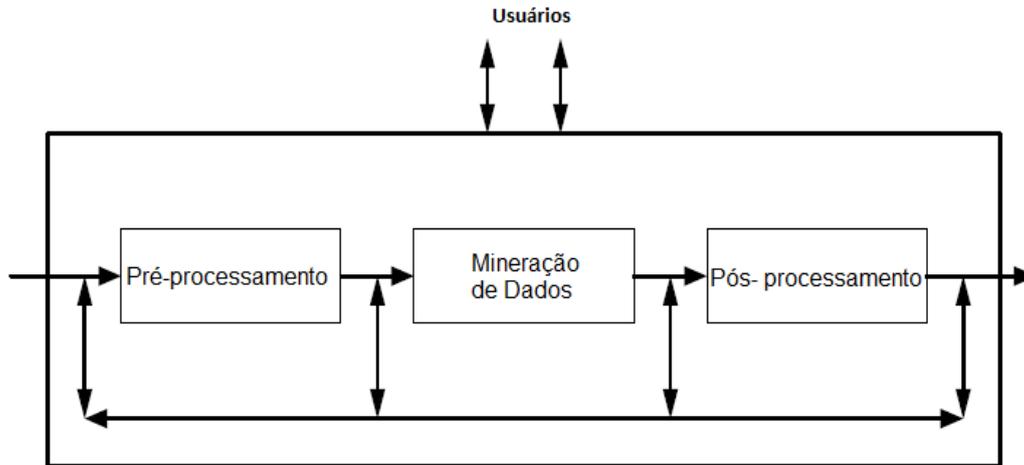
A Descoberta de Conhecimento em Base de Dados, é um processo que possibilita encontrar padrões e relacionamentos em grandes volumes de dados. KDD é um processo de várias etapas, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados (GOLDSCHMIDT; PASSOS, 2015).

O processo de KDD é composto por três etapas operacionais: Pré-processamento, Mineração de Dados e Pós-processamento. A primeira etapa compreende as funções relacionadas à captação, à organização e ao tratamento dos dados e tem como objetivo a preparação dos dados para os algoritmos da etapa seguinte. Na Mineração de Dados, é realizada a busca efetiva por conhecimentos úteis e, são definidas as técnicas e os algoritmos a serem utilizados no problema em questão. A última etapa abrange o tratamento do conhecimento obtido com o objetivo de viabilizar o conhecimento descoberto (GOLDSCHMIDT; PASSOS, 2015). A Fig.1 mostra a relação entre Pré-processamento, Mineração de Dados e Pós-processamento.

No pré-processamento estão relacionadas todas as atividades referentes ao tratamento dos dados antes do processamento. Algumas das tarefas do Pré-processamento são: limpeza de informações ausentes, tratamento de valores distorcidos, enriquecimento, no qual consiste em agregar mais informações aos registros existentes para que estes forneçam mais elementos

para o processo de descoberta de conhecimentos (GOLDSCHMIDT; PASSOS, 2015).

Figura 1 - Relação entre as três etapas do KDD.



Fonte: Adaptado de (GOLDSCHMIDT; PASSOS, 2015).

A Mineração de Dados é uma das alternativas mais eficientes para extrair conhecimento a partir de grandes volumes de dados, descobrir relações ocultas e padrões, gerando regras para prever e correlacionar dados, auxiliando nas tomadas de decisões.

A etapa de pós-processamento, envolve a análise, interpretação e visualização do modelo de conhecimentos gerado pela etapa de Mineração de Dados.

3. AMBIENTE DE DADOS

Dados são considerados abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, a exigências que visem preservar sua proveniência e sua abertura (FEDERAL, Acesso em: 2 jul. 2017d).

No sítio da Câmara Federal Brasileira estão presentes os dados referentes às despesas pela Cota para Exercício da Atividade Parlamentar, separados por ano, de 2009 a 2017, com extensões XML, JSON, CSV, XLSX e ODS. Nesta pesquisa, foram escolhidos os dados referentes às Cotas Parlamentares relativas aos anos de 2011 e 2013 segundo a legislatura 54, ano inicial e final da última legislatura concluída.

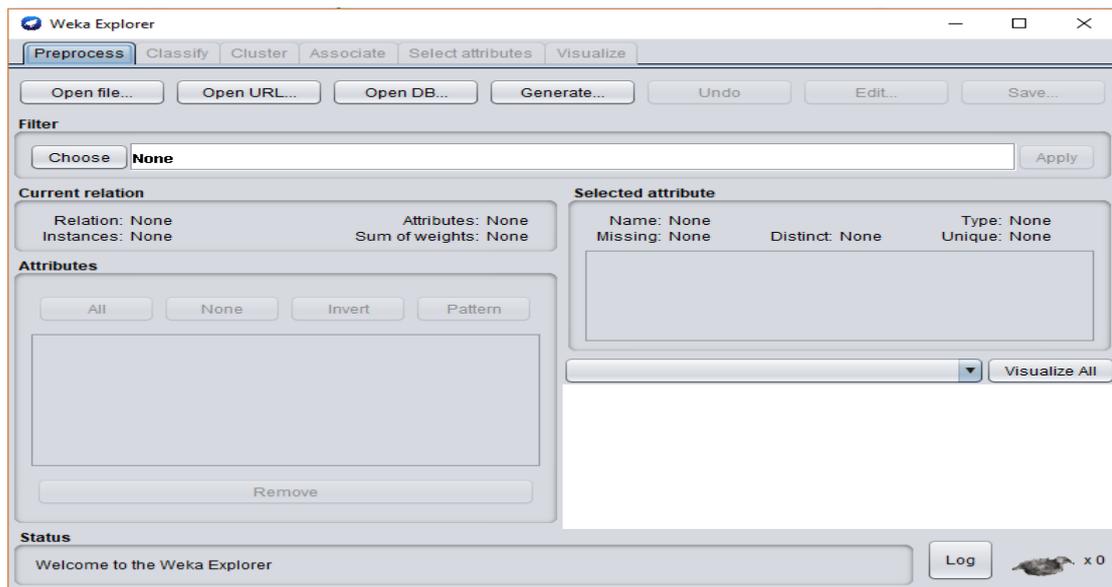
4. METODOLOGIA

Para a realização das etapas operacionais do KDD, foi utilizado principalmente a ferramenta computacional WEKA com seus filtros para a etapa de Pré-processamento e os algoritmos de Mineração de Dados supervisionado J48 e não supervisionado *Apriori*. Para a etapa de Pós-processamento foi gerada um grafo para auxiliar na visualização das principais regras construídas pela árvore de decisão utilizando o software Graphviz versão 2.38.

4.1. FERRAMENTA WAIKATO ENVIRONMENT FOR KNOWLEDGE ANALYSIS – WEKA

O WEKA é uma suíte composta por diversos algoritmos de aprendizagem de máquina para Mineração de Dados. Os algoritmos podem ser aplicados diretamente na base de dados ou chamados por um código Java. O WEKA possui também ferramentas para Pré-processamento de dados. A Fig.2 mostra a ferramenta computacional WEKA 3.8.1.

Figura 2 - Ferramenta computacional WEKA.



Fonte: Próprio autor.

A ferramenta de Mineração de Dados WEKA foi escolhida para este projeto por ser um software livre onde o download pode ser feito diretamente da página do desenvolvedor, conter várias versões de algoritmos de Mineração de Dados e ferramentas para realização do Pré-processamento.

4.2 PRÉ-PROCESSAMENTO

Neste trabalho, o pré-processamento foi realizado em duas etapas: a primeira etapa com auxílio do software livre LibreOffice versão: 5.4.1.2 e a segunda etapa no software livre

WEKA versão 3.8.1 (aba Preprocess).

4.2.1 PRIMEIRA ETAPA – LIBREOFFICE CALC

Nesta etapa utiliza-se o LibreOffice Calc, que é uma planilha de cálculo da Suíte do LibreOffice, para remover todas as instâncias das tabelas Ano-2011.ods e Ano-2013.ods que não pertencesse ao conjunto da legislatura 54. A segunda tarefa foi a remoção de todos os assentos e a substituição pela respectiva letra sem o acento. Haviam atributos que representavam valores numéricos e que eram separados por vírgula e devido à grande quantidade de valores distintos, o WEKA transformou este atributo em um tipo string colocando o valor entre aspas simples. Assim sendo, os campos que continham valores numéricos e que eram separados por vírgula foram substituídos por ponto e, ao ser lido pelo WEKA, esses atributos passavam a ser identificados pelo tipo numeric. Outra alteração nos valores numéricos foi remoção do sinal diacrítico hífen (-), alterando os valores negativos para positivos. Isto foi realizado pois, acredita-se que não seja correto um valor negativo em um documento oficial e que, provavelmente tenha sido erro de digitação. O arquivo gerado foi salvo no formato CVS (Ano-2011.csv e Ano-2013.csv), pois é um dos formatos aceitos como entrada pelo WEKA.

