
Identificando o grau de importância dos programas eleitorais a partir da mineração de dados

Lucinéia Rodrigues Magalhães (Bacharel)

Curso de Bacharelado em Sistemas de Informação – Universidade Tuiuti do Paraná

Deborah Ribeiro Carvalho (Doutor)

Curso de Bacharelado em Ciência da Computação – Universidade Tuiuti do Paraná

Resumo

O presente trabalho tem como objetivo identificar o potencial de uso das técnicas de Data Mining, bem como algumas formas de pós-processar os padrões descobertos para apoio ao processo decisório. Para tanto, foi construído um experimento com vistas a avaliar o quanto a população julga importante o programa eleitoral. Dessa forma, realizou-se uma pesquisa de campo a partir da aplicação de um formulário contendo 15 questões, com o propósito de identificar o entendimento não apenas da importância desta forma de comunicação durante o processo eleitoral, mas também se o mesmo chega a influenciar o eleitor. A fim de melhor explorar os dados, construiu-se um plano tabular para extração de informações, identificaram-se as regras de associação (APRIORI) e as regras de exceção. As regras gerais (de senso comum) comumente representam padrões conhecidos pelo usuário/gestor, ao contrário das regras de exceção. Sendo assim, as regras de exceção tendem a ser mais interessantes/surpreendentes, dado o fato de representarem uma contradição em relação à regra geral.

Palavras-chave: Mineração de Dados; Descoberta de Conhecimento; Regras de exceção; Horário Político.

Abstract

This work has as a goal to identify the potential use of Data Mining techniques, as well as some forms of post processing the patterns found to support the decisory process. For this, a experiment was built to evaluate how important is the electoral program for population. A field research was made from the application of a form enclosing 15 questions with the purpose of identifying the understanding not only the importance of this way of communication during electoral process, but if it influences the voter. In order to explore data better, a tabular plan was built to obtain information. Association rules (APRIORI) and exception rules were identified. Differently of exception rules, general rules (of common sense) usually represent standards know by the user/supervisor. In this way, exception rules tend to be more interesting/surprising, due the fact of representing a contradiction to general rule.

Key Word: Data Mining, Knowledge Discovery, Rules Exception; Political Time

Introdução

Em geral, a campanha dos candidatos a algum cargo público ganha maior visibilidade a partir dos programas eleitorais na TV e rádio, principalmente em 2006, quando diversas outras formas de divulgação foram proibidas pela lei eleitoral. A partir destes programas, é possível o acesso às diversas propostas. Alguns candidatos chegam a afirmar: “A nossa campanha efetivamente deslanchou a partir do programa eleitoral na TV e no rádio. A falta de visibilidade nas placas fez com que as pessoas não tivessem conhecimento de nossa candidatura; acarretando um grande número de indecisos”. Porém até que ponto a população valoriza e assiste a estes programas? Estes programas chegam a influenciar na hora do voto?

Na época das eleições surgem questionamentos diversos, tais como: O voto deve ser obrigatório? Até que ponto o horário eleitoral é eficiente e atrativo para os eleitores? As opiniões são diversas sobre estas discussões, por exemplo, Serrano (2006) defende o fim do voto obrigatório e do horário eleitoral gratuito. A justificativa para tal posição é que “O horário eleitoral gratuito, por sua vez, também se transformou em ferramenta de manipulação e mistificação. Nasceu como um grito de liberdade em um momento em que a ditadura sufocava o debate político e teve um

papel importantíssimo no processo de abertura e na reconstrução da democracia”.

Um dos pontos de defesa de tais programas reside no fato de que uma grande parcela dos brasileiros não tem acesso a outros meios de informação, como, por exemplo, jornais, internet, constituindo assim, a TV, a única fonte de informação sobre o candidato, suas propostas, vida política, etc.

