

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE
CAMPINAS

CENTRO DE CIÊNCIAS EXATAS, AMBIENTAIS E DE
TECNOLOGIAS

ALAN MENK DOS SANTOS

SMART MARKETING NA TV DIGITAL INTERATIVA
ATRAVÉS DE UM SISTEMA DE RECOMENDAÇÃO
DE ANÚNCIOS

CAMPINAS
2012

ALAN MENK SANTOS

SMART MARKETING NA TV DIGITAL INTERATIVA
ATRAVÉS DE UM SISTEMA DE RECOMENDAÇÃO
DE ANÚNCIOS

Dissertação apresentada como exigência para obtenção do Título de Mestre em Engenharia Elétrica, ao programa de Pós-Graduação em Gestão de Redes de Telecomunicações, Pontifícia Universidade Católica de Campinas.

Orientador: Prof. Dr. David Bianchini

Ficha Catalográfica
Elaborada pelo Sistema de Bibliotecas e
Informação - SBI - PUC-Campinas

t005.74
S237s

Santos, Alan Menk dos.
SMART marketing na TV digital interativa através de um sistema
de recomendação de anúncios / Alan Menk dos Santos. – Campinas:
PUC-Campinas, 2012.
144p.

Orientador: David Bianchini.
Dissertação (mestrado) - Pontifícia Universidade Católica de
Campinas, Centro de Ciências Exatas, Ambientais e de Tecnologias,
Pós-Graduação em Engenharia Elétrica.
Inclui bibliografia.

1. SMART (Sistema de recuperação da informação). 2. TV Digital.
3. Sistemas de recuperação da informação. 4. Teledifusão. I. Bianchi-
ni, David. II. Pontifícia Universidade Católica de Campinas. Centro de
Ciências Exatas, Ambientais e de Tecnologias. Pós-Graduação em
Engenharia Elétrica. III. Título.

22.ed.CDD – t005.74

ALAN MENK DOS SANTOS

**SMART MARKETING NA TV DIGITAL INTERATIVA
ATRAVÉS DE UM SISTEMA DE RECOMENDAÇÃO DE
ANÚNCIOS**

Dissertação apresentada ao Curso de Mestrado Profissional em Gestão de Redes de Telecomunicações do Centro de Ciências Exatas, Ambientais e de Tecnologias da Pontifícia Universidade Católica de Campinas como requisito parcial para obtenção do título de Mestre em Gestão de Redes de Telecomunicações.

Área de Concentração: Gestão de Redes e Serviços.

Orientador: Prof. Dr. David Bianchini

Dissertação defendida e aprovada em 03 de dezembro de 2012 pela Comissão Examinadora constituída dos seguintes professores:



Prof. Dr. David Bianchini
Orientador da Dissertação e Presidente da Comissão Examinadora
Pontifícia Universidade Católica de Campinas



Prof.ª Dr.ª Lia Toledo Moreira Mota
Pontifícia Universidade Católica de Campinas



Prof. Dr. João Pedro Albino
Universidade Estadual Paulista Júlio de Mesquita

Dedico esta dissertação a minha amada esposa que me apoiou a todo o momento e ajudou na correção. E também a minha querida mãe Mari Estela e ao meu Irmão William Menq que sempre acreditaram em mim. Em especial ao responsável por tornar meus sonhos realidade, o meu pai Luiz Alves, que lá de cima a todo o momento me acompanhou e iluminou meus passos.

AGRADECIMENTOS

Ao Prof. Dr. David Bianchini,
Meu amigo e grande orientador que sempre acreditou que era possível tornar este sonho uma realidade.

Aos Professores Alexandre de Assis Mota e Lia Toledo Moreira Mota, por apoiarem e incentivarem a realização deste projeto.

Ao Prof. Dr. Marcelo Luís Francisco Abbade
Prof. e Coordenador do Programa de Mestrado da PUC-Campinas

Ao Prof. Ms. Manoel Campos da Silva Filho, pela grande ajuda na programação do aplicativo *Smart Marketing*.

Aos Professores Dres. Juan Manuel Adán Coello, Fabio Santos da Silva e Maria Augusta Silveira Netto Nunes pelas sugestões no desenvolvimento deste trabalho.

A Prof. Dra. Maria Madalena Dias, pelas sugestões no desenvolvimento do sistema de recomendação e ajuda no entendimento dos resultados.

A Prof^a. Simone Regina Silva a instituição de ensino FCV por colaborarem na realização deste trabalho.

Aos alunos e funcionários e professores da FCV por participarem do experimento.

Aos Amigos Shyrles Monteiro, Felipe José Rocha Vieira.
Eduardo Correa e Ricardo Augusto Giaviti pela troca de experiências em TV Digital e Mineração de Dados.

"Não sabendo que era impossível, foi lá e fez".
Jean Cocteau

RESUMO

SANTOS, Alan Menk. *Smart Marketing* na TV Digital Interativa através de um sistema de recomendação de anúncios. 2012. 143f. Dissertação (Mestrado em Engenharia Elétrica). Pontifícia Universidade Católica de Campinas, Centro de Ciências Exatas, Ambientais e de Tecnologias, Programa de Pós – Graduação em Engenharia Elétrica, Campinas, 2012.

Com a implantação do Sistema Brasileiro de TV Digital (SBTVD), inicia-se uma gama de novas oportunidades e possibilidades tanto para o telespectador quanto as emissoras de TV. Para os Telespectadores, eles terão uma imensa quantidade de canais, programas e propagandas interativas. Para as emissoras de TV, aumenta a possibilidade de propagandas em novos meios de comunicação. Neste contexto, surge a oportunidade de um sistema de recomendação para os aplicativos e portais de interatividade.

Esta dissertação apresenta uma proposta de personalização de propaganda em aplicativos e portais do ambiente de TV Digital com o objetivo de trazer uma melhor experiência ao telespectador, uma nova forma de obtenção de recursos por parte das teledifusoras e também uma maior aceitação de produtos especializados, para uso.

Este trabalho desenvolve um aplicativo para a TV Digital interativa denominado *Smart Marketing* capaz de capturar os dados de navegação do telespectador tanto por meio implícito quanto explícito, realizando a apresentação de publicidades personalizadas a partir do processo de descoberta do conhecimento.

Elaborado a partir do middleware AstroTV, compatível com a especificação brasileira, sua aplicação foi avaliada por meio do experimento que se utilizou, de usuários com perfis variados, aplicando na base de dados gerada o processo de descoberta de conhecimento, o qual utilizou-se das tarefas de classificação e agrupamento. Os resultados obtidos indicaram a qualidade da recomendação gerada pelo *Smart Marketing*.

Termos de Indexação: TV Digital, SBTVD, Interatividade, Sistemas de Recomendação, Personalização, Smart Marketing, Mineração de Dados, KDD.

ABSTRACT

SANTOS, Alan Menk. Smart Marketing on Interactive Digital TV through an advertising recommendation system. 2012. 135f. Dissertation (Master's Degree in Electrical Engineering). Pontifical Catholic University of Campinas, Center of Exact and Environmental Sciences and Technologies, Post-Graduation Program in Electrical Engineering, Campinas, 2012.

With the implementation of the Brazilian Digital TV System (SBTVD) comes a range of new opportunities and possibilities both for viewer and TV stations. For the viewers, they will have an immense amount of channels, programs and interactive advertisements. For TV stations, it increases the possibility of advertising in new media. In this context, the opportunity arises for a recommendation system for applications and interactivity portals. This dissertation presents a proposal of advertising personalization into applications and portals of digital TV environment in order to bring a better experience to the viewer, a new form of income for the broadcasters and also a greater acceptance of specialized products for use.

This work develops an application for interactive Digital TV called Smart Marketing capable of capturing viewer navigation data through both implicit and explicit means by performing customized advertising from the process of knowledge discovery.

Developed from AstroTV middleware, compatible with the Brazilian specification, its application was evaluated by means of experiment that used varied user profiles, applying into the generated database the process of knowledge discovery, which used tasks of classification and grouping. The results indicated the quality of the recommendation generated by Smart Marketing.

Indexing terms: Digital TV, Interactivity, Recommendation Systems, Personalization, Smart Marketing.

LISTA DE FIGURAS

Figura 1. Padrões de TV Digital adotados pelos países	27
Figura 2. Cronograma de implantação do SBTVD no Brasil.....	30
Figura 3. Modelo contendo os componentes do SBTVD.	31
Figura 4. Etapas para difusão do sinal	32
Figura 5. Etapas da Recepção	33
Figura 6. Funcionamento do carrossel de dados.....	34
Figura 7. Arquitetura do <i>Middleware</i> Ginga	36
Figura 8. Exemplo de coleta explícita.....	40
Figura 9. Exemplo de coleta sensível ao contexto.....	41
Figura 10. Proporção de dados em 2009 e em 2020.....	42
Figura 11. Modelo do processo de KDD.....	45
Figura 12. Árvore de decisão para a compra de um computador	50
Figura 13. Exemplo de agrupamento em excelentes condições.....	53
Figura 14. K-means com grupos de tamanhos diferentes	54
Figura 15. Exemplo de um ambiente de aplicações TCP/IP	56
Figura 16. Tecnologias e evolução na segurança	57
Figura 17. Diagrama de Caso de Uso.	65
Figura 18. Modelo do Usuario utilizado no Smart Marketing.....	68
Figura 19. Diagrama do Banco de Dados do Smart Marketing.....	72
Figura 20. Arquitetura do Sistema Smart Marketing.	75
Figura 21. Trecho do Perfil do Telespectador enviado por NCLua SOAP.....	78
Figura 22. Tela de cadastro Smart Marketing.....	77
Figura 23. Teclado apresentado no campo Login.....	78
Figura 24. Tela de validação do telespectador.	80
Figura 25. Apresentação da tela de notícias no Smart Marketing.....	81
Figura 26. Dados Capturados na estrutura de Notícias em XML.....	82
Figura 27. Log Gerado após o acesso a uma notícia.	82
Figura 28. Exemplo de Retorno de publicidade no Smart Marketing	84
Figura 29. Tela Segurança e Privacidade.....	85
Figura 30. Tela Direitos de Uso.....	86
Figura 31. Estrutura de ajuda definida pela TOTVS.	87
Figura 32. Formulário de cadastro da publicidade.....	88
Figura 33: Diagrama das atividades no processo de recomendação.....	90
Figura 34. Gráfico de usuários por sexo.....	95