4.2.2 SEGUNDA ETAPA - WEKA

Carregados os arquivos Ano-2011.csv e Ano-2013.csv o pré-processamento no WEKA, ocorreu na aba Preprocess, onde pode ser analisado cada atributo separadamente. Com isso, foi possível remover colunas cujos dados faltantes eram predominantes, colunas que eram equivalentes entre si e colunas que apresentavam valores com pouca divergência. Os arquivos gerados possuem a extensão arff (extensão própria do WEKA) e foram salvos como Ano-2011_weka.arff e Ano-2013_weka.arff.

O objetivo da etapa de pré-processamento foi fazer uma limpeza dos dados removendo todos atributos que não tinham significado a base de dados. Com isso houve a redução de 34 atributos para 05 atributos.

4.3 MINERAÇÃO DE DADOS

A Mineração de Dados é uma das alternativas mais eficientes para extrair conhecimento a partir de grandes volumes de dados, descobrir relações ocultas e padrões, gerando regras para prever e correlacionar dados, para auxiliar nas tomadas de decisões (GALVÃO; MARIN, 2009).

A literatura propõe muitos algoritmos e tarefas que são utilizados de acordo com os objetivos de estudo, a fim de obter uma resposta para os problemas em questão. As técnicas de aprendizagem podem ser divididas em duas categorias (CAMILO; SILVA, 2009): algoritmos de aprendizado não supervisionados e algoritmos de aprendizado supervisionados.

4.3.1 APRENDIZADO SUPERVISIONADO

No aprendizado supervisionado o conjunto de dados possui uma variável pré-definida (classe) e os registros são categorizados ou rotulados em relação esta classe

4.3.2 APRENDIZADO NÃO SUPERVISIONADO

Nesta categoria o conjunto de dados não precisa de uma pré-categorização para os registros, ou seja, não é necessária uma variável alvo.

Para esta aplicação, foram escolhidas duas tarefas de Mineração de Dados, a tarefa supervisionada de classificação e a tarefa não supervisionada de descoberta de associações. A tarefa supervisionada de classificação foi escolhida pois desejava-se classificar o partido de acordo com o valor gasto, o período e o com que foi gasto. Segundo (GOLDSCHMIDT; PASSOS, 2005).

A classificação é a técnica mais importante e mais popular. O algoritmo C4.5 foi escolhido devido ter a construção da árvore de decisão e ao fato de ser mais seguro devido ao processo de poda, capaz de eliminar classificações avaliadas como irrelevantes (MACIEL et al., 2015).

A tarefa não supervisionada de descoberta de associações foi escolhida pois desejava-se encontrar gastos, períodos e partidos que ocorriam de forma simultaneamente e frequentes na base de dados. O algoritmo Apriori foi escolhido por ser um algoritmo clássico de Mineração de Dados (GOLDSCHMIDT; PASSOS, 2005).

Devido ao fato de que muitos parlamentares não constarem o nome na planilha de gastos, a análise deste trabalho será orientada a gastos gerados pelos partidos. Considerando ainda aproximadamente 58% dos gastos foram realizados por cinco (05) partidos, os demais partidos foram removidos do banco de dados a fim de que, as regras de associações e a árvore de decisão sejam mais condensadas e objetivas.

Houve a inclusão do atributo numTrimestre, cujo objetivo foi classificar os meses em que o gasto ocorreu por trimestres, isso se deve a observação de que, gastos mais elevados ocorriam com menos frequência e, quando agrupados, ofereciam maior suporte as regras de associações e melhor visualização na árvore de decisão do que quando considerados mensais.

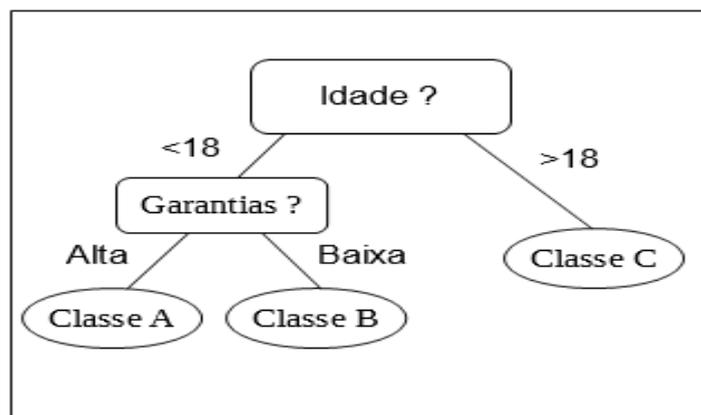
4.4 ALGORITMO C4.5

No WEKA, o algoritmo C4.5 é nomeado como J48. O algoritmo J48 é a recodificação do algoritmo C4.5, que, originalmente, é escrito na linguagem C, para a linguagem Java (LIBRELOTTO; MOZZAQUATRO, 2014).

A tarefa de classificação é uma tarefa supervisionada que emprega de um conjunto de instâncias previamente classificadas, as quais são utilizadas como base para construir um modelo capaz de classificar novas instâncias (MACIEL et al., 2015). Esta tarefa consiste em buscar uma função que permita associar corretamente cada registro de um bando de dados a uma única classe (GOLDSCHMIDT; PASSOS, 2015).

O algoritmo C4.5 é um algoritmo baseado em árvores de decisão. Uma árvore de decisão é um padrão que consiste em um conjunto hierarquicamente ordenado de conceitos, que iniciam em um nó raiz, sendo cada conceito dividido recursivamente em sub-conceitos no próximo nível mais inferior da árvore. Cada elemento terminal de uma árvore (conhecido como nó folha) está associado a um conceito objetivo (classe). (VIANNA et al., 2010). “Árvores de decisão constituem um dos principais e mais populares métodos de Mineração de Dados” (DE MORAES SOUSA; FIGUEIREDO, 2014). A Fig.3 apresenta um modelo de árvore de decisão.

Figura 3 - Modelo de árvore de decisão



Fonte: Adaptado de (CAMILO; SILVA, 2009).

4.5 DESCOBERTA DE ASSOCIAÇÕES (APRIORI)

A tarefa de descoberta de associações consiste em encontrar conjuntos de itens que ocorram simultaneamente e de forma frequente em uma base de dados gerando regras de associações do tipo $X \rightarrow Y$ onde X e Y são conjuntos tais que $X \cap Y \neq \emptyset$ (GOLDSCHMIDT; PASSOS,

2015)

O *Apriori* é um algoritmo não supervisionado clássico de mineração de regras de associações. Algoritmos como GSP, DHP, DIC, Eclat, MaxEclat e Clique foram baseados no funcionamento do *Apriori* (GOLDSCHMIDT; PASSOS, 2015). Desenvolvido por Agrawal, Imielinski e Swami, baseia-se na seguinte propriedade: se um conjunto de itens (k-itensets) é frequente, então todos os seus subconjuntos (k-1 itensets) também devem ser frequentes e, se um conjunto de itens é infrequente então todos os seus superconjuntos também serão. (CARVALHO, 2015).

5. ANALISE DOS RESULTADOS

A partir da árvore de decisão construída pelo algoritmo J48, pode-se inferir que em 2011:

1. As cotas referentes a manutenção de escritório de apoio a atividade parlamentar, combustíveis e lubrificantes, telefonia e serviço de segurança prestado por empresa especializada foram usadas predominantemente pelo PT independente da época e valor.
2. As cotas referentes a divulgação da atividade parlamentar e serviços postais foram usados em maior proporção pelo PMDB independente da época e valor.
3. A cota referente a fornecimento de alimentação do parlamentar quando de janeiro a setembro foi em maioria, gasta pelo PT, enquanto de outubro a dezembro pelo PSB independentemente do valor.
4. A cota referente a locação de veículos automotores ou fretamento de embarcações, independente da época, quando o valor era de até R\$7200.01, o partido PT era o de maior predominância e, quando o valor era de R\$7200.01 até R\$14400.09 era predominantemente do PMDB. Os demais valores não obtiveram ocorrências significativas.
5. A cota referente a consultorias, pesquisas e trabalhos técnicos quando o valor era de até R\$7200.01, o partido PT era o de maior predominância e, quando o valor era de R\$7200.01 até R\$14400.09 era predominantemente do PMDB. Os demais valores não obtiveram ocorrências significativas.
6. As cotas emissões de emissão de bilhetes aéreos e passagens aéreas se referem a mesma ação: comprar passagem aérea, apenas escrita de duas formas diferentes e

assim tratadas pelo WEKA. Entretanto elas foram analisadas e forma conjunta e, predominantemente o PT foi o partido quem mais usufruiu desta cota.

As demais cotas não citadas tiveram ocorrências insignificante perante o banco de dados.

Para o algoritmo Apriori pode-se analisar que:

1. As cotas para emissão de bilhete aéreo, telefonia e combustíveis e lubrificantes, estiverem quase 100% no intervalo de R\$ 0,00 até R\$ 7200,01;
2. Os partidos PSB, PT e PMDB tiveram seus gastos predominantemente no intervalo de R\$ 0,00 até R\$ 7200,01;
3. Cerca 24% e 34% dos gastos de R\$ 0,00 até R\$ 7200,01 foi realizado pelo PMDB e PT respectivamente;
4. As demais regras inferiram com grande precisão que durante o ano, a maioria dos gastos se mantiveram até R\$ 7200,01.

Para a base de dados Ano-2013_weka.arff e o algoritmo J48 pode-se inferir que em 2013:

1. As costas referentes a combustíveis e lubrificantes, fornecimento de alimentação do parlamentar, manutenção de escritório de apoio a atividade parlamentar, telefonia, serviço de taxi pedágio e estacionamento, assinatura de publicações, foram usadas predominantemente pelo PT independente da época e valor.
2. A cota referente a serviço de segurança prestado por empresa especializada foi gasta, predominantemente pelo PSD independente da época e valor.
3. A cota referente a divulgação da atividade parlamentar, quando o valor era de até R\$ 9701,00 o gasto foi predominantemente do PMDB. Para os demais valores, não houve ocorrência significativa para algum determinado partido.
4. A cota referente a hospedagem, exceto do parlamentar no distrito federal, quando de janeiro a setembro, foi utilizada, em maioria, pelo PT, já para os meses de outubro a dezembro, foi gasta predominantemente pelo PMDB.
5. A cota referente a locação de veículos automotores ou fretamento de embarcações, quando o valor era de até R\$9701,00 o gasto foi em maioria realizado pelo PT. Para os demais valores, não houve ocorrência significativa para algum determinado partido.
6. A cota referente a consultoria, pesquisas e trabalhos técnicos, quando o valor era de

até R\$19402,00 foi gasto predominantemente pelo PT. Para os demais valores, não houve ocorrência significativa para algum determinado partido.

7. A cota referente a emissão de bilhete aéreo foi utilizada predominantemente pelo PT. Já a cota passagens aéreas quando no valor de até R\$9701,00 foi gasta predominantemente pelo PMDB. Para os demais valores, não houve ocorrência significativa para algum determinado partido.

Para o algoritmo *Apriori* pode-se analisar que:

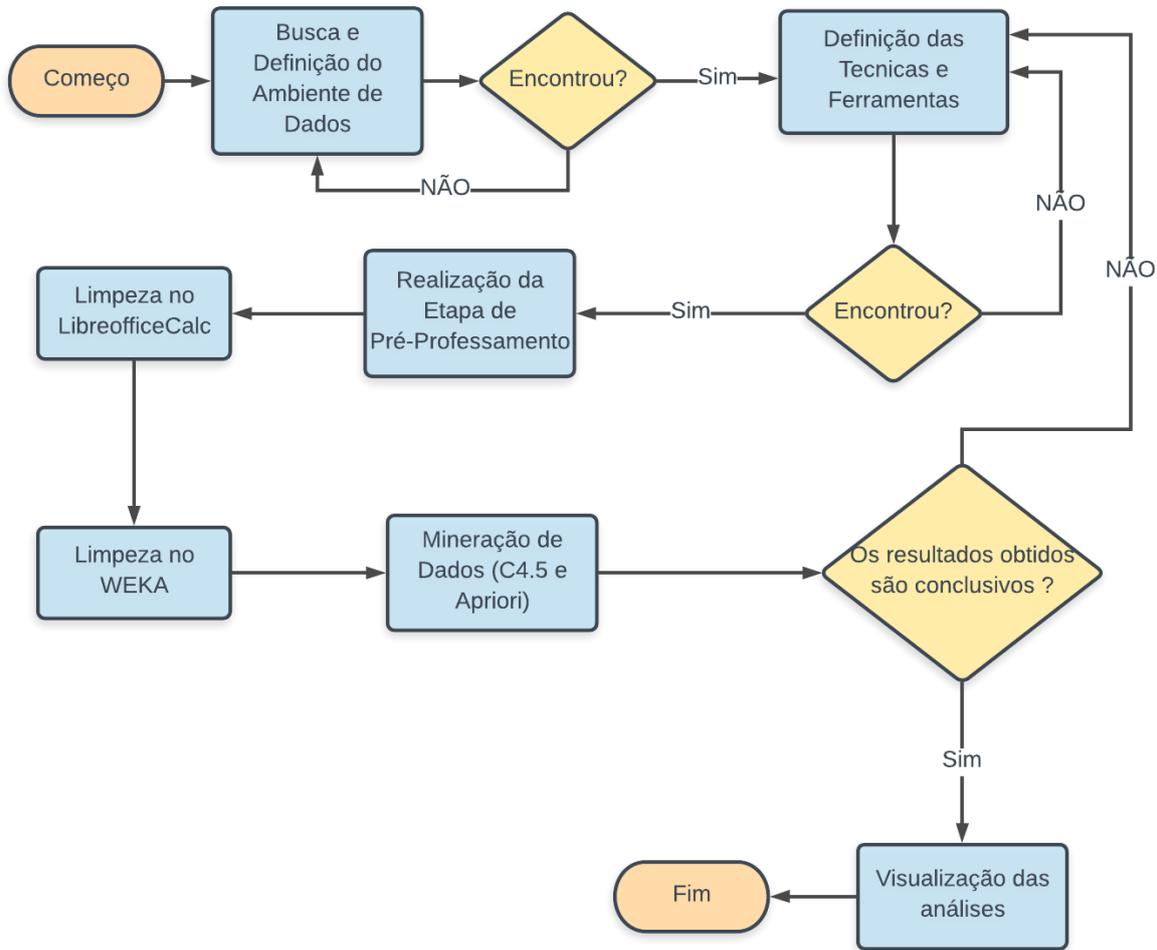
1. As cotas para emissão de bilhete aéreo, telefonia e combustíveis e lubrificantes, estiveram 100% no intervalo de R\$ 0,00 até R\$ 7200,01;
2. Os cinco partidos (PR, PSB, PSD, PT e PMDB) tiveram seus gastos predominantemente no intervalo de R\$ 0,00 até R\$ 7200,01;
3. Cerca de 36% dos gastos de R\$ 0,00 até R\$ 7200,01 foi realizado pelo PT;
4. Aproximadamente 28% dos gastos de R\$ 0,00 até R\$ 7200,01 foi para a emissão de bilhetes aéreos;
5. As demais regras inferiram com grande precisão que durante o ano, a maioria dos gastos se mantiveram em até R\$ 7200,01.

6.PÓS-PROCESSAMENTO

Os grafos gerados para os anos 2011 e 2013 a partir das principais regras do algoritmo C4.5 através da ferramenta Graphviz, estão no Apêndice deste artigo, representados respectivamente pelas Fig. 4 e Fig. 5.

Na Fig. 6 é apresentado um fluxograma das principais atividades realizadas nesta pesquisa.

Figura 6—Fluxograma de atividades



Fonte: Próprio autor.

7. CONCLUSÕES E TRABALHOS FUTUROS

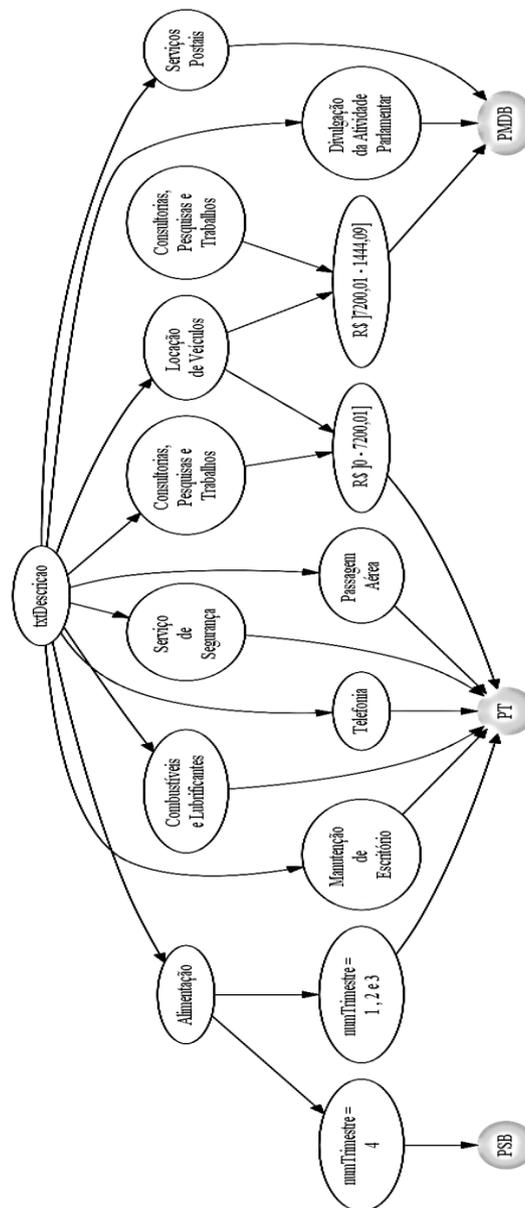
Com a análise dos resultados dos algoritmos finalizada, pode-se inferir que as regras de associações (algoritmo não supervisionado) obtiveram melhor desempenho em relação a árvore de decisão (algoritmo supervisionado) que alcançaram, em média, 36% de acerto. Acredita-se que a principal razão para o baixo acerto da árvore de decisão seja devido ao volume de dados apresentado, enquanto nos trabalhos relacionados a maior base de dados continha 4 mil instâncias, a base de dados da Câmara Federal possui em média 88 mil instâncias. Outro ponto relevante é devido à variabilidade dos dados por se tratar de um comportamento humano, que provavelmente não houve um padrão. O desempenho melhor do algoritmo *Apriori* pode concluir que, provavelmente, os algoritmos não supervisionados se enquadram melhor no problema proposto. Ainda pode-se destacar como possíveis causas para o baixo acerto o fato de o WEKA ser uma ferramenta de aprendizado de Mineração Dados e,

portanto, possuir os algoritmos já implementados, permitindo apenas alterações em alguns parâmetros de entrada, o que impossibilita ao analista da aplicação fazer modificações no código original e otimizações que melhorem o desempenho da Mineração de Dados.

Como propostas para trabalhos futuros, sugere-se a implementação dos algoritmos na linguagem Python e C e no software R fazendo uma comparação entre tempos de execução, taxa de acertos e algoritmos otimizados e não otimizados. Outra proposta é utilizar outra técnica de Mineração de Dados presente na biblioteca PANDAS do Python para verificar o desempenho em relação ao WEKA.

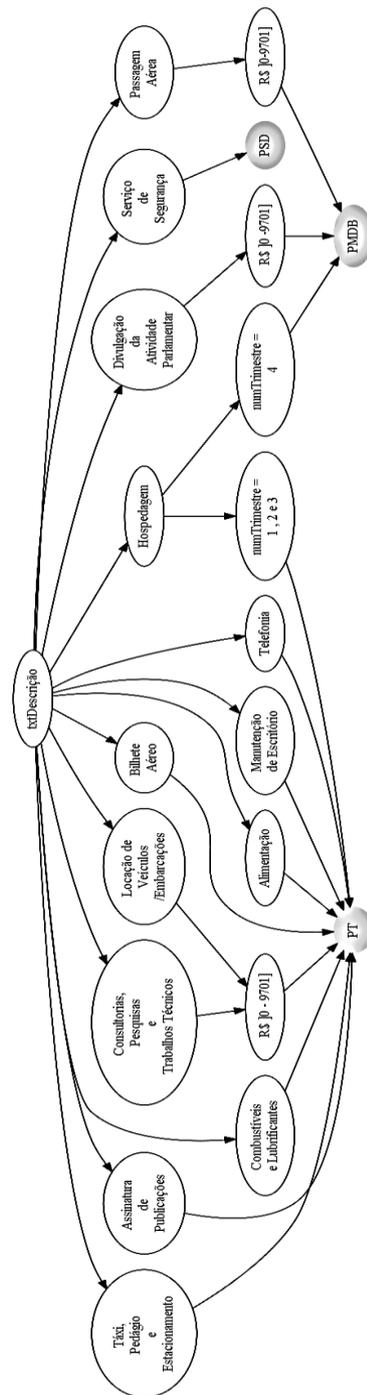
APÊNDICE

Figura 4 - Grafo para Ano 2011



Fonte: Próprio autor.

Figura 5 - Grafo para Ano 2013



Fonte: Próprio autor.

REFERENCIAS

BRAGA, L. P. V. Introdução à Mineração de Dados-2a edição: Edição ampliada e revisada. [s.l.] Editora E-papers, 2005.

CAMILO, C. O.; SILVA, J. C. DA. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. Universidade Federal de Goiás (UFG), p. 1–29, 2009.

CARVALHO, J. V. DE. Descoberta de conjuntos de itens frequentes com o modelo de programação MapReduce sobre contextos de incerteza. 2015.

DE MINAS, E. Deputado custa até R\$ 2 milhões por ano aos brasileiros. Disponível em: <http://www.em.com.br/app/noticia/politica/2017/01/23/interna_politica,841629/deput%0Aad-o-pode-custar-ate-r-2-milhoes-por-ano-aos-brasileiros.shtml%0A>. Acesso em: 20 jul. 2017.

DE MORAES SOUSA, M.; FIGUEIREDO, R. S. ANÁLISE DE CRÉDITO POR MEIO DE MINERAÇÃO DE DADOS: APLICAÇÃO EM COOPERATIVA DE CRÉDITO. Revista de Gestão da Tecnologia e Sistemas de Informação, v. 11, n. 2, p. 379–396, 2014.

FEDERAL, C. O Papel da Câmara dos Deputados. Disponível em: <<http://www2.camara.leg.br/a-camara/conheca/o-papel-da-camara-dos-deputados>>. Acesso em: 23 jun. 2017a.

FEDERAL, C. Saiba o que faz um deputado federal: atribuições, direitos, deveres e benefícios. Disponível em: <<http://www2.camara.leg.br/camaranoticias/noticias/POLITICA/474310-SAIBA-OQUE-%0AFAZ-UM-DEPUTADO-FEDERAL-ATRIBUICOES,-DIREITOS,-DEVERES-EBENEFICIOS.%0Ahtml%0A>>. Acesso em: 20 jul. 2017b.

FEDERAL, C. Quantos são e de que forma é definido o número de Deputados. Disponível em: <<http://www12.senado.leg.br/noticias/glossariolegislativo/%0Aproposicao%0A>>. Acesso em: 2 jul. 2017c.

FEDERAL, C. O que são dados abertos?. Disponível em: <<http://dados.gov.br/pagina/dados-abertos>>. Acesso em: 2 jul. 2017d.

GOLDSCHMIDT, R.; PASSOS, E. Data mining: um guia prático, conceitos, técnicas, ferramentas, orientações e aplicações. Editora Elsevier, 2.ed,2015.

LIBRELOTTO, S. R.; MOZZAQUATRO, P. M. Análise dos algoritmos de mineração J48 e Apriori aplicados na detecção de indicadores da qualidade de vida e saúde. Revista Interdisciplinar de Ensino, Pesquisa e Extensão, v. 1, n. 1, 2014.

MACIEL, T. V. et al. Mineração de dados em triagem de risco de saúde. Revista Brasileira de Computação Aplicada, v. 7, n. 2, p. 26–40, 2015.

NOTÍCIAS, S. Proposição. Disponível em: <<https://www12.senado.leg.br/noticias/glossario-legislativo/proposicao>>. Acesso em: 2 jul. 2017.

VIANNA, R. C. X. F. et al. Mineração de dados e características da mortalidade infantil. Cad Saude Publica, p. 535–542, 2010.