Dada a diversidade de pontos de vista, este artigo tem como objetivo realizar uma pesquisa a partir da aplicação de um formulário e potencializar o uso dos dados a partir de técnicas de Mineração de Dados, identificando padrões de opinião entre os entrevistados. Vale destacar que o principal objetivo deste artigo é comparar o quanto essas técnicas de Mineração de Dados podem complementar as tradicionais extrações de informações a partir de bases de dados oriundas de pesquisas de opinião.

2 Mineração de Dados

É muito frequente a tradicional forma de exploração das bases de dados não ser suficientemente eficiente para auxiliar o gestor no processo decisório. Uma das razões é o fato dos métodos tradicionais dependerem da definição de hipóteses por parte do usuário/gestor, em geral consolidadas em um plano tabular, que nada mais

é do que a definição das informações a serem obtidas. Outras vezes essa atividade de avaliação ultrapassa a capacidade humana de análise e interpretação. Para facilitar a recuperação e o uso desses dados, uma das alternativas possíveis é a aplicação do processo de Descoberta de Conhecimento a partir de Bases de Dados (*Knowledge Discovery in Data Base - KDD*). Segundo (Fayyad *et al*, 1996), esse processo é composto das seguintes etapas: Pré-processamento - prevê a coleta e seleção dos dados, da incorporação de novos dados criados a partir dos já existentes; Mineração de Dados – aplicação de algoritmos que procurem por padrões/relações e regularidades em um determinado conjunto de dados; Pós-processamento - Interpretação e avaliação - verifica a qualidade do conhecimento (padrões) descoberto, procurando identificar se o mesmo auxilia a resolver o problema original que motivou a realização do processo KDD.

Os algoritmos de Mineração de Dados podem ser identificados em três tarefas principais: Classificação, Descoberta de Regras de Associação e Agrupamento.

Neste artigo será aplicada tarefa de Descoberta de Regras de Associação e posterior identificação de regras de exceção sobre conjunto regras originalmente descoberto. A opção pela tarefa de Regras de Associação se deve ao fato da natureza da investigação

estar relacionada a encontrar padrões em itens de dados de mesma natureza, como, por exemplo, um formulário de pesquisa de opinião.

A Descoberta de Regras de Associação prevê identificar afinidades entre os itens de dados mediante um determinado número de transações (Agrawal *et al.*, 1993). Tendo-se um conjunto de registros, em que cada qual constitui um conjunto de itens de dados, uma regra de associação é uma expressão do tipo $X \rightarrow Y$ (Se X, então Y), em que X e Y representam conjuntos de itens independentes, na forma $X \cap Y = \emptyset$.

Um exemplo, segundo Carvalho (2004), de uma regra do tipo $X \rightarrow Y$ poder ser: 90% dos consumidores que compram pneus e acessórios automotivos também utilizam serviços de manutenção do carro. O valor 90% é dito a confiança da regra, ou seja, representa a probabilidade condicional do conseqüente ocorrer dado que o antecedente também ocorreu.

Outra medida atribuída às regras de associação é o suporte, o qual representa a probabilidade de ocorrência do antecedente X.

O Suporte e a Confiança, segundo (Agrawal *et al.*, 1993), são definidos como:

$$\text{Suporte} = |X| / N \quad (1)$$

$$\text{Confiança} = |X \cup Y| / |X| \quad (2)$$

onde N é o número total de exemplos.

Neste artigo, a formulação do suporte difere

daquela proposta originalmente (expressão 3) por Agrawal e seus colegas (1993), por se acreditar que a definida pela expressão 1 é mais intuitiva e de fácil interpretação.

$$\text{Suporte} = |X \cap Y| / N \quad (3)$$

3 Pós-Processamento do Conjunto de Regras Descobertas

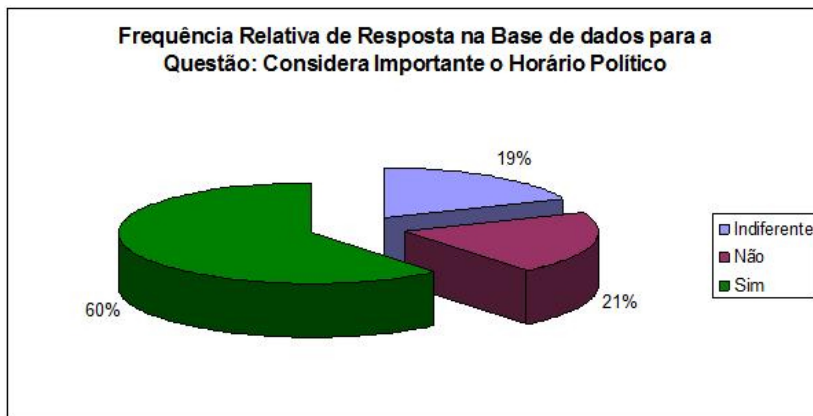
Embora os conjuntos de regras de associação contenham padrões interessantes, por permitirem a percepção da relação entre itens de dados, em geral apresentam inconvenientes, tais como: a quantidade de regras pode ser muito grande, nem todas serem efetivamente relevantes, etc. Uma das alternativas para minimizar esse problema é a adoção de técnicas de pós-processamento. Desse modo, o número de regras descobertas pode ser reduzido, ranqueando pelo grau de interesse atribuído, etc. facilitando assim a sua avaliação.

Neste artigo são relatadas não apenas as regras descobertas, mas também situações de exceção, que constitui uma das alternativas possíveis de pós-processamento. É importante observar que uma regra de exceção é uma especialização de uma regra de senso comum (geral) e nega o conseqüente previsto por esta regra. Esse método assume que regras de senso comum

representam padrões conhecidos pelo usuário, tendo em vista que elas têm uma grande cobertura (suporte), ao contrário das regras de exceção, que em geral são desconhecidas, uma vez que são de baixa cobertura. Em geral, as regras de exceção tendem a ser mais surpreendentes. Por exemplo, a regra “se X, então Y” corresponde à regra geral, tendo como exceção a regra “se (X e B), então (não) Y”. Essa notação de seta invertida (\leftarrow) é o padrão de saída do algoritmo adotado <APRIORI> (Borgelt, 2004).

$Y \leftarrow X$ regra de senso comum (alta cobertura e alta confiança).
 $\neg Y \leftarrow X, B$ regra de exceção (baixa cobertura, alta confiança).

Gráfico 1- Frequência relativa das respostas para a questão: considera importante o horário político?



$\neg Y \leftarrow B$ regra de referência (baixa cobertura e/ou baixa confiança).

4 Resultados

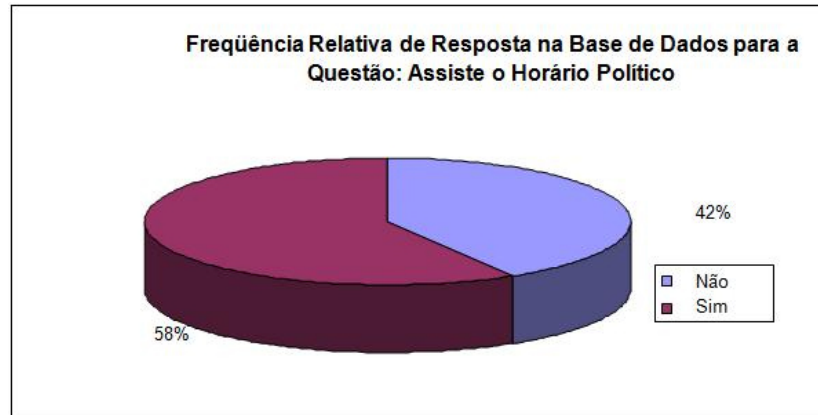
Para a pesquisa foi elaborado um formulário a ser respondido pelos 110 entrevistados que participaram da pesquisa (anexo). A partir da consolidação dos dados é possível perceber que 53% são do sexo feminino e 47% masculino; mais de 68% são solteiros, aproximadamente 30% são casados, e outros 2% viúvos ou divorciados; acima de 65% trabalham.

Em relação ao grau de instrução, 5% concluíram o ensino fundamental; 15% ensino médio; 65% ainda cursam o ensino superior; 3% possuem apenas graduação e 12% concluíram cursos de pós-graduação.

Quanto à importância do horário político, a grande maioria, 60% dos entrevistados, julga importante, 19% demonstram indiferença e outros 21% não o consideram (gráfico 1).

Sobre o fato de assistirem o horário político, 58% dos entrevistados responderam que assistem, atingindo aproximadamente o percentual das pessoas que entendem que este programa seja importante (gráfico 1), e 42% não o assiste.

Gráfico 2. Frequência relativa das respostas para a questão: assiste o horário político?



Considerando que a formação acadêmica tenha alguma influência na opinião dos entrevistados, em relação ao grau de importância deste programa, extraiu-se a tabela 1.

Analisando a tabela 1 não é possível perceber uma relação evidente entre estes dois atributos contemplados, enquanto no grupo dos Especialistas 66.7% consideram o programa eleitoral importante, já entre os mestres a maior frequência é dos que não o consideram importante.

Em geral, os métodos mais tradicionais para processar pesquisas são baseados em hipóteses previamente idealizadas de cruzamentos entre as variáveis, que venham a contribuir na análise do objeto em questão.

Neste artigo, a proposta é exatamente verificar como a adoção de Mineração de Dados, mais especificamente,

Descoberta de Regras de Associação, pode aumentar este potencial de análise. A motivação decorre do fato deste tipo de alternativa ser menos dependente da definição de requisitos dos especialistas do problema de gestão em questão.

Para tal, foi descoberto um conjunto de regras, a partir da base de dados contendo as respostas à pesquisa. Na parametrização do algoritmo foram utilizados valores

para o suporte e confiança superiores ou iguais a 10% e 80% respectivamente, resultando um conjunto de 4.048 regras de associação.

Tabela 1. Frequência relativa das respostas para a questão: considera importante o Horário Político, segundo o grau de formação.

Formação	Considera Importante o Hor. Político (%)			Total
	Indiferente	Não	Sim	
ENS.FUNDAMENTAL	40,0	20,0	40,0	5
ENS.MEDIO	12,5	25,0	62,5	16
ENS.SUP.INCOMPLETO	20,0	15,7	64,3	70
ENS.SUP	0,0	33,3	66,7	3
ESPECIALISTA	0,0	33,3	66,7	3
MESTRADO	12,5	50,0	37,5	8
DOCTORADO	50,0	0,0	50,0	2
Total	18,7	20,6	60,7	107

Apesar dos gráficos 1 e 2 indicarem que aproximadamente os mesmos percentuais de entrevistados consideram importante o horário político e o assistirem, não significa que sejam exatamente as mesmas pessoas. A regra 1 nos permite identificar que dos 57,9% que assistem o horário político obrigatório na televisão, 82,3% acreditam na sua importância e aproximadamente 18% não acreditam. Ou seja, a regra de associação complementa a informação obtida a partir dos gráficos gerados.

Quadro 1 - Regra 1

Regra 1 – Importância_HP_Sim ← Assiste_SIM	(57.9, 82.3)
--	--------------

Da mesma forma a regra 2 nos permite identificar que das pessoas entrevistadas, 60,2% acham que é importante o horário político obrigatório na televisão; destas, 86,2% dizem que assistir o horário político influencia na sua escolha de voto. Neste caso, apesar de complementar a informação já disponível a partir da interpretação do gráfico 1, pode-se concluir que esta associação representa algo esperado, ou seja, quem acredita que o programa seja importante também crê que este influencie o voto.

Quadro 2 - Regra 2

Regra 2 – Influência_voto_Sim ← Importância_HP_Sim	(60.2, 86.2)
--	--------------

Quadro 3 - Regra 3

Regra 3 - Influência_voto_Sim ← Tempo_disponível_Nao Importância_HP_Sim	(34.3, 89.2)
---	--------------

A partir da regra 3, é possível perceber que a regras de associação descobertas podem contemplar antecedentes compostos por mais itens de dado, como por exemplo Tempo_disponível_Não e Importância_HP_Sim. Neste tipo de representação de padrões, é muito frequente encontrarmos regras com mais itens, quer sejam no antecedente e/ou no conseqüente. A análise desta regra indica que 34,3% das pessoas não estão disponíveis para assistir o horário político, no tempo reservado na televisão, mas considera importante este programa, das quais 89,2% acreditam que o programa influencia na decisão do voto. Neste caso, pode-se inferir que mesmo as pessoas que não têm tempo para assistir, mas entendem que o programa seja importante, julgam que este influencie o voto.

Dos entrevistados, 12% consideram péssima a qualidade do programa político, porém também o consideram importante; destes, 84,6% acreditam que ele influencia na decisão do voto (regra 4). Da mesma forma que na regra 3, pode-se também inferir que mesmo as pessoas julgando como péssima a qualidade destes programas, mas entendendo que estes sejam importantes, acreditam que o mesmo influencia o voto.

Analisando estas três últimas regras em conjunto, é possível perceber que as regras 3 e 4 são redundantes em relação à primeira, ou seja, a associação que de fato

está evidenciada é estabelecida entre a importância do horário político e a sua influência sobre o voto. Conforme foi mencionado na seção 1, a etapa de Pós-processamento, que sucede a Mineração de Dados, auxilia na avaliação e análise dos padrões descobertos. Para tal, é possível encontrar na literatura diversas alternativas para concretizar esta etapa, como por exemplo, a eliminação de regras redundantes, a generalização a partir do uso de taxonomias, a busca por regras de exceção, etc.

A partir do conjunto de regras descobertas, foi possível identificar a existência de regras gerais e suas respectivas regras de exceção. Porém, foi necessário executar novamente o algoritmo APRIORI alterando a parametrização, baixando o suporte, conforme já foi exposto na seção 3, a regra de exceção, em geral, tem suporte mais baixo.

A regra 5 demonstra que 29,9% das pessoas trabalham, acreditam que o programa eleitoral influencia o voto, e também entendem que o horário político seja importante, das quais 62,5% qualificam o programa como razoável. A exceção a esta regra ocorre para os profissionais da área de Assessoria Administrativa que o qualificam como bom.

Avaliando a regra 6, percebe-se que 13,1% dos entrevistados acreditam que a melhor alternativa seja a TV, assistem, bem como tem tempo disponível, dos

Quadro 4 - Regra 4

Regra 4 - Influência_voto_SIM ← Qualidade_Programa_Pessimo Importância_HP_SIM (12.0, 84.6)
--

Quadro 5 - Regra 5

Regra 5 Qualidade_Programa_RAZOAVEL ← trabalha_SIM Influência_voto_SIM Importância_HP_SIM (29.9%, 62.5%)
Exceção à Regra 5 Qualidade_Programa_BOM ← trabalha_SIM Influência_voto_SIM Importância_HP_SIM profissao_ASSESS.ADM (3.7%, 75.0%)

Quadro 6 - Regra 6

Regra 6 Influência_voto_SIM ← opcao_TV Assiste_SIM Tempo_disponivel_SIM (13.1%, 71.4%)
Exceção à Regra 6: Influência_voto_NAO ← sexo_M Tempo_disponivel_SIM Assiste_SIM opcao_TV (2.8%, 66.7%)

quais 71,4% julgam que o programa influencia o voto. A exceção a esta regra ocorre em relação aos homens que em iguais situações já não acreditam nesta influência.

Quadro 7 - Regra 7

Regra 7 Influência_voto_SIM ← Qualidade_RAZOAVEL Importância_HP_SIM (35.5%, 81.6%)
Exceção à Regra 7: Influência_voto_NAO ← Qualidade_RAZOAVEL Tempo_aceitavel_SIM Importância_HP_SIM formacao_ENS.MEDIO (2.8%, 66.7%)

A regra 7 mostra que 35,5% acreditam na importância do programa e atribui o grau razoável a sua qualidade, dos quais 81,6% julgam que o programa influencia o voto. A exceção a esta regra é identificada quando os entrevistadores têm como grau de instrução o Ensino Médio.

Quadro 8 - Regra 8

Regra 8 Importância_HP_SIM ← opção_TV trabalha_SIM (24.3%, 80.8%) Exceção à Regra 8: Importância_HP_INDIFERENTE ← opção_TV trabalha_SIM Influência_voto_NAO formacao_ENS.SUP_INCOMPLETO (4.7%, 80.0%)
--

A regra 8 mostra que 24,3% optam pela TV como alternativa mais indicada ao candidato para expor as suas idéias e trabalham, dos quais 80,0% entendem que o Horário Político seja importante. A exceção a esta regra é identificada quando os entrevistados têm como grau de instrução o Ensino Superior Incompleto e não acreditam que o programa político influencie o voto.

Complementarmente à tabela 1, pode-se inferir a partir do conjunto de regras 9, que 65,4% dos

Quadro 9 – Regra 9

Regra 9 Importância_HP_SIM ← formacao_ENS.SUP_INCOMP (65.4%, 64.3%) Exceções à Regra 9: Importância_HP_NÃO ← formacao_ENS.SUP_INCOMP opção_DEBATE (1.9%, 100.0%) Importância_HP_NÃO ← formacao_ENS.SUP_INCOMP opção_TRABALHOS (0.9%, 100.0%) Importância_HP_NÃO ← formacao_ENS.SUP_INCOMP profissao_MILITAR (0.9%, 100.0%) Importância_HP_NÃO ← formacao_ENS.SUP_INCOMP profissao_BANCARIO (0.9%, 100.0%) Importância_HP_INDIFERENTE ← formacao_ENS.SUP_INCOMP profissao_TI (0.9%, 100.0%) Importância_HP_INDIFERENTE ← formacao_ENS.SUP_INCOMP opção_RADIO (0.9%, 100.0%)

Quadro 10 - Regra 10

Regra 10 Importância_HP_SIM <- formacao_ESPECIALISTA (2.8%, 66.7%) Exceções à Regra 10: Importância_HP_NÃO <- formacao_ESPECIALISTA estado_civil_SOLTEIRO (0.9%, 100.0%) Importância_HP_NÃO <- formacao_ESPECIALISTA frequencia_assiste_NUNCA (0.9%, 100.0%) Importância_HP_NÃO <- formacao_ESPECIALISTA opção_COMICIO (0.9%, 100.0%) Importância_HP_NÃO <- formacao_ESPECIALISTA profissao_ANAL_SUPORTE (0.9%, 100.0%)

entrevistados têm ensino superior incompleto, dos quais 64,3% entendem que o horário político na televisão seja importante. Porém, quando se considera a profissão militar, bancário ou técnico em tecnologia da informação, esta situação se associa a não entender o programa eleitoral na televisão como importante ou mesmo ser indiferente. A mesma situação se evidencia ao combinarmos a formação com as alternativas mais indicadas para o candidato expor suas idéias como, por exemplo, debates ou mesmo trabalhos na comunidade.

Ainda sobre a tabela 1, ao longo do artigo já foi comentado que 66,7% dos Especialistas entendem o programa como importante. Porém, o conjunto de regras 10 permite entender melhor os fatores que levam a contrariar esta associação. Por exemplo, o estado civil, frequência com a qual o assistem, a profissão e terem como melhor opção os comícios.

A partir do conjunto de regras descobertas avaliado, é possível verificar que dentre as variáveis constantes no formulário aplicado na pesquisa, destacam-se: profissão, nível de graduação e sexo “provocando” exceções em relação ao comportamento geral. Apesar das alternativas mais indicadas para o candidato expor suas idéias também gerar exceções ao padrão, trata-se de uma situação esperada, ou seja, estas pessoas tendem a não entender o programa como importante.

Conclusão

O objetivo deste artigo é exatamente demonstrar a importância da complementação das técnicas tradicionais na extração de informações, a partir das técnicas de Mineração de Dados, para melhor aproveitar o potencial dos dados disponíveis.

Para tal, foi aplicada uma pesquisa sobre o entendimento das pessoas entrevistadas quanto à importância dos programas eleitorais. A partir desta, foi possível perceber que a maioria dos entrevistados entende que o horário político seja importante (gráfico 1), bem como o assistem (gráfico 2), porém a regra de associação 1 evidencia que não existe uma interseção plena destes dois conjuntos, ou seja, 18% dos que assistem não entendem que o horário político seja importante.

Muitos dos resultados obtidos indicam que as pessoas, na maioria das vezes, consideram razoável a qualidade do horário político, não acham que seja justificável a quantidade de tempo a ele reservado, e a maioria não entende não ser influenciada pelo seu conteúdo.

Algumas pessoas consideram importante o horário político, dizem que este influencia na opção do voto, mesmo considerando-o péssimo (Regra 3).

A partir deste estudo de caso, foi possível perceber que a Descoberta de Regras de Associação, considerando os dados levantados, de fato agregou valor ao potencial de análise sobre a opinião das pessoas entrevistadas em relação aos programas políticos. Algumas das variáveis levantadas demonstraram especial destaque, como por exemplo, profissão (Regras 5, 9 e 10), sexo (Regra 6) e formação (Regra 7 e Regra 8).

Referências

AGRAWAL, R.; IMIELINSK, T.; SWAMI, (1993) A. Mining Associations Between Sets of Items in Massive Databases. Proc. of the ACM-SIGMOD *Int'l Conference on Management of Data*, Washington D.C., May 1993, p.207-216.

BORGELT, C. *Working Group Neural Networks and Fuzzy Systems*, Departament of knowledge Processing and Language Engineering. Otto-von-Guericke-University of Magdeburg, Alemanha. Disponível em: <http://www-ics.cs.uni-magdeburg.de/iws.html> Acesso em: 05 jun. 2004.

CARVALHO, D. R. Gestão Pedagógica de Cursos de Graduação a partir de Data Mining In: *IV Congresso Brasileiro de Computação*, IV Congresso Brasileiro de Computação – SCBCOMP-2004. ITAJAI: SCBCOMP, 2004.

FAYYAD, U. M; PIATETSKY-SHAPIRO, G.; SMYTH, P.; (1996). Uthurusamy, R. *Advances in Knowledge Discovery and Data Mining*. USA: American Association for Artificial Intelligence.

REZENDE, S. O. *Sistemas Inteligentes*. Editora Manole, 2003.

Anexo

Questionário: Horário político obrigatório na televisão

1) Idade: _____

2) Sexo: Masculino Feminino

3) Estado civil:

Casado Solteiro Divorciado Outros. Qual: _____

4) Formação:

Ens. Fundamental Ens. Médio Ens. Superior
 Especialista Mestrado Doutorado

5) Trabalha: Não Sim.

6) Profissão: _____

7) Você considera importante o horário político obrigatório na televisão?

Sim Não Indiferente

8) Você assiste o horário político?

Sim Não

9) Em relação à questão anterior, com qual frequência, na semana, você assiste o horário político na televisão?

Nunca Raramente Algumas vezes Frequentemente
 Sempre

- 10) No horário reservado a propaganda eleitoral, você está com seu tempo disponível para assistir?
 Sim Não
- 11) Você considera importante o horário político ser apresentado duas vezes diariamente na televisão?
 Sim Não Indiferente
- 12) Você considera aceitável o tempo reservado para o horário político? (100 minutos diários).
 Sim Não Indiferente
- 13) Assistir o horário político influencia na sua escolha do voto?
 Sim Não
- 14) Sob seu ponto de vista, a qualidade do horário político na televisão é:
 Péssimo Razoável Bom Ótimo
- 15) Qual seria a alternativa que você mais indicaria ao candidato à eleição expor as suas idéias:
 TV Rádio Outdoors Folder - Santinhos Comícios
 Outros. Qual: _____