Figura 35. Gráfico de usuários por classificação etária	96
Figura 36. Gráfico de usuários por classe social.	97
Figura 37. Gráfico de usuários por estado civil.....	97
Figura 38. Gráfico de usuários com e sem filhos.....	98
Figura 39. Gráfico de usuários com ou sem veículos.	99
Figura 40. Gráfico de usuários por tipo de residência.....	99
Figura 41. Amostra do arquivo ARFF gerado	101
Figura 42. Arvore de decisão gerada através do algoritmo J48.....	103
Figura 43. Clusters gerados através do algoritmo Kmeans.	106
Figura 44. Clusters gerados através do algoritmo Kmeans.	108
Figura 45. Clusters gerados através do algoritmo Kmeans.	109
Figura 46. Publicidade apresentada para o publico alvo.	110
Figura 47. Publicidade apresentada para o publico alvo.	111
Figura 48. Gráfico de relevância da publicidade com a recomendação desabilitada.	114
Figura 49. Gráfico de relevância da publicidade baseada na tarefa de Classificação....	115
Figura 50. Gráfico de relevância da publicidade baseada na tarefa de Agrupamento. ...	116
Figura 51. Comparativo entre os resultados obtidos.....	117

LISTA DE TABELAS

Tabela 1. Normas definidas pela ABNT	28
Tabela 2. Exemplo de transações de cestas de compras.....	47
Tabela 3. Exemplo de conjunto de dados dos vertebrados.	48
Tabela 4. Exemplo de modelagem preditiva.....	48
Tabela 5. Tabela de valores da estatística Kappa.	55
Tabela 6. Comparação de Middiwares para o projeto Smart Marketing.	70
Tabela 7. Classificação por estrutura etária.	91
Tabela 8. Classificação social.	92
Tabela 9. Categorias e suas composições	93
Tabela 10. Descrição das atividades realizadas por dia	95

LISTA DE ABREVIATURAS E SIGLAS

ABERT	= Associação Brasileira de Emissoras de Rádio e Televisão
ABNT	= Associação Brasileira de Normas Técnicas
ANATEL	= Agencia Nacional de Telecomunicações
API	= <i>Application Programming Interface</i>
ARFF	= <i>Attribute-Relation File Format</i>
ARIB	= <i>Association of Radio Industries and Businesses</i>
ATSC	= <i>Advanced Television Systems Committee</i>
AVC	= <i>Advanced Video Coding</i>
CDMA	= <i>Code Division Multiple Access</i>
CPQD	= Centro de Pesquisa e Desenvolvimento em Telecomunicações
DSM-CC	= <i>Digital storage media command and control</i>
DVB	= <i>Digital Video Broadcasting</i>
DVB-T	= <i>Digital Video Broadcasting – Terrestrial</i>
EDGE	= <i>Enhanced Data rates for GSM Evolution</i>
ENEM	= Exame Nacional do Ensino Médio
EUA	= Estado Unidos da América
FGTS	= Fundo de Garantia por Tempo de Serviço
FIES	= Financiamento Estudantil
GSM	= <i>Global System for Mobile Communications</i>
GPL	= <i>General Public License</i>
GPS	= Global Positioning System
GPRS	= Serviço de Rádio de Pacote Geral
HTML	= <i>HyperText Markup Language</i>
IBGE	= Instituto Brasileiro de Geografia e Estatística
ICP-Brasil	= Infraestrutura de Chaves Públicas Brasileira
IDEB	= Índice de Desenvolvimento da Educação Básica
IEC	= <i>International Engineering Consortium</i>
IP	= <i>Internet Protocol</i>
IPSec	= <i>Internet Protocol Security</i>
ISDB-T	= <i>International Service Digital Broadcasting – Terrestrial</i>
ISDB-TB	= <i>International Service Digital Broadcasting – Terrestrial Brazil</i>
ISDN	= <i>Integrated Services Digital Network</i>
ISO	= <i>International Organization for Standardization</i>
ITI	= Instituto de Tecnologia da Informação

JVM	= <i>Java Virtual Machine</i>
KDD	= <i>Knowledge-Discovery in Databases</i>
LAVID	= Laboratório de Aplicações de Vídeo Digital
LTE	= <i>Long Term Evolution</i>
MPEG	= <i>Moving Picture Experts Group</i>
NBR	= Norma Brasileira
NCL	= <i>Nested Context Language</i>
OFMD	= <i>Orthogonal frequency-division multiplexing</i>
PCI	= <i>Peripheral Component Interconnect</i>
P&D	= Pesquisa e Desenvolvimento
PHP	= <i>Personal Home Page</i>
PLC	= <i>Power Line Communication</i>
PROUNI	= Programa Universidade para Todos
PSN	= <i>PlayStation Network</i>
PUC-Rio	= Pontifícia Universidade Católica Rio de Janeiro
RSS	= <i>Really Simple Syndication</i>
SBTVD	= Sistema Brasileiro de TV Digital
SET	= Sociedade Brasileira de Engenharia de Televisão
SGBD	= Sistema de Gerenciamento de Banco de Dados
SOAP	= <i>Simple Object Access Protocol</i>
SQL	= <i>Structured Query Language</i>
SMTVI	= Serviço Multiplataforma de TV Interativa
SSL	= <i>Secure Sockets Layer</i>
STB	= <i>Set-Top-Box</i>
TCP	= <i>Transmission Control Protocol</i>
TDMA	= <i>Time Division Multiple Access</i>
TS	= <i>Transport Systems</i>
TVDI	= TV Digital Interativa
UOL	= Universo OnLine
URL	= <i>Uniform Resource Locator</i>
UFPB	= Universidade Federal da Paraíba
UFPR	= Universidade Federal do Paraná
UML	= <i>Unified Modeling Language</i>
UTMS	= <i>Universal Mobile Telecommunications System</i>
WIFI	= <i>Wireless Fidelity</i>
WIMAX	= <i>Worldwide Interoperability for Microwave Access</i>
WEKA	= <i>Waikato Environment for Knowledge Analysis</i>

WWW = *World Wide Web*
XHTML = *eXtensible Hypertext Markup Language*
XML = *eXtensible Markup Language*
XDSL = *Digital Subscriber Line*

SUMÁRIO

1. INTRODUÇÃO.....	20
1.1. Motivação e Justificativas.....	22
1.2. Objetivos	22
1.3. Organização do Trabalho.....	23
2. SBTVD – SISTEMA BRASILEIRO DE TV DIGITAL.....	25
2.1. Implantação e Normas.....	26
2.2. Componentes de um sistema de TV Digital	30
2.2.1. Difusão	32
2.2.2. Recepção.....	32
2.2.3. Carrossel de Dados	33
2.2.4. Canal de Retorno	34
2.3. Middleware Ginga	35
2.3.1. Desenvolvimento e Aplicações.....	37
3. SISTEMAS DE RECOMENDAÇÃO.....	38
3.1. Coleta de Dados.....	39
3.1.1. Implícita – Baseada no Histórico de Uso	39
3.1.2. Explícita - Baseada em preenchimento de formulários	39
3.1.3. Sensível ao Contexto.....	40
3.2. Filtragem de Informações	41
3.2.1. Filtragem Baseada em Conteúdo	42
3.2.2. Filtragem Colaborativa	43
3.2.3. Filtragem Híbrida	44
3.3. Descoberta de Conhecimento em Banco de Dados	44
3.3.1. Tarefa de Associação	46
3.3.2. Tarefa de Classificação	47
3.3.2.1. Técnica de árvores de decisão	49
3.3.3. Tarefa de Regressão/Estimativa	52
3.3.4. Tarefa de Sumarização	52
3.3.5. Tarefa de Agrupamento (<i>Clustering</i>).....	52
3.3.5.1. Estatística Kappa	54
4. SEGURANÇA E PRIVACIDADE	56
4.1. Legislação e Aspectos Legais.....	58
4.2. Considerações	60

5. METODOLOGIA	62
5.1. Descrição de Cenários de Uso	62
5.2. Requisitos Funcionais e Não Funcionais	63
5.2.1. Requisitos Funcionais.....	63
5.2.2. Requisitos Não Funcionais.....	64
5.3. Diagrama de Casos de Uso	65
5.4. Modelagem de Usuário	67
5.5. Tecnologias e Ferramentas Utilizadas	68
5.5.1. Definição do Middleware para aplicações.....	69
5.5.2. Definição do Banco de Dados.....	71
5.5.3. Definição do software de descoberta de conhecimento.....	72
5.6. Arquitetura do Aplicativo	73
5.6.1. Módulo Perfil do Telespectador (Captura Explícita).....	76
5.6.2. Módulo de Validação do Telespectador	79
5.6.3. Módulo de Notícias (Captura Implícita)	81
5.6.4. Módulo de Comunicação.....	83
5.6.5. Módulo de Exibição de Conteúdo (Recomendação)	83
5.6.6. Segurança e Privacidade.....	84
5.6.7. Direitos de Uso	85
5.6.8. Ajuda (help).....	86
5.6.9. Módulo Cadastro e Armazenamento da Publicidade	87
5.7. Processo de Recomendação da Publicidade Personalizada	89
6. APLICAÇÃO DO SMART MARKETING – RESULTADOS E ANÁLISES	91
6.1. Definição dos Dados	91
6.2. Realização do Experimento	93
6.2.1. Aplicação da Descoberta do Conhecimento.....	100
6.2.2. Definição da apresentação das publicidades	110
6.3. Resultados	112
6.4. Trabalhos Relacionados	118
7. CONSIDERAÇÕES FINAIS	119
7.1. Aprimoramentos	120
7.2. Trabalhos Futuros	120
8. REFERÊNCIAS	123
9. APÊNDICES	131

APÊNDICE A – Termo de consentimento da instituição	131
APÊNDICE B - Termo de consentimento do usuário.....	133
APÊNDICE C – Planilha de Frequência dos usuários	135
APÊNDICE D – Selects da tarefa de classificação do algoritmo J48.....	136
APÊNDICE E – Resultado da aplicação do algoritmo J48.....	139
APÊNDICE F – Selects da tarefa de agrupamento do algoritmo K-Means.....	140
APÊNDICE G – Resultado da aplicação do algoritmo KMeans.....	143
APÊNDICE H - Formulário de avaliação da publicidade.....	144

1. INTRODUÇÃO

A TV Digital não significa apenas uma TV com melhor som e imagem, onde pode-se ver com maior realismo as transmissões. Existe também a interatividade entre os telespectadores e as produtoras de conteúdo para a TV (CRAIDE, 2012). Todavia, o significado da implantação do Sistema Brasileiro de TV Digital (SBTVD) transcende à aspiração de uma melhor imagem, alcançando a interatividade entre os telespectadores e a televisão. Os desafios de torná-la realmente interativa são muitos, mas os primeiros passos já foram dados, como o desenvolvimento do Ginga (GINGA, 2012), *Middleware*¹ aberto do Sistema Brasileiro de TV Digital, desenvolvido pela Pontifícia Universidade Católica - Rio de Janeiro (PUC-RIO) (CPQD, 2009). No que diz respeito às normas e padrões, a Associação Brasileira de Normas Técnicas (ABNT) publicou uma série de normas que padronizam a TV Digital e sua interatividade, conforme apresentado na seção 2.1, tabela 1.

Com os padrões já definidos, a obrigatoriedade dos aparelhos de TV em aderir ao padrão Ginga está seguindo um cronograma de implantação definido pelos Ministério do Desenvolvimento, Indústria e Comércio Exterior e a Ciência, Tecnologia e Inovação. Segundo a Portaria Interministerial nº140, de 16 de março de 2006, no período entre o dia 1º de Janeiro e 31 de Dezembro de 2013, 75% das TVs produzidas deverão conter o Ginga, e a partir de 1º de Janeiro de 2014 este número sobe para 90% das TVs produzidas.

Tendo em vista a futura disseminação das TVs Digitais com Ginga nos lares brasileiros, o desenvolvimento de ferramentas, softwares e novas técnicas para o SBTVD tornar-se-ão cada vez mais necessários, de forma a atender as necessidades e exigências de seus telespectadores e, em contrapartida, os anunciantes e estações de teledifusão.

Uma das possíveis inovações com o advento da TV interativa é a possibilidade de geração de dados decorrentes do acesso aos menus de

¹ Middleware é a camada de software localizada entre o código das aplicações e a infraestrutura de execução (hardware e sistema operacional, (GINGA, 2012)).

interação, como: home shopping, e-mail, governo eletrônico, ensino à distância, serviços de saúde, pesquisas de opinião, dentre outros conteúdos adicionais. Dessa forma, a apresentação de conteúdos personalizados baseados na navegação, se torna possível. Aranha (2000) considera que o objetivo da personalização de conteúdo é garantir que a pessoa certa receba a informação certa no momento certo.

Os benefícios oriundos da personalização publicitária na TV interativa são inúmeros, tanto para as empresas publicitárias quanto para as emissoras de TV e seus telespectadores. Para as empresas publicitárias, a personalização é uma forma especial de diferenciação de produto. Ela transforma um produto padrão em uma solução especializada para um indivíduo. Hanson (1999) afirma que diferentes graus de personalização criam um *continuum* em busca do desenvolvimento de um relacionamento, requerendo confiança e cooperação dos clientes. Já as emissoras de TV terão mais uma ferramenta de apoio para os seus anunciantes, tornando-se diferenciais no mercado atual. Por fim, para os telespectadores, trata-se de uma nova forma de interatividade, que lhes permitirá a apresentação de material publicitário de maior relevância, gerando satisfação, interesse e identificação com os produtos e serviços apresentados.

Essas estratégias de personalização demandam a aplicação criteriosa e objetiva de técnicas de descoberta do conhecimento e mineração de dados, determinando padrões de comportamento a partir de variadas fontes de dados; conseqüentemente, esses padrões transformam-se em serviços personalizados, os quais podem resultar em aumento de lucratividade ou eficácia desses serviços. (MEIRA JR *et al.*, 2002, p. 179).

A realização de personalização a partir de informações exatas, oportunas e relevantes sobre os consumidores será o grande diferencial do mercado. Sendo assim, a TV Digital irá criar outras necessidades para as empresas e consumidores em geral, deixando, para sempre, de ser apenas um equipamento de entretenimento para a família e tornando-se mais um canal de interatividade e negócios.

Portanto, este trabalho veio propor uma nova possibilidade e atratividade na utilização da TV Digital Interativa (TVDI), fazendo com que a TV Digital possa trazer benefícios não apenas aos telespectadores, mas também para os anunciantes e emissoras, já que se trata de um novo meio de publicação focado na necessidade de seus telespectadores na TV aberta brasileira.

1.1. Motivação e Justificativas

A motivação deste trabalho se dá pelo fato do SBTVD estar em fase inicial no Brasil, e a utilização da interatividade pelo Ginga ainda ser muito tímida. Desse modo, as pesquisas nesta área são relevantes, pois a partir de 1º de Janeiro de 2013 mais de 75% dos televisores fabricados no Brasil estarão equipados com o *Middleware* Ginga (POSSEBON, 2012).

Outro ponto importante, segundo Becker (2012) é que as emissoras utilizam apenas 20% dos recursos de interatividade do Ginga. Elas desenvolvem pouco conteúdo interativo porque ainda não encontraram uma forma de comercializar este novo recurso junto aos anunciantes. Com isso, os meios de obtenção de lucros com essa nova tecnologia são necessários para justificar tal aderência à tecnologia e ao investimento. Sendo assim, destacam-se, como foco neste trabalho, sistemas de recomendação com os quais, além de contemplar a personalização de conteúdos publicitários para aplicativos, poder-se-á oferecer serviços diferenciados. Isso poderá tornar o uso dos aplicativos Ginga mais atraente aos telespectadores e rentável às difusoras e anunciantes.

1.2. Objetivos

Conforme a motivação apresentada na seção anterior, referente aos desafios de uma melhor utilização e retorno financeiro dos aplicativos Ginga no SBTVD, o objetivo deste trabalho é desenvolver um aplicativo intitulado *Smart Marketing* capaz de acessar notícias via web e, por meio implícito realizar a captura e armazenamento dos dados de navegação do telespectador pelo

aplicativo, além de criar um formulário para armazenamento do perfil do telespectador.

E por fim, o aplicativo deve ser capaz de realizar a validação do telespectador através de um login e senha, apresentando-lhe publicidade personalizada no aplicativo, de acordo com o resultado do processo da descoberta do conhecimento, contribuindo, assim, para a área de personalização de publicidade em aplicativos Ginga do Sistema Brasileiro de TV Digital Interativa.

Para atingir o objetivo principal, objetivos específicos devem ser cumpridos, sendo eles:

- Integrar o aplicativo *Smart Marketing* de forma online com um Sistema de Gerenciamento de Banco de Dados (SGBD);
- Investigar as variações de *middleware* que melhor se adequam as necessidades do aplicativo a ser desenvolvido e utiliza-la.
- Aplicar padrões de metadados para a descrição de conteúdo em aplicativos padrão Ginga;
- Empregar diferentes tarefas de recomendação;
- Dirigir experimentos para analisar e avaliar a qualidade da publicidade apresentada, geradas através do *Smart Marketing*.

1.3. Organização do Trabalho

Esta dissertação está estruturada em sete capítulos.

No primeiro capítulo, foram apresentadas as motivações e justificativas do trabalho, além dos objetivos.

O capítulo 2 apresenta os conceitos da TV Digital Interativa (TVDI) suas normas, características e técnicas.

No capítulo 3, são abordados os fundamentos do sistema de recomendação e apresentadas as principais técnicas de filtragem e descoberta de conhecimento em bancos de dados.

O capítulo 4 aborda a questão de segurança e privacidade em sistemas de recomendação.

A metodologia utilizada é apresentada no capítulo 5. Este capítulo descreve possíveis cenários de uso, os requisitos funcionais e não funcionais, as tecnologias, ferramentas e softwares utilizados no desenvolvimento do aplicativo e também apresenta a arquitetura do aplicativo, contendo os seus respectivos módulos.

O capítulo 6 apresenta o experimento realizado para verificação da qualidade das recomendações no aplicativo *Smart Marketing*.

Considerações finais, aprimoramentos e trabalhos futuros são apresentados no capítulo 7 e, finalmente, o capítulo 8 encerra este trabalho apresentando as referências.

2. SBTVD – SISTEMA BRASILEIRO DE TV DIGITAL

O Sistema de TV no Brasil está passando por uma importante fase de migração, da tecnologia analógica para a digital, e com isso muitas mudanças o acompanham como: flexibilidade para oferecer multiprogramação, alta qualidade em imagem e som, acesso à internet e principalmente a interatividade através da execução de aplicativos Ginga (DTV, 2011).

Desde o surgimento das primeiras TVs em solo brasileiro, com a chegada de apenas duzentos equipamentos, não se imaginava que quarenta e nove anos depois, 95% dos lares teriam pelo menos um aparelho de TV (IBGE, 2009).

O início desta evolução aconteceu no Brasil, nos anos 1950, com a chegada dos primeiros aparelhos de televisão trazidos por Assis Chateaubriand que, na época, espalhou-as em vitrines pela cidade de São Paulo para que as pessoas pudessem assistir. Nos anos 1970, 27% das residências brasileiras já contavam com aparelhos de TV, e nesta mesma década uma evolução da tecnologia chegava ao país, a TV em cores, iniciando uma nova fase na história da TV. Outra grande evolução ocorreu no Brasil apenas em 2007, ano de lançamento da TV Digital Brasileira na cidade de São Paulo (OLIVEIRA, 2012).

O início da TV Digital Aberta Brasileira não foi animador, com preços proibitivos para a população em geral e área de cobertura restrita a alguns pontos de São Paulo sendo que menos de mil set-top-boxes² (STB) foram comercializados no ano de seu lançamento. Mas com a adesão das teledifusoras, produção de equipamentos digitais em massa, aumento da área de cobertura, no ano de 2011 já se somavam mais de 16 milhões de televisores aptos a receber o sinal digital, além de este estar disponível para 45% da população brasileira. (ANATEL, 2011).

² O Set-Top Box é o equipamento utilizado em televisões que não possuem um sintonizador digital embutido. Este trabalho adotou este termo para suas citações.

2.1. Implantação e Normas

O início dos estudos sobre o sistema de TV Digital no Brasil começou com a Sociedade Brasileira de Engenharia de Televisão (SET) e Associação Brasileira de Emissoras de Rádio e Televisão (ABERT), porém foi em 1999, com a Agência Nacional de Telecomunicações (Anatel) que tais estudos foram detalhados, a fim de estabelecer um padrão.

Firmou-se com o Centro de Pesquisa e Desenvolvimento em Telecomunicações (CPqD) uma cooperação técnica, dando início ao processo de avaliação da qualidade e viabilidade econômica para a definição do padrão. Os padrões analisados foram:

- *Digital Video Broadcasting/Terrestrial (DVB/T)*: padrão europeu, difundido pela maioria dos países do mundo como Austrália, Índia e Nova Zelândia;
- *Advanced Television Systems Committee (ATSC)*: padrão adotado pelos Estados Unidos, Canadá, México e Coréia do Sul;
- *International Service Digital Broadcasting – Terrestrial (ISDB-T)*: padrão japonês desenvolvido desde a década de 1970
- Digital Terrestrial Multimedia Broadcast (DTMB) - padrão criado e adotado na China, Hong Kong e Macau;

O resultado dos estudos indicava que o melhor padrão de TV Digital a implantar no Brasil seria o japonês. (TAVARES, 2012). Na Figura 1, são apresentados os padrões adotados pelos países; em cinza, estão os países sem padrão definido.

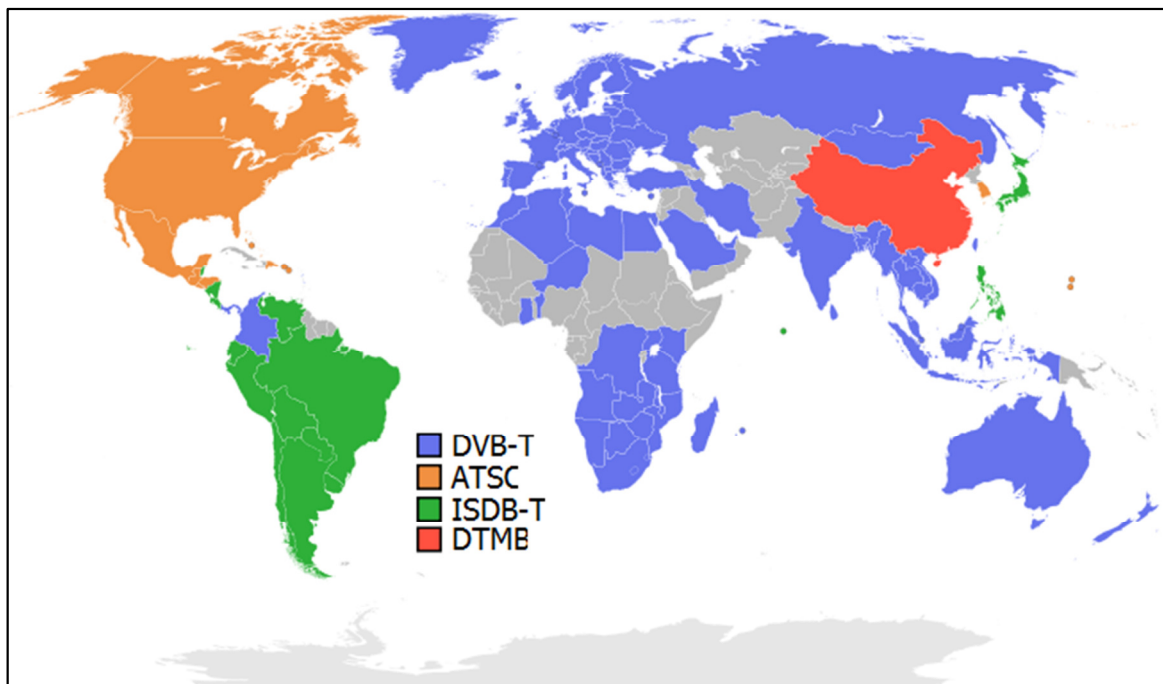


Figura 1. Padrões de TV Digital adotados pelos países (ITVBR, 2011)

Após a publicação do Decreto Presidencial em junho de 2006 (BRASIL, 2012), foi criado o fórum SBTVD composto por ministérios brasileiros, Instituto de Tecnologia da Informação (ITI), Universidades e Institutos de Pesquisa e Desenvolvimento (P&D), ABERT, fabricantes do setor eletroeletrônico e mais de 1200 pesquisadores, com o objetivo de auxiliar na melhoria do padrão, com acréscimo de novos recursos e funções, tornando o SBTVD adequado às necessidades do Brasil.

Tendo como base o ISDB-T, novos recursos foram acrescentados durante três anos de pesquisas e desenvolvimentos, resultando no *International Service Digital Broadcasting – Terrestrial Brazilian* (ISDB-TB). Foram acrescentados (ITVBR, 2012):

- Compressão de vídeo *Moving Picture Experts Group-4 Advanced Video Coding* (MPEG-4 AVC) (H.264), possibilitando uma maior capacidade de dados em num canal;
- Apresentação do *Middleware* Ginga que utiliza ambiente declarativo e procedural, permitindo aplicações interativas e complexas.

Além do acréscimo de novos recursos, adaptações foram realizadas para o SBTVD, as principais foram:

- Máscaras de emissão dos transmissores foram alteradas para atender aos mais diversos cenários, melhorando a questão de interferência entre emissoras, visto que o Brasil contém o espectro de frequência relativamente congestionado;
- Inclusão de conjuntos de caracteres para línguas latinas nos sistemas de multiplexação, sinalização e as estruturas de dados;
- Estabelecimento de 30fps como taxa de apresentação de quadros para receptores portáteis, significando uma maior qualidade para os aparelhos móveis como TVs portáteis, celulares, *tablets* etc.

Para atender às normas estabelecidas no Decreto 5820/2003 (BRASIL, 2006), o fórum de desenvolvimento do SBTVD solicitou à ABNT a elaboração de documentos para padronização da TV Digital Brasileira (Tabela 1). Para isso, a ABNT criou uma Comissão de Estudo Especial Temporária (CEET) (ABNT, 2006). Em 2012, vinte e quatro documentos técnicos em português, espanhol e inglês, com mais de três mil páginas, haviam sido publicadas pela ABNT, detalhando todo o padrão SBTVD, incluindo: transmissão, codificação, multiplexação, receptores, segurança, *middleware*, canal de interatividade, guia de operação e acessibilidade.

Tabela 1. Normas definidas pela ABNT (Adaptado de ABNT 2012)

Assunto	ABNT NBR	Partes
Transmissão	15601	Parte 1
Codificação	15602	Parte 1
		Parte 2
		Parte 3
Multiplexação	15603	Parte 1
		Parte 2
		Parte 3
Receptores	15604	Parte 1
Segurança	15605	Parte 1
Middleware	15606	Parte 1
		Parte 2
		Parte 3

		Parte 4
		Parte 5
		Parte 6
		Parte 7
		Parte 8
		Parte 9
Canal de Interatividade	15607	Parte 1
Guia de Operação	15608	Parte 1
		Parte 2
		Parte 3
Acessibilidade	15610	Parte 1
		Parte 2

Após a definição e implantação do SBTVD, as transmissões comerciais com o padrão ISDB-TB se iniciaram em dezembro de 2007. Em 2012, a TV Digital já está presente em todas as capitais do Brasil, atendendo a mais de 45% da população (ANATEL, 2011); em estados como São Paulo e Rio de Janeiro, este valor salta para 70%.

De acordo com o cronograma oficial (Figura 2) apresentado pela DTV, em junho de 2013 todo o país estará coberto com o sinal digital e, em junho de 2016, o sinal analógico será desativado (DTV, 2011).

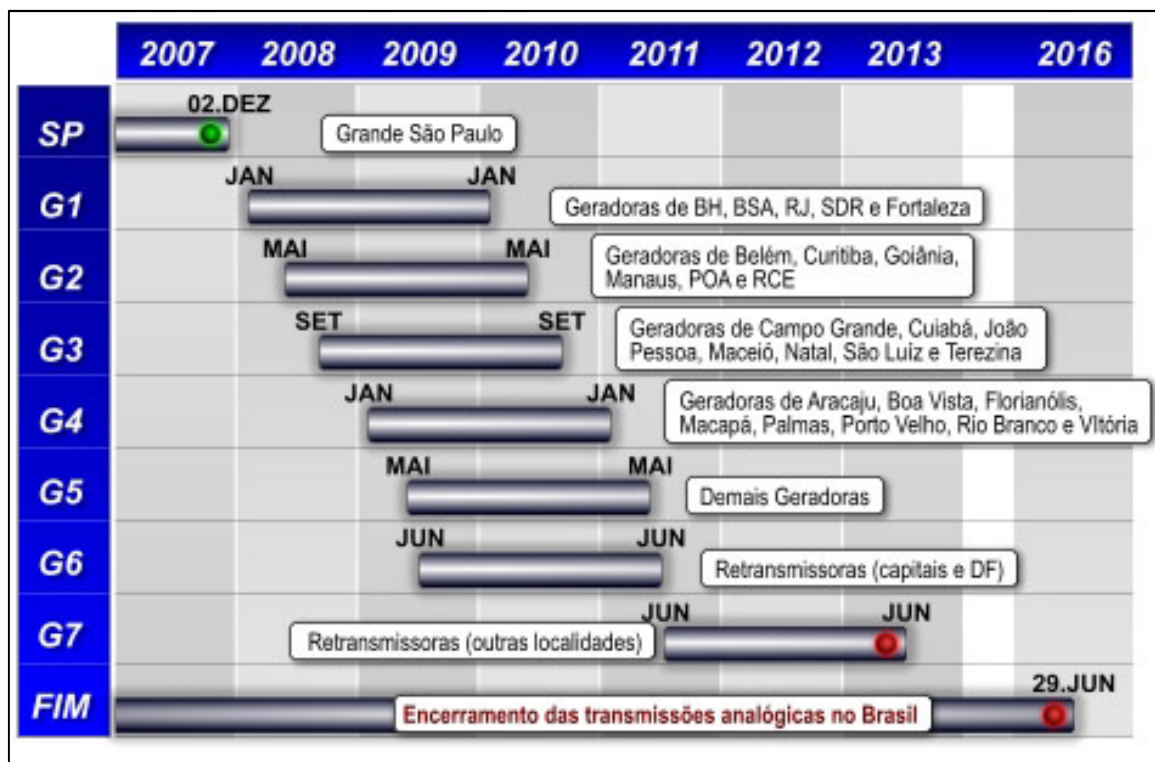


Figura 2. Cronograma de implantação do SBTVD no Brasil (Fonte DTV, 2011).

Com qualidade, normas definidas e pioneirismo em relação à América do Sul na adoção de um padrão de TV Digital, o Brasil, em conjunto com o Japão, passou a oferecer o SBTVD a outros países.

Em 2012, países como: Argentina, Bolívia, Chile, Costa Rica, Equador, Paraguai, Peru, Uruguai e Venezuela adotaram o padrão nipo-brasileiro, e mais de 45 países da África estão em testes para a adesão ao sistema, que se difere dos demais em aspectos sociais como inclusão digital, qualidade de imagem e som, robustez na captura do sinal, além da mobilidade sem custos e interatividade.

2.2. Componentes de um sistema de TV Digital

A Norma Brasileira (NBR) 15607-1 apresenta um modelo de referência completo, adotado pelo padrão brasileiro (Figura 3). O modelo é dividido em cinco camadas, ilustrando os componentes do SBTVD (ABNT, 2008):

- Aplicação: realiza a execução dos aplicativos multimídia desenvolvidos em NCL/Lua e Java;
- *Middleware*: provê um *Application Programming Interface (API)*, fazendo com que as aplicações sejam executadas independentes do modelo de hardware;
- Compressão: responsável pela compressão e descompressão do fluxo de áudio e vídeo;
- Transporte: realiza a multiplexação e demultiplexação dos programas exibidos na TV;
- Transmissão: responsável pela codificação/decodificação do sinal além da sintonia e modulação/demodulação;

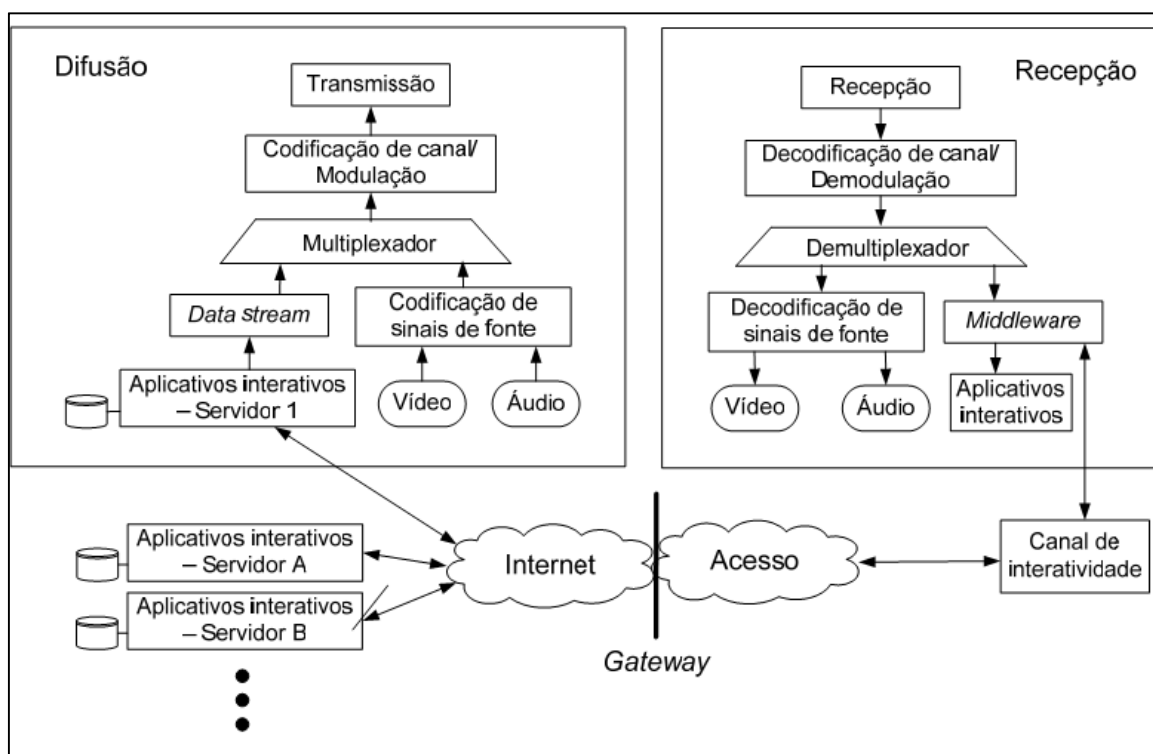


Figura 3. Modelo contendo os componentes do SBTVD (ABNT NBR 15607-1 2008).

2.2.1. Difusão

Para o processo de transmissão, a ABNT NBR 15601/2007 define que as entradas, contendo feixe de dados *Transport System* (TS) (definidas no sistema MPEG-4), devem ser demultiplexadas para que se tornem apenas um único TS; com isso, este será enviado para codificação no canal múltiplo, modulado, amplificado e então enviado como um sinal *Orthogonal Frequency - Division Multiplexing* (OFDM). A Figura 4 ilustra esse processo.

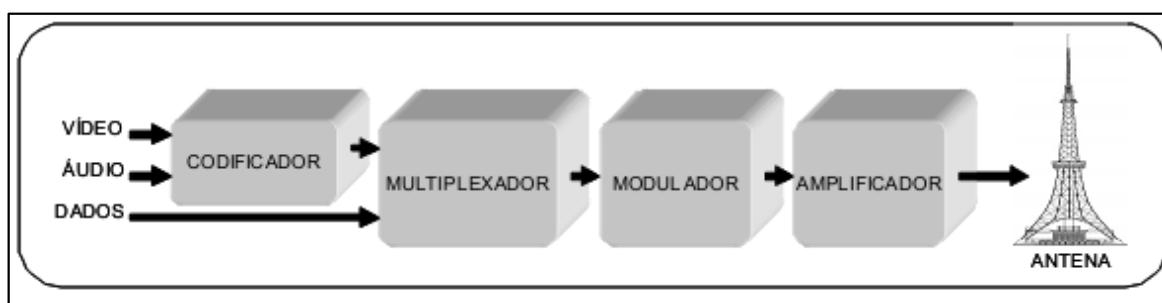


Figura 4. Etapas para difusão do sinal (ABNT NBR 15601/2007).

2.2.2. Recepção

A Figura 5 exibe o fluxo do processo de recepção, partindo da captura do sinal e transmissão pela difusora, seja por meio de antenas ou cabos. O sinal é recebido pelo sintonizador digital e demodulado, para que seja extraído o fluxo de transporte MPEG-4. O fluxo é enviado pelo demultiplexador, que realiza a extração dos fluxos elementares, e um sistema de acesso condicional os decifra. Após esse procedimento, os fluxos de áudio e vídeo são encaminhados para o decodificador e convertidos para o formato de exibição adequado, de acordo com o equipamento de TV utilizado.



Figura 5. Etapas da Recepção - Adaptado de Becker, 2005.

2.2.3. Carrossel de Dados

A definição de carrossel de dados, segundo a ABNT NBR 15604/2007, é enviar qualquer conjunto de dados ciclicamente, para que esses possam ser obtidos, via difusão, em um intervalo de tempo tão longo quanto necessário (ABNT NBR 15604/2007).

A TV Digital utiliza o método apresentado na figura 6, Carrossel de Dados, do Padrão *Digital Storage Media Command and Control* (DSM-CC), sendo

que seus dados são representados por objetos que contêm os seus respectivos atributos.

A vantagem em relação aos demais protocolos de download se dá pelo fato de suas operações serem rápidas e leves, atendendo a equipamentos que contém pouca memória, como é o caso do STB da TV Digital Brasileira.

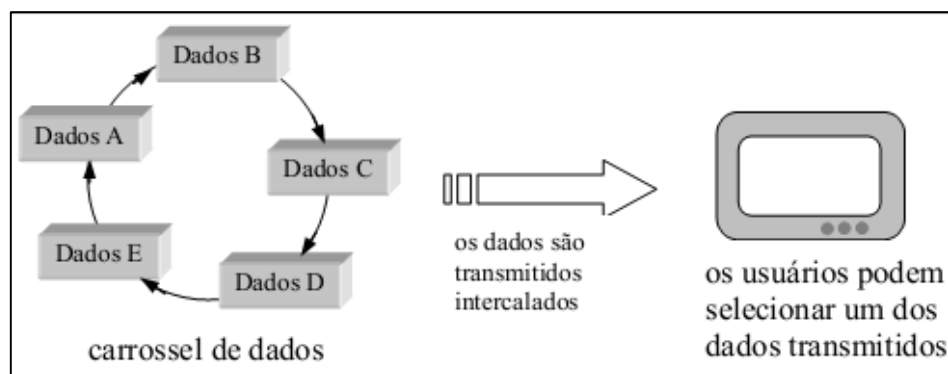


Figura 6. Funcionamento do carrossel de dados (Becker, 2005)

O funcionamento é baseado na transmissão realizada pelas difusoras de forma periódica ou cíclica dos dados de modo intercalado. E, quando recebido em sua totalidade pelo receptor, é apresentado em forma de aplicativo interativo. Com isso, o receptor apenas aguarda o próximo envio quando precisar de uma determinada informação adicional (BECKER, 2005).

2.2.4. Canal de Retorno

Um dos principais recursos do Sistema Brasileiro de TV Digital é a oferta de interatividade, recurso que transcende à forma de uma simples ferramenta tecnológica, tornando-se um meio de inclusão social e digital, como previsto no Decreto Presidencial 4901 (BRASIL, 2006), já que possibilita o acesso a jogos, leitura de livros, pesquisas, programas com cunho social como Extratos de Fundo de Garantia por Tempo de Serviço (FGTS), acesso aos programas de baixa renda, agendamento de consultas médicas, além de alertas de catástrofes como tsunamis, chuvas torrenciais e muitas outras possibilidades que tal tecnologia proporciona (ITVBR, 2011).

Para o funcionamento pleno da interatividade, é necessário o acesso à internet, seja por meio da telefonia celular (*Code Division Multiple Access* (CDMA), *Enhanced Data Rates for GSM Evolution* (EDGE), *Time Division Multiple Access* (TDMA), *Global System for Mobile Communications* (GSM), *Universal Mobile Telecommunications System* (UTMS) ou *Long Term Evolution* (LTE)), pela telefonia fixa através da *Digital Subscriber Line* (xDSL), *Integrad Services Digital Network* (ISDN), Radio (exemplo: *Wireless Fidelity* (Wi-fi), *Worldwide Interoperability for Microwave Access* (WiMAX), Satélite, *Power Line Communication* (PLC) ou por cabo, através das operadoras de TV a cabo.

Segundo Benneton (2003), o canal de interação não deve ser homogêneo. Portanto, cada telespectador deve aderir ao meio de comunicação que melhor se adapte às suas necessidades.

2.3. Middleware Ginga

O *Middleware* é uma camada de software intermediária entre o código das aplicações e a infraestrutura de execução. Composto por um conjunto de APIs padronizadas, o *Middleware* é um dos principais componentes da TV Digital, sendo responsável pela execução das linguagens oferecidas. O Sistema Brasileiro de TV Digital utiliza o *Middleware* Ginga, projeto desenvolvido pela Telemídia da PUC-Rio e do Laboratório de Aplicações de Vídeo Digital (LAVID) da Universidade Federal da Paraíba (UFPB).

A constituição do conjunto de tecnologias e inovações para o SBTVD foi definida pelas normas ABNT NBR 15606-1 a 15606-7; na figura 7 é apresentada a arquitetura do *Middleware* Ginga do Sistema Brasileiro de TV Digital. No que se refere à segurança, a ABNT NBR 15605 apresenta as normas, políticas de acesso e execução dos componentes.

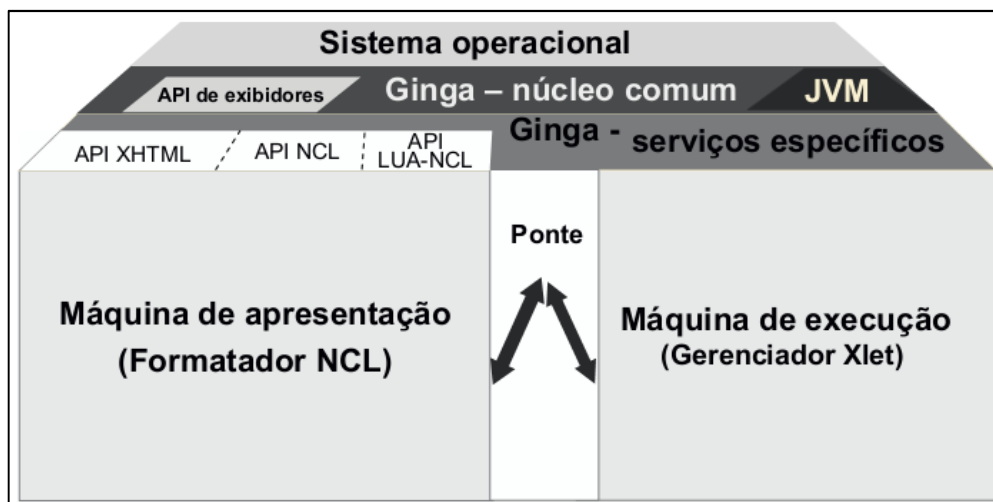


Figura 7. Arquitetura do *Middleware* Ginga (ABNT 15606-2).

Segundo a ABNT 15606-2, conforme apresentado na figura acima, a arquitetura do Ginga é composta por:

- Máquina de execução, contendo um subsistema responsável pela avaliação e execução de aplicações imperativas, consistindo em instruções em linguagem de computador, conteúdos de mídia associados e outros dados. Exemplo: Ginga-J;
- Máquina de apresentação, a qual consiste em um subsistema responsável por analisar e apresentar aplicações declarativas, com conteúdos como áudio, vídeo, gráficos e texto, por exemplo, o Ginga-NCL;
- A ponte contém um mecanismo que permite o mapeamento bidirecional entre APIs Java e os objetos e métodos do DOM, ECMAScript e LUA Script;
- O núcleo comum ginga é constituído de decodificadores de conteúdo comuns, bem como de procedimentos para obter conteúdos transportados em fluxos de transportes MPEG-2, utilizando-se do canal de interatividade;
- As APIs de exibidores são módulos plug-in da máquina de apresentação (XHTML, NCL, Lua-NCL).

2.3.1. Desenvolvimento e Aplicações

Com a utilização de um *middleware* próprio e de código aberto, um novo mercado se abriu no país: o desenvolvimento de aplicativos para a TV Digital. Seu desenvolvimento pode utilizar-se de dois ambientes: o procedural ou declarativo.

O Ambiente Declarativo Ginga (Ginga-NCL), também conhecido como máquina de apresentação, é capaz de interpretar aplicativos desenvolvidos a partir da linguagem NCL (*Nested Context Language*) criado pela PUC-Rio, possibilitando aplicações multimídia contendo vídeos, áudios, imagens e textos (SANT'ANNA, 2009). Atualmente, algumas ferramentas de criação estão disponíveis para utilização como o Composer, Berimbau iTV e mais recentemente o IRIS, lançado em 2012 pelo CPqD por meio do projeto de Serviços Multiplataforma de TV Interativa (SMTVI).

No que diz respeito à linguagem Lua, por se tratar de uma linguagem de script imperativa, procedural, pequena, reflexiva e leve, ela foi adotada pelo NCL para a TV Digital Interativa. Com isso, o desenvolvimento de aplicativos embarcados, softwares complexos, jogos, entre outras possibilidades tornaram-se viáveis no SBTVD. A linguagem Lua se assemelha às linguagens: Python, Ruby e Icon (LUA, 2012).

Já o ambiente procedural utiliza-se do Ginga-J, tendo como componente principal o JVM (Java Virtual Machine), contendo algumas APIs definidas, dentre elas a JavaTV desenvolvida pela Sun. (LAVID, 2012).

3. SISTEMAS DE RECOMENDAÇÃO

Os sistemas de recomendação de publicidade aplicados na TV Digital Interativa têm como principal objetivo apresentar itens de interesse pessoal, tornando o produto apresentado de maior relevância ao telespectador e fazendo com que a publicidade atinja o seu público alvo.

Burke (2002) define o sistema de recomendação como um sistema que direciona o usuário (telespectador), de maneira personalizada para objetos relevantes ou de interesse a partir de uma lista de opções possíveis. Resnick e Varian (1997) apresentam os sistemas de recomendação como uma forma de auxiliar no aumento da capacidade e eficácia desse processo de indicação, já bastante conhecido na relação social entre seres humanos.

Eirinaki *et al.* (2004) definem personalização como qualquer ação que relacione informações para o usuário. Herlocker *et al.* (2004) define um sistema de recomendação como um antecipador de itens que um usuário achará relevante e apresenta sugestões ao usuário, o qual é beneficiado pela filtragem de itens.

O processo de recomendação na internet já está difundido, sendo utilizado principalmente por grandes empresas de vendas online, locadoras virtuais, sites de notícias, entre outros nichos de mercado. Já no domínio da TV Digital Interativa, os sistemas de recomendação devem seguir os mesmos padrões da web, sendo necessária a utilização de processos de captura de dados.

Neste capítulo, são apresentadas três formas de se realizar a captura de dados: a implícita, a explícita e a sensível ao contexto. Também são apresentadas as técnicas mais utilizadas de sistemas de recomendação, além da descoberta de conhecimento em banco de dados contendo suas tarefas, técnicas e algoritmos.

3.1. Coleta de Dados

Com o advento dos sistemas computacionais, a captura e armazenamento de dados pessoais podem ser considerados rotineiros. Por exemplo, ao abrir uma conta bancária, realizar um exame de rotina, realizar uma compra online ou até mesmo acessar um site de relacionamentos, os dados de navegação são capturados tanto de forma direta (explícita) por meio de formulários, quanto implícita (através de algoritmos de forma transparente). Com tais dados armazenados, os sistemas de recomendação apresentam itens personalizados (REATEGUI *et al*, 2005). A seguir, são apresentados detalhes das formas de obtenção de dados.

3.1.1. Implícita – Baseada no Histórico de Uso

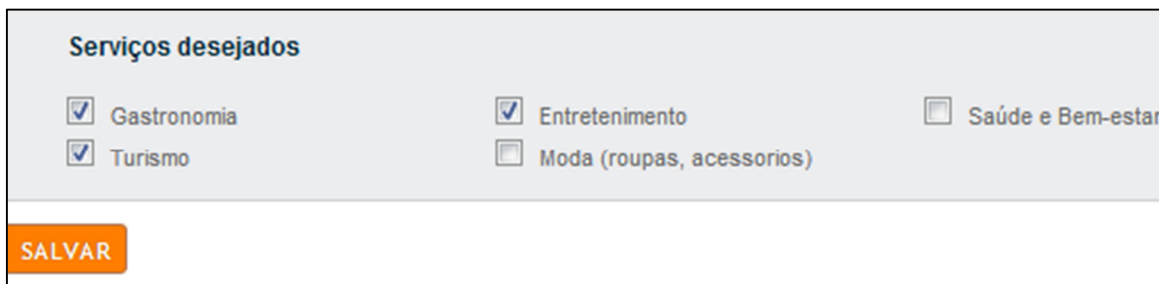
De uma maneira simples, a obtenção dos dados é realizada por meio da captura e armazenamento dos dados, os quais são obtidos de modo transparente ao telespectador, diretamente no código da página ou software acessado (de forma oculta) por vários canais como, por exemplo: últimas compras (*e-commerce*), histórico de acessos, menus navegados, canais assistidos, aplicativos baixados, entre outros (REATEGUI *et al*, 2005).

Essa forma de captura muitas vezes ocorre sem o conhecimento do telespectador, sendo considerada uma invasão de privacidade por especialistas como Neto (2012), Hamann (2011) e Rava (2012). O capítulo 4 aborda melhor esse tema e apresenta uma discussão em relação à segurança e privacidade dos dados.

3.1.2. Explícita - Baseada em preenchimento de formulários

A forma explícita de obtenção de dados (Figura 8) é feita de forma espontânea pelo telespectador, indicando o(s) seu(s) interesse(s) através de

formulários, críticas ou *ratings* (classificadores), respostas a perguntas rápidas etc. (REATEGUI *et al*, 2005).



The image shows a user interface titled "Serviços desejados" (Desired Services). It contains five checkboxes with corresponding labels: "Gastronomia" (checked), "Turismo" (checked), "Entretenimento" (checked), "Moda (roupas, acessórios)" (unchecked), and "Saúde e Bem-estar" (unchecked). Below the checkboxes is an orange button labeled "SALVAR" (Save).

Figura 8. Exemplo de coleta explícita (Fonte: Peixe Urbano, 2011)

3.1.3. Sensível ao Contexto

A abordagem sensível ao contexto explora as preferências explícitas, implícitas e também as informações contextuais, como dia, horário, dados da programação de TV, localização geográfica do usuário, etc.

A Figura 9 apresenta o exemplo da empresa Google que aplica o meio de obtenção de dados sensível ao contexto em seus produtos. Todavia, o usuário pode, a qualquer momento, editar o que foi capturado por meio de um painel de preferências.

Empregos e educação	Remover
Empregos e educação - Empregos - Listas de emprego	Remover
Internet e telecomunicações - E-mail e mensagens	Remover
Internet e telecomunicações - Serviços da web - Marketing e otimização de mecanismos de pesquisa	Remover
Localidades mundiais - América Latina - América do Sul - Brasil - Sul do Brasil - Paraná	Remover
Referência - Diretórios e listagens - Listas de contatos comerciais e pessoais	Remover
Referência - Referência geral - Como fazer, faça você mesmo e conteúdo profissional	Remover
Referência - Referência geral - Ferramentas de referência e calculadoras	Remover
Viagens - Empresas aéreas	Remover
Viagens - Hotéis e acomodações	Remover
Adicionar ou editar interesses	
Suas informações demográficas	
Veja abaixo as informações demográficas inferidas que o Google associou a seu cookie. Inferimos sua faixa etária e seu sexo com base nos sites que você visitou.	
Idade: 25 a 34 anos	Remover
Gênero: masculino	Remover
Idioma: espanhol	Remover
Idioma: português (Portugal)	Remover

Figura 9. Exemplo de coleta sensível ao contexto (Traduzido de: Google Ads, 2012)

3.2. Filtragem de Informações

A quantidade de informações como: textos, vídeos, áudios, entre outros conteúdos gerados pelo mundo cresce vertiginosamente; segundo Bloch (2011), em 2020, a quantidade de informações gerada será aproximadamente 44 vezes maior que em 2009, conforme pode se observar na figura 10.

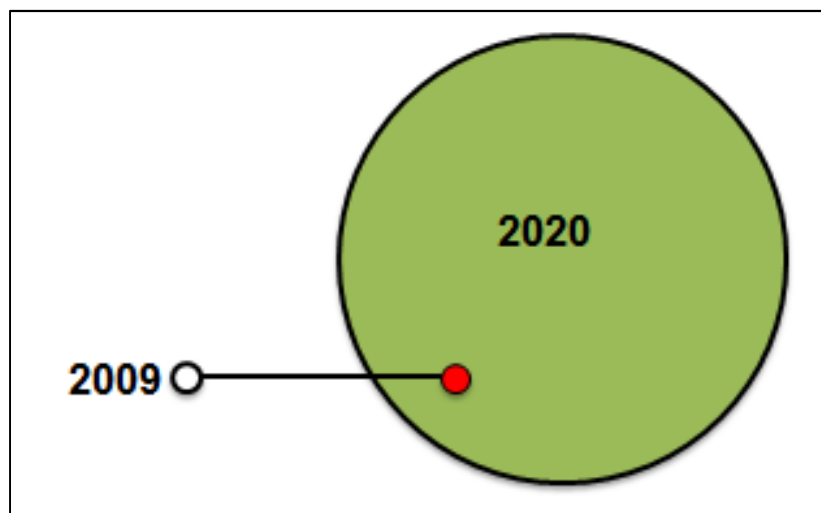


Figura 10. Proporção de dados em 2009 e em 2020 (Adaptado de BLOCH 2011)

Com tal crescimento, é necessário filtrar a informação que chega às pessoas (RADFAHRER, 2012). A área de recuperação de informações em sistemas de recomendação tem papel fundamental nesse processo, já que, através de suas técnicas de filtragem, o problema de sobrecarga de informações pode ser minimizado. Os capítulos 3.2.1, 3.2.2 e 3.2.3 apresentam, respectivamente, as técnicas de filtragem baseada em conteúdo, filtragem colaborativa e filtragem híbrida.

3.2.1. Filtragem Baseada em Conteúdo

A técnica de filtragem baseada em conteúdo utiliza-se das preferências do usuário por meio da recuperação de informação, podendo, assim, realizar a apresentação de novos itens. Segundo Herlocker *et al.* (2004), as preferências são fornecidas pelo próprio usuário e também pela avaliação de conteúdos ou itens que o usuário tenha recomendado ou consumido anteriormente. Sendo assim, baseando-se em registros do histórico, podem-se apresentar recomendações de acordo com a sua similaridade.

Todavia, a filtragem baseada em conteúdo apresenta algumas limitações. Adomavicius e Tuzhilin (2005) citam problemas de análise limitada do conteúdo, como por exemplo, a complexidade em empregar a técnica em

imagens, vídeos ou áudios, limitando-se a textos. Outro empecilho consiste no que diz respeito à superespecialização de recomendação, tornando a apresentação apenas de itens com semelhanças aos já adquiridos ou avaliados, comprometendo a diversidade da recomendação.

Outro problema conhecido é a “partida a frio” (*cold-start problem*), isto é, o algoritmo de filtragem baseado em conteúdo perde em eficiência nas estimativas até que um número suficiente de avaliações seja realizado (FILHO *et al.*, 2006).

3.2.2. Filtragem Colaborativa

Ao contrário da filtragem baseada em conteúdo, a filtragem colaborativa busca explorar experiências de outros usuários em relação a um determinado item. Portanto, não é considerado o conteúdo do item e sim a opinião dos demais usuários. Com o advento das redes sociais, compras coletivas e adesão de novos internautas, a filtragem colaborativa ganha mais força, transferindo o "boca a boca" para o meio digital.

As experiências dos usuários podem ser expostas em forma de avaliações, classificações e outros meios. Com a utilização desses dados, a filtragem colaborativa busca realizar a similaridade das avaliações do usuário com os demais, a fim de determinar outros usuários com perfis semelhantes, realizando futuras recomendações. Herlocker *et al.* (2004) destaca que a filtragem colaborativa possibilita a criação de comunidades, visto que os usuários possuem interesses comuns. Adomavicius e Tuzhilin (2005) apontam limitações em relação à filtragem colaborativa, como o problema de um novo item: como não há recomendações desse item, o mesmo só será indicado a outros usuários a partir de um número considerável de indicações. Para que o sistema de recomendação baseado em filtragem colaborativa possa atingir seu objetivo, é preciso que um número mínimo de usuários realize suas avaliações; como a avaliação de produtos não é obrigatória nos meios digitais, essa técnica acaba se tornando pouco eficiente.

3.2.3. Filtragem Híbrida

A utilização da filtragem híbrida busca utilizar as técnicas de filtragem baseada em conteúdo e filtragem colaborativa, além de outras técnicas existentes, com o intuito de minimizar as suas limitações, ressaltando as vantagens de cada uma. Burke (2002) afirma que a utilização de várias técnicas busca trazer bons resultados nas recomendações, visto que as desvantagens de uma podem ser supridas por outra.

3.3. Descoberta de Conhecimento em Banco de Dados

O processo de descoberta de conhecimento em bases de dados, em inglês, *Knowledge-Discovery in Databases* (KDD), necessita de várias etapas e apoio de ferramentas computacionais para que se possa cumpri-las. Sua aplicação pode ser feita nas mais diversas áreas, como publicidade, detecção de fraudes bancárias, instituições governamentais, controle de qualidade, instituições financeiras, medicina, dentre outras. Devido ao fato de ser multidisciplinar, a descoberta de conhecimento contém diversas propostas para a definição das etapas do processo de KDD.

A definição dada por Han e Kamber (2001) diz que o KDD consiste em sete passos: Limpeza, Integração, Seleção, Transformação, Mineração de Dados, Avaliação e Apresentação. Já Dias (2011) propõe a divisão do processo de KDD em seis passos: Definição de Objetivos, Preparação de Dados, Definição de um estudo, Construção de um modelo, Análise do Modelo e Predição.

Fayyad (1996) por sua vez apresenta o processo de KDD mais conhecido da literatura, apresentando-a em cinco passos (Figura 11): Seleção, Pré-Processamento, Transformação, Mineração e Interpretação/Avaliação.

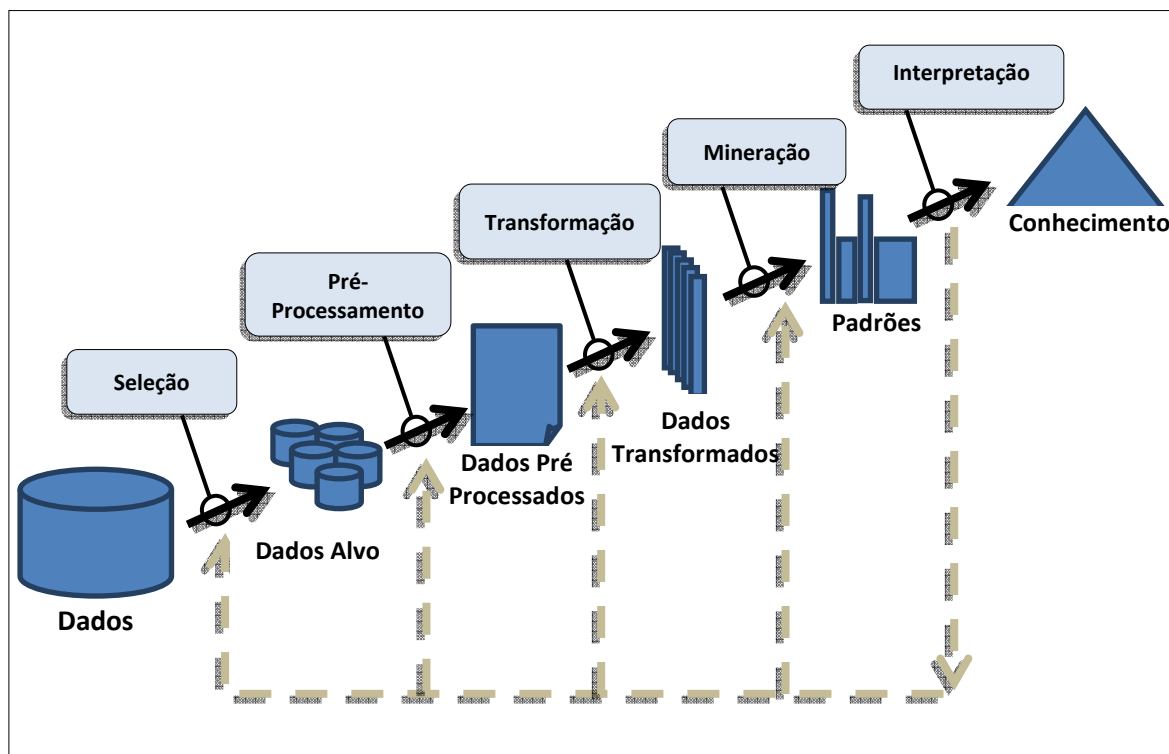


Figura 11. Modelo do processo de KDD adaptado de Fayyad (1996).

Seleção - O entendimento bem como os objetivos do problema devem estar claros para que se torne viável a seleção das bases de dados alvo. Ou seja, deve ser feita a escolha do conjunto ou subconjunto de dados considerados úteis para serem processados.

Pré-Processamento - Com os dados já definidos, o pré-processamento se encarrega de realizar a limpeza dos dados, eliminando ruídos, registros duplicados ou vazios, acentuação ou dados similares. Por exemplo, campo UF (Unidade Federativa) pode ser preenchido de diferentes formas como: Paraná, PR, Parana, etc. Nesta etapa, também é realizada a integração dos dados em bases distintas, tornando seus valores mais confiáveis.

Transformação - Com a base de dados consolidada, a transformação de dados é responsável, de modo geral, pela redução da base, já que em sua maioria são grandes bases de dados, tornando o processo ineficiente e lento em seu processamento. Para isso, é feita uma representação eficiente dos dados, por

meio de uma amostragem (*sampling*), além de outras técnicas e algoritmos de redução de dados.

Mineração de Dados (*Data mining*) – É feita a busca efetiva por aplicação dos métodos e definição de técnicas e algoritmos de mineração de dados a ser aplicado.

Interpretação e avaliação - Pode-se utilizar o conhecimento adquirido, todavia a necessidade de especialistas nesta etapa é fundamental para a descoberta do conhecimento, visto que a interpretação e avaliação são feitas por meio de conhecimento de mundo do especialista e, caso após a interpretação o resultado obtido não seja satisfatório, pode-se retornar a etapas anteriores e reformular as escolhas utilizadas.

Por fim, para o máximo aproveitamento de um sistema de descoberta de conhecimento, o especialista necessita ter um sólido entendimento do negócio da organização a ser trabalhada, para que seja capaz de selecionar corretamente os conjuntos/subconjuntos de dados e as classes de padrões relevantes (RABELO, 2007).

3.3.1. Tarefa de Associação

Segundo Dias (2001), a tarefa de associação consiste em localizar conjuntos de itens que ocorrem simultânea e frequentemente em um sistema de banco de dados, por exemplo: Leite \Rightarrow Pão.

Outra situação seria Pão \wedge Manteiga \Rightarrow Café. A definição dada por Tan *et al.* (2009) é que a tarefa de associação é um meio de descobrir relacionamentos interessantes escondidos em conjuntos grandes de dados. Conforme a Tabela 2, existe associação entre as vendas de fraldas e cervejas, onde a maioria dos clientes que compram fraldas também compraram cervejas.

Tabela 2. Exemplo de transações de cestas de compras

Identificador	Itens
1	{Pão, Leite}
2	{Pão, Fraldas, Cerveja, Ovos}
3	{Leite, Fraldas, Cerveja, Cola}
4	{Pão, Leite, Fraldas, Cerveja}
5	{Pão, Leite, Fraldas, Cola}

{Fraldas \Rightarrow Cerveja}

Fonte: TAN *et al.* (2011)

O exemplo foi baseado em uma venda por varejo, porém áreas como a bioinformática, diagnósticos médicos, mineração online, processamento de dados científicos, também podem utilizar-se de tal tarefa.

3.3.2. Tarefa de Classificação

A definição dada por Tan *et al.* (2009) para a Tarefa de Classificação é aprender uma função alvo 'f' que mapeie cada conjunto de atributos 'x' para um dos rótulos de classes y predeterminados.

Sua utilização é conhecida nas mais diversas áreas, desde a detecção de *spams* em caixas de e-mails considerando o conteúdo do cabeçalho, até a categorização de células cancerígenas (malignas ou benignas), possibilitando também a classificação de galáxias de acordo com a sua estrutura.

A Tarefa de Classificação trabalha com um conjunto de registros em sua entrada. Cada registro, também chamado de instância, se caracteriza por utilizar 'x' e 'y', sendo 'x' o conjunto de atributos e 'y' o atributo especial, também conhecido como atributo alvo ou de categorização.

Modelagem Descritiva: Tem como objetivo ser utilizada para a distinção entre objetos e classes distintas. O exemplo apresentado por Tan *et al.* (2009) na

Tabela 3 apresenta as características que definem um vertebrado nas categorias: mamífero, réