# Mineração de dados eleitorais: descoberta de padrões de candidatos a vereador na região da campanha do Rio Grande do Sul

Alex Camargo<sup>1</sup>
Roger Silva<sup>1</sup>
Érico Amaral<sup>2</sup>
Milton Heinen<sup>2</sup>
Francisco Pereira<sup>2</sup>

Resumo: Atualmente o Brasil é visto como um país onde o sistema de votação é bastante avançado em termos de informatização. A cada final de eleição uma grande quantidade de dados é gerada e, com base em números estatísticos, são efetuados estudos a fim de se identificar quais fatores levaram um candidato a ser eleito. Este trabalho propõe um processo de descoberta de padrões de perfil de candidatos a vereador nas cidades da região da campanha do Rio Grande do Sul. Com o auxílio da ferramenta WEKA (Waikato Environment for Knowledge Analysis) foi aplicado o algoritmo J48 utilizando como base de dados os registros disponibilizados pelo TSE (Tribunal Superior Eleitoral), referente as eleições do ano de 2012.

Palavras-chave: Classificação. Descoberta de padrões. Mineração de dados.

**Abstract:** Actually Brazil has been seen as a country in which the vote's system is very advanced in terms of data processing. At the final of every vote's period a great data amount is generated, and, based on statistic numbers, studies are done with the purpose to identify which factors had made a candidate win it. The present work proposes a discovery process of standards of candidates' profiles from southwest's region of the State of Rio Grande do Sul. With the help from WEKA tool (Waikato Environment for Knowledge Analysis) it was applied the J48 algorithm, using data base made available by TSE (Electoral Supreme Tribunal), and it refers to 2012's Elections.

Keywords: Classification. Data mining. Pattern discovery.

## 1 Introdução

Estudos sobre a influência direta da comunicação e da propaganda política em processos eleitorais, realizados nas duas últimas décadas, apontam e consolidam algumas teorias sobre as motivações psicológicas, políticas e econômicas na condução estratégica do eleitorado [1]. Nas campanhas eleitorais, os partidos propõem políticas e lançam seus candidatos [2]. Também podemos mencionar que decifrar o fenômeno eleitoral vai além deste aspecto pois se busca identificar as causas que motivam, ou levam, o indivíduo ou o grupo a tomar esta importante decisão. Identificando o comportamento nas eleições, os fenômenos sociais, mais precisamente os eleitorais, são elucidados [3]. Além das fontes comuns de divulgação eleitoral, como televisão, rádio e carros de som, a Internet está sendo cada vez mais utilizada, onde, através dela, são criados espaços conversacionais nãohierárquicos, mudando a experiência de campanha e divulgando, de certa maneira, opiniões políticas através das mídias sociais [4]. As informações sobre os candidatos, portanto, são buscadas facilmente, o que tornam a vida

{ericohoffamaral, miheinen, fcpbage @gmail.com}

http://dx.doi.org/10.5335/rbca.2015.5395

<sup>&</sup>lt;sup>1</sup> Programa de Pós-Graduação em Computação, FURG, Campus Carreitos - KM 8 - Rio Grande (RS) - Brasil {alexcamargo, rogersilva @furg.br}

<sup>&</sup>lt;sup>2</sup> Especialização em Sistemas Distribuídos com Ênfase em Banco de Dados, UNIPAMPA, Campus Bagé - Travessa 45 - Bagé (RS) - Brasil

política ou profissional do candidato uma grande influência na escolha do mesmo. Além deste fator importante outros também são relevantes, como o total de investimento na campanha política que, muitas vezes, proporciona um maior gasto em propaganda e as características do candidato, como idade, sexo e profissão, que podem influenciar na escolha do mesmo [5].

Considerando o número elevado de dados disponíveis em cada eleição, torna-se possível avaliar o quanto que cada fator supra citado é decisivo para o resultado final. Para isso a mineração de dados é fundamental, pois através desta técnica pode-se extrair informações antes não conhecidas se utilizando de estruturas computacionais onde é definido um escopo para estudo, ou seja, a escolha de quais dados serão utilizados. A proposta é, portanto, explorar os dados dos candidatos a vereador das cidades que compõem a região da campanha do Rio Grande do Sul, composta por sete municípios, sendo eles: Aceguá, Bagé, Caçapava do Sul, Candiota, Dom Pedrito, Hulha Negra e Lavras do Sul para, então, descobrir fatores que levam, ou não, um candidato a eleição.

# 2 Mineração de dados

Como a quantidade de informação disponível aumenta a cada dia, é essencial tentar aproveitar o máximo possível desse recurso. A forma mais sensata de utilizar essas informações é verificar se há algum conhecimento, padrão ou alguma direção dentro delas. Mineração de dados, ou Data Mining em inglês, diz respeito à extração não-trivial de informação implícita, previamente desconhecida e potencialmente útil em grandes conjuntos de dados [6]. A mineração de dados é uma etapa da Descoberta de Conhecimento em Bases de Dados, em inglês KDD (Knowledge Discovery in Databases), onde também podem ser incluídas etapas de extração de informações operacionais e etapas de geração de bases gerenciais de acordo com os objetivos e o contexto envolvido [7]. O processo de KDD, ilustrado na Figura 1, é constituído por cinco etapas, isto pode ser feito por meio de métodos supervisionados como classificação e também por métodos não-supervisionados, por exemplo, agrupamento de dados [8].

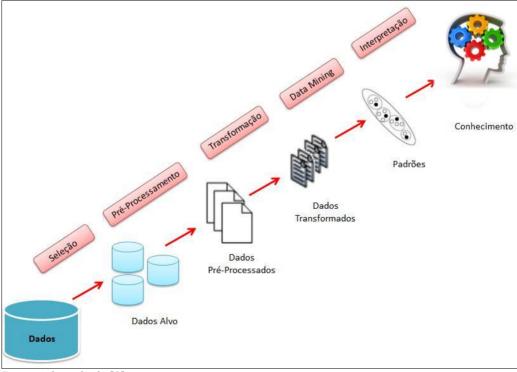


Figura 1: Etapas que constituem o processo de KDD

Fonte: Adaptado de [8]

## 3 Técnicas de mineração de dados

O volume de dados passíveis a mineração estão cada vez mais diversificados quanto a sua estrutura e distribuição. Os dados podem ser divididos em tradicionais (número e pequenas cadeias de caracteres) e complexos (imagens, cadeias de DNA, vídeos, etc). Para esses tipos de dados, a mineração de dados possuí algoritmos bastante consolidados [9]. Tradicionalmente, os dados a explorar são representados como atributos numéricos ou categóricos em uma tabela que descreve em cada tupla um caso de teste do conjunto sob análise.

A escolha do método adequado para realizar a mineração dos dados é um passo importante para alcançar o conhecimento pretendido [5]. Algumas das técnicas de mineração de dados mais significativas são: Classificação, Regras de Associação e Agrupamento.

#### 3.1. Classificação

A técnica de classificação é uma atividade preditiva que consiste na análise de dados em busca da definição de padrões que descrevem tendências futuras para esses dados [10]. Uma abordagem geral para o aprendizado deste modelo consiste, primeiramente, em fornecer dados de treinamento, cujas classes são conhecidas. Os dados de treinamento são então usados para gerar o modelo de classificação, que é posteriormente aplicado aos dados de teste, cujas classes são desconhecidas. O objetivo é criar um modelo capaz de categorizar corretamente tanto os dados utilizados em seu treinamento, como dados nunca vistos antes, ou seja, um modelo com boa capacidade de generalização [6]. O interesse principal na classificação está na precisão do resultado obtido, cuja avaliação é baseada na contagem do número de instâncias classificadas corretamente (acurácia) ou incorretamente (taxa de erro) [11].

### 3.2. Regras de associação

A tarefa dessa técnica envolve a descoberta de regras que indiquem correlações relevantes entre objetos de uma base de dados. Uma regra de associação é uma implicação da forma X => Y, onde X e Y são subconjuntos dos atributos de uma base de dados [12]. Tal técnica pode ser utilizada em diversas aplicações, como, por exemplo, a análise do carrinho de compras de um website. Nesse tipo de aplicação há o interesse em descobrir associações entre os itens comprados pelos consumidores, ou seja, quais produtos são comprados juntamente com outros [13].

## 3.3. Agrupamento

A tarefa de agrupamento identifica a classe de cada objeto de maneira que os objetos dentro de uma mesma classe apresentem alta similaridade entre si e, ao mesmo tempo, baixa similaridade em relação aos objetos das outras classes. A medida de similaridade é feita baseada nos valores dos atributos que descrevem os objetos da base de dados através de métodos de agrupamento [7]. A análise de agrupamentos busca semelhanças entre padrões, organizando os dados em grupos de forma que as características dos objetos pertencentes ao mesmo grupo sejam mais evidentes entre si e distintas dos objetos presentes em outros grupos [14].

# 4 Algoritmo J48

O algoritmo J48 foi desenvolvido por Ross J. Quinlan, representando uma versão do algoritmo C4.5. Basicamente, é construído um modelo de árvore de decisão, baseado em um conjunto de dados de treinamento, e esse modelo é usado para classificar outras instâncias em um conjunto de teste [15]. As árvores de decisão estão entre os mais populares algoritmos de inferência, sendo aplicados em várias áreas pois pode-se extrair regras do tipo "se então" que são facilmente compreendidas [16]. A indução por árvore de decisão é um algoritmo que habitualmente aprende um conjunto de regras com elevada facilidade de compreensão [17]. Com o algoritmo J48 é criada uma árvore de decisão simples, ilustrando de maneira clara o objetivo deste trabalho.

## 5 Ferramentas para mineração de dados

A escolha da ferramenta para se trabalhar com os dados disponíveis é de fundamental importância para o processo de descoberta de conhecimento em bases de dados. Entre os diversos softwares disponíveis, as sessões subsequentes abordam: o WEKA (utilizado no presente trabalho) e o Rapidminer.

### **5.1. WEKA**

WEKA é uma ferramenta livre, sob a licença GPL (General Public License), desenvolvida na Universidade de Waikato. Ela foi desenvolvida em JAVA e contém uma interface gráfica para interagir com arquivos de dados e produzir resultados visuais (árvores, curvas e tabelas) [18]. O programa possui, como interface principal, o "Explorer", onde cada painel representa uma tarefa de mineração de dados diferente, além de um painel denominado "Pré-processo" (preprocess, em inglês), onde as fontes de dados são carregadas e os atributos são selecionados.

## 5.2. Rapidminer

Rapidminer, segundo Costa et. al [19], é um software livre e com código aberto, distribuído de forma independente para análise de dados, permitindo a integração com outros produtos desenvolvidos pelo mesmo projeto. Algumas características destacadas são: funcionamento na maioria das plataformas, interface gráfica intuitiva e integração com diferentes fontes de dados.

### 6 Trabalhos relacionados

Compreende-se que um problema a ser tratado em KDD é a transformação de um conjunto de dados em um modelo utilizável, com determinado propósito, e que possibilite o entendimento humano [20]. A análise de dados tem sido realizada desde seus primórdios da computação por meios estatísticos, mas somente esse método matemático não proporciona a descoberta de informações desconhecidas em uma base de dados [10].

Um trabalho de proposta similar ao aqui apresentado é o de Pereira [5], cujo principal objetivo foi encontrar possíveis padrões de candidatos a prefeito, indicando prováveis tendências relevantes, porém, em um âmbito nacional, onde foram feitas comparações com os resultados das urnas dos principais estados do Brasil bem como uma breve descrição sobre a situação política do país.

A mineração de dados pode ser considerada como um dos passos cruciais da KDD, logo há trabalhos que tratam sobre o tema em diferentes aspectos, como vemos no artigo de Neto et al. [21], que envolveu a aplicação de técnicas de mineração de dados para identificação de padrões de comportamento dos clientes que causam perdas comerciais à uma concessionária de energia elétrica.

# 7 Materiais e métodos

O algoritmo de classificação aplicado neste trabalho é baseado em árvores de decisão. Tais árvores respeitam uma sequencia hierárquica de testes construídos ao longo de uma estrutura de condições. Uma árvore pode expor distintas classificações, onde cada regra é uma direção da árvore, que vai desde a raiz até um nó folha. Como uma maneira de validar as pesquisas foi desenvolvida a implementação do tema proposto através da ferramenta WEKA.

Os experimentos foram aplicados em um computador com o sistema operacional Linux Mint, 8GB RAM, 350GB HD e um processador CORE I3 2,40 GHZ . Para processamento e filtragem dos dados foi utilizada a planilha eletrônica LibreOffice Calc. Após a definição do escopo, a próxima etapa foi a seleção dos dados disponibilizados no portal Repositório de Dados Eleitorais <a href="http://www.tse.jus.br/eleicoes/repositorio-de-dados-eleitorais/">http://www.tse.jus.br/eleicoes/repositorio-de-dados-eleitorais/</a>, conforme mostra a Figura 2.

Originalmente, os dados se encontraram no formato texto (TXT), ilustrado na Figura 3, logo se fez necessária a sua conversão para um arquivo de valores separados por vírgula (CSV).

Figura 2: Repositório de Dados Eleitorais



Fonte: Adaptado de [21]

Figura 3: Arquivo consulta cand 2012 RS.txt

Fonte: Elaboração dos autores com base nos dados da pesquisa

Na etapa subsequente foram definidos quais atributos seriam usados. Os registros de cidades que não fazem parte da região da campanha do Rio Grande do Sul foram excluídos, facilitando, assim, o préprocessamento dos dados. Posteriormente, na etapa de transformação dos dados, foi preciso a adequação de alguns atributos pois não estavam em um padrão ideal para a mineração dos dados. Algoritmos de classificação requerem que os dados estejam na forma de atributos categorizados, por exemplo, dados com alta granulidade (muitos detalhes) intereferem negativamente no resultado. A Figura 4 exibe o resultado final desta etapa.

Figura 4: Arquivo consulta cand 2012 RS.txt

	Α	В	С	D	E	F	G	Н
1	NASC NO MUNICIPIO	POLITICO	IDADE	SEXO	ESCOLARIDADE	CASADO	DESPESA CAMPANHA	ELEITO
2	NAO	NAO	BAIXA(0-30)	FEMININO	MEDIO	NAO	BAIXA(<35mil)	NAO
3	NAO	NAO	BAIXA(0-30)	MASCULINO	SUPERIOR	NAO	BAIXA(<35mil)	NAO
4	SIM	NAO	BAIXA(0-30)	MASCULINO	SUPERIOR	NAO	BAIXA(<35mil)	SIM
5	SIM	NAO	BAIXA(0-30)	MASCULINO	SUPERIOR	NAO	BAIXA(<35mil)	NAO
6	NAO	SIM	MEDIA(31-56)	FEMININO	SUPERIOR	NAO	BAIXA(<35mil)	SIM
7	NAO	NAO	MEDIA(31-56)	FEMININO	MEDIO	SIM	BAIXA(<35mil)	NAO
	SIM	NAO	MEDIA(31-56)	FEMININO	MEDIO	SIM	BAIXA(<35mil)	NAO
a	NAO	NAO	MEDIA/31-56)	CEMININO	CUMPAMENTAL	NAO	BAIVA/<35mil)	NAO

Fonte: Elaboração dos autores com base nos dados da pesquisa

Realizada a etapa de adaptação dos atributos, se fez necessária a conversão do arquivo CSV (Commaseparated values) para o formato ARFF (Attribute-Relation File Format), Figura 5. Para isso, foi escolhido o serviço online chamado csv2arff (CSV to ARFF), disponível em <a href="http://slavnik.fe.uni-lj.si/markot/csv2arff/">http://slavnik.fe.uni-lj.si/markot/csv2arff/</a>. Por fim, o arquivo na extensão "ARFF" foi submetido ao WEKA, conforme ilustra a Figura 6.

Figura 5: Arquivo convertido para ARFF

```
consulta cand ...S campanha.arff *
       @relation vereadores campanha
  2
  3
       @attribute NASC NO MUNICIPIO {SIM,NAO}
  4
       @attribute POLITICO {SIM,NAO}
  5
       @attribute IDADE {BAIXA(0-30),MEDIA(31-56),ALTA(57-85)}
  6
       @attribute SEXO {MASCULINO,FEMININO}
  7
       @attribute ESCOLARIDADE {FUNDAMENTAL, MEDIO, SUPERIOR}
  8
       @attribute CASADO {SIM,NAO}
  9
       @attribute DESPESA CAMPANHA {BAIXA(<35mil),MEDIA(36mil-70mil),ALTA(71mil-150mil)}</pre>
 10
       @attribute ELEITO {SIM,NAO}
```

Fonte: Elaboração dos autores com base nos dados da pesquisa

Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Open fil... Open U... Open D... Genera... Undo Edit... Save... Filter Choose None **Apply Current relation** Selected attribute Relation: vereadores\_campanha Instances: 544 Attribu Name: FLFITO Type: Nominal Attributes: 8 Missina: 0 (0%) Distinct: 2 Unique: 0 (0%) Attributes Label Count 1 SIM 74 All None Invert Pattern 2 NAO 470 No Name NASC\_NO\_MUNICIPIO POLITICO IDADE Visualize All Class: ELEITO (Nom) SEX0 4 ESCOLARIDADE ICASADO DESPESA CAMPANHA 8 ELEITO Remove Status Log x 0 OK

Figura 6: Informações prévias da mineração de dados

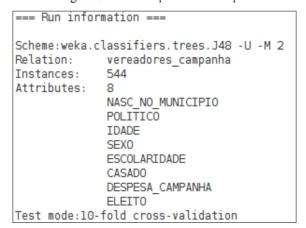
Fonte: Elaboração dos autores com base nos dados da pesquisa

Dentro da ferramenta WEKA, em sua tela principal exibida na Figura 6, é permitida uma visualização geral das informações a serem mineradas. A execução da mineração dos dados foi realizada através da opção "Classify", tendo em vista que a técnica aplicada neste trabalho é a classificação dos dados.

### 8 Resultados obtidos

Com o ambiente de simulação elaborado para o experimento foi possível identificar, na prática, os conceitos previamente estudados neste trabalho. Procurou-se, à medida do possível, documentar os procedimentos realizados de maneira simples e objetiva. A Figura 7 exibe o cabeçalho do resultado da execução do algoritmo J48. Foi utilizada a técnica 10-fold cross validation, que consiste na divisão do conjunto total de dados em 10 subconjuntos mutuamente exclusivos objetivando a validação cruzada da acurácia do modelo.

Figura 7: Informações de execução



Fonte: Elaboração dos autores com base nos dados da pesquisa

O próximo resultado a ser exibido, na Figura 8, mais importante deste trabalho, é a árvore de decisão. O atributo classe, neste caso, é "ELEITO" assumindo "SIM" ou "NÃO".

Figura 8: Árvore de decisão

```
=== Classifier model (full training set) ===
J48 unpruned tree
POLITICO = SIM
   IDADE = BAIXA(0-30): SIM(0.0)
   IDADE = MEDIA(31-56)
       ESCOLARIDADE = FUNDAMENTAL: NAO (1.0)
       ESCOLARIDADE = MEDIO: NAO (3.0/1.0)
       ESCOLARIDADE = SUPERIOR
            NASC NO MUNICIPIO = SIM: SIM (6.0/1.0)
            NASC NO MUNICIPIO = NAO: NAO (7.0/3.0)
    IDADE = ALTA(57-85)
       NASC NO MUNICIPIO = SIM: NAO (3.0/1.0)
       NASC NO MUNICIPIO = NAO: SIM (5.0)
POLITICO = NAO
   SEX0 = MASCULINO
       ESCOLARIDADE = FUNDAMENTAL: NAO (136.0/13.0)
        ESCOLARIDADE = MEDIO: NAO (124.0/18.0)
        ESCOLARIDADE = SUPERIOR
            NASC NO MUNICIPIO = SIM: NAO (49.0/6.0)
            NASC NO MUNICIPIO = NAO
                DESPESA CAMPANHA = BAIXA(<35mil)
                    IDADE = BAIXA(0-30): NAO(1.0)
                    IDADE = MEDIA(31-56)
                        CASADO = SIM: NAO (11.0/5.0)
                        CASADO = NAO: SIM (3.0/1.0)
                    IDADE = ALTA(57-85): NAO(8.0/3.0)
                DESPESA CAMPANHA = MEDIA(36mil-70mil): NAO (10.0/2.0)
                DESPESA CAMPANHA = ALTA(71mil-150mil): NAO (7.0/3.0)
   SEXO = FEMININO: NAO (170.0/7.0)
Number of Leaves
                        17
Size of the tree :
                        28
```

Fonte: Elaboração dos autores com base nos dados da pesquisa

Através da árvore de decisão é possível a análise e descoberta de padrões. Neste trabalho serão abordados somente os principais. Pode-se observar que, o fator mais importante é se o candidato é político (POLITICO = SIM). Também, fatores como idade entre 31 e 56 anos (IDADE = MEDIA(31-56)) juntamente com um grau de instrução mais elevado (ESCOLARIDADE = SUPERIOR), na maioria dos casos obtiveram sucesso na eleição.

Outro fator relevante que a mineração dos dados apresentou foi que se o candidato não for político (POLITICO = NAO) e possuir somente o ensino fundamental (ESCOLARIDADE = FUNDAMENTAL), em grande parte dos casos não é eleito. Esse padrão se mostrou ainda mais evidente referente ao sexo do candidato (SEXO = FEMININO).

A seguinte informação a ser exibida é o sumário de execução do algoritmo J48, nele consta, basicamente, o total de instâncias classificadas corretamente e incorretamente juntamente com o seu nível de precisão, conforme é apresentado na Figura 9, já a Figura 10 exibe a precisão para cada classe.

Figura 9: Sumário de execução

=== Summary ===		
Correctly Classified Instances	480	88.2353 %
Incorrectly Classified Instances	64	11.7647 %
Kappa statistic	0.2398	
Mean absolute error	0.1897	
Root mean squared error	0.3079	
Relative absolute error	80.3543 %	
Root relative squared error	89.8248 %	
Total Number of Instances	544	

Fonte: Elaboração dos autores com base nos dados da pesquisa

Figura 10: Acurácia detalhada por classe

=== Detailed Accuracy By Class ===								
	TP Rate 0.122 0.962	FP Rate 0.038 0.878	Precision 0.333 0.874	Recall 0.122 0.962	F-Measure 0.178 0.916	ROC Area 0.654 0.654	Class SIM NAO	
Weighted Avg.		0.764	0.801	0.847	0.816	0.654	1470	

Fonte: Elaboração dos autores com base nos dados da pesquisa

Então, das 544 instâncias classificadas, 480 foram classificas corretamente e 64 incorretamente, resultando em 88,2% e 11,7%, respectivamente. Com isso alcançou-se uma boa margem de acerto. Após a análise da matriz de confusão, exibida na Figura 11, percebeu-se que o classificador melhor aprendeu a classificar "ELEITO = NAO", isso geralmente ocorre devido ao desbalanceamento das instâncias em uma classe.

Figura 11: Matriz de confusão

==	== (	Confu	IS:	ion	Mat	trix		
	а	Ь					ified	as
		65 452				SIM NAO		

Fonte: Elaboração dos autores com base nos dados da pesquisa

Adicionalmente, foram realizados testes com outros algorimos de classificação disponíveis no WEKA, entretanto, nenhum dos métodos testados atingiu um resultado superior ao do J48 na classificação correta de instâncias . As Figuras 12 e 13 apresentam o sumário de execução dos algoritmos SimpleCart e REPTree, respectivamente.

Figura 12: Sumário de execução adicional (SimpleCart)

=== Summary ===		
Connectly Classified Instances	461	84.7426 %
Correctly Classified Instances	461	
Incorrectly Classified Instances	83	15.2574 %
Kappa statistic	0.1777	
Mean absolute error	0.202	
Root mean squared error	0.3536	
Relative absolute error	85.5269 %	
Root relative squared error	103.1341 %	
Total Number of Instances	544	

Fonte: Elaboração dos autores com base nos dados da pesquisa

Figura 13: Sumário de execução adicional (REPTree)

=== Summary ===		
Correctly Classified Instances	455	83.6397 %
Incorrectly Classified Instances	89	16.3603 %
Kappa statistic	0.1182	
Mean absolute error	0.2061	
Root mean squared error	0.3628	
Relative absolute error	87.267 %	
Root relative squared error	105.8163 %	
Total Number of Instances	544	

Fonte: Elaboração dos autores com base nos dados da pesquisa

# 9 Considerações finais e trabalhos futuros

O objetivo deste trabalho foi apresentar a mineração de dados aplicada em uma situação real, para isso, foram usados dados disponibilizados publicamente pelo portal Repositório de Dados Eleitorais. De acordo com a bibliografia estudada, a escolha de um algoritmo de árvore de decisão mostrou-se apropriada por apresentar facilidade na interpretação dos resultados. Com ajuda da ferramenta WEKA, comumente utilizada para mineração de dados, conseguiu-se o objetivo inicial almejado. Como proposta de trabalhos futuros é sugerida uma solução para o desbalanceamento do atributo classe, tendo em vista que o número de candidatos não eleitos "ELEITO = NAO" é bastante superior comparado ao de candidatos eleitos "ELEITO = SIM". Também estimase uma maior exploração dos parâmetros do algoritmo J48, através do WEKA, e aplicação de novos algoritmos de mineração de dados, comparando-se os resultados. Nesta pesquisa foram abordados os resultados aplicados a uma região específica do Brasil, porém nada impede a busca por padrões de candidatos de outras regiões. Também, o estudo torna possível a aplicabilidade do problema proposto nas demais eleições estaduais e federais, como, para deputado, senador, governador ou até mesmo presidente da república.

## Referências

- [1] FIGUEIREDO, Marcus; ALDÉ, Alessandra; DIAS, Heloisa; JORGE, Vladimir L. Estratégias de persuasão em eleições majoritárias: uma proposta metodológica para o estudo da propaganda eleitoral. Rio de Janeiro: IUPERJ, 1998. (Série Estudos, 100).
- [2] MANIN, Bernard; PRZEWORSKI, Adam; STOKES, Susan C. **Eleições e representação**. Lua Nova, São Paulo, n. 67, p. 105-138, 2006. Disponível em: <a href="http://www.scielo.br/scielo.php?script=sci\_arttext&pid=S0102-64452006000200005&lng=en&nrm=iso">http://www.scielo.br/scielo.php?script=sci\_arttext&pid=S0102-64452006000200005&lng=en&nrm=iso</a>. Acesso em: 05 set 2014.
- [3] OLIVEIRA, Adriano. O estado da arte dos determinantes do voto no Brasil e as lacunas existentes. Sociedade e Cultura, v. 15, n. 1, p. DOI: 10.5216/sec. V15i1. 20684, 2012.
- [4] ROSSINI, Patricia Gonçalves; LEAL, Paulo Roberto Figueira. Efeitos da campanha virtual no universo das mídias sociais: o comportamento do eleitor no Twitter nas Eleições 2010. Compolítica, v. 3, n. 1, p. 7-28, 2013.
- [5] PEREIRA, Francisco C. **Descoberta de conhecimento a partir de dados das eleições municipais das regiões do Brasil**. Trabalho de Conclusão de Curso (Especialização em Sistemas distribuídos com Ênfase em Banco de Dados) Universidade Federal do Pampa, Bagé, 2013.
- [6] PANG-NING, Tan et al. Introduction to data mining. Library of Congress. Harlow: Pearson, 2006.
- [7] HAN, Jiawei; KAMBER, Micheline. **Data mining: concepts and techniques**. The Morgan Kaufmann Series in data management systems. Waltham, MA: Elsevier, 2000.

- [8] WINCK, Ana Trindade. **Processo de KDD para auxílio à reconfiguração de ambientes virtualizados**. Dissertação (Mestrado em Ciência da Computação) Faculdade de Informática, PUCRS, Porto Alegre, 2007.
- [9] APPEL, Ana Paula. **Métodos para o pré-processamento e mineração de grandes volumes de dados multidimensionais e redes complexas**. Tese de Doutorado. Instituto de Ciências Matemáticas e de Computação. 2010.
- [10] PELEGRIN, Diana. et al. A Shell de Data Mining Orion: Classificação, Clusterização e Associação. In: I Congresso Sul Catarinense de Computação, v. 1, n. 1. p. 1-4. Criciúma. Anais... Criciúma: UNESC, 2005.
- [11] ALENCAR, Aretha Barbosa. **Mineração e visualização de coleções de séries temporais**. 2007. 127 f. Dissertação (Mestrado em Ciências Ciências de Computação e Matemática Computacional] Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Paulo, 2007.
- [12] CHEN, Ming-Syan; HAN, Jiawei; YU, Philip S. **Data mining: an overview from a database perspective. Knowledge and data Engineering**, IEEE Transactions on, Washington, DC v. 8, n. 6, p. 866-883, 1996.
- [13] BARIONI, Maria Camila Nardini. **Visualização de operações de junção em sistemas de bases de dados** para mineração de dados. Tese (Doutorado em Ciências de Computação e Matemática Computacional) Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Paulo, 2002.
- [14] GIL, Vanessa de Oliveira; FERRARI, Fabricio; EMMENDORFER, Leonardo. Investigação da aplicação de algoritmos de agrupamento para o problema astrofísico de classificação de galáxias. Revista Brasileira de Computação Aplicada, v. 7, n. 2, p. 52-61, 2015.
- [15] MACHADO, Marcos et al. **Web mining aplicado à detecção de perfis de estudantes**. *In: II Congresso Sul Catarinense de Computação*, v. 2, n. 2. p. 1-4. Criciúma. Anais... Criciúma: UNESC, 2006.
- [16] FUNCHAL, João Paulo; MADSEN, Carlos Alberto Cruz; ADAMATTI, Diana Franscisca. Classificação automática de dados para descoberta de conhecimento: um estudo de caso para classificação de risco na área da saúde. Revista Brasileira de Computação Aplicada, v. 7, n. 2, p. 41-51, 2015.
- [17] MARTINS, António Cardoso; MARQUES, M. J.; COSTA, Paulo Dias. Estudo comparativo de três algoritmos de machine learning na classificação de dados electrocardiográficos. Trabalho (Mestrado em Informática Médica) Universidade do Porto, Porto, 2009.
- [18] ROMA JUNIOR, Douglas Nassif. **Uma ferramenta para mineração de dados de projetos de software livre e criação de redes sócio-técnicas**. Trabalho de Conclusão de Curso (Graduação) Universidade Tecnológica Federal do Paraná, Campo Mourão, 2013.
- [19] COSTA, Evandro; BAKER, Ryan.; AMORIM, Lucas; MAGALHÃES, Jonathas; MARINHO, Tarsis. Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. Jornada de Atualização em Informática na Educação, v. 1, n. 1, p. 1-29, 2013.
- [20] MACIEL, Thales Vaz; SEUS, Vinícius; MACHADO, Karina; BORGES, Eduardo. Mineração de dados em triagem de risco de saúde. Revista Brasileira de Computação Aplicada, v. 7, n. 2, p. 26-40, 2015.
- [21] NETO, Manoel G. de Mendonça; VIEIRA, Ricardo E. P.; ANDION, Izabel C.; MA, Janaina F.; NASCIMENTO, Paulo V.; DANTAS, Pedro; SILVEIRA, Mônica; BRITO, Alfredo J. Garimpando Fraudadores: O Uso da Mineração de Dados na Descoberta de Padrões de Consumidores que Causam Perdas Comerciais em uma Concessionária de Energia Elétrica. In: V Simpósio Brasileiro de Sistemas de Informação, p. 145-156. Anais ... Marabá: SBC, 2010.
- [22] BRASIL. Tribunal Superior Eleitoral (TSE). Repositório de dados eleitorais, 2012. Disponível em: <a href="http://www.tse.jus.br/eleicoes/repositorio-de-dados-eleitorais/">http://www.tse.jus.br/eleicoes/repositorio-de-dados-eleitorais/</a>>. Acesso em: 10 abr 2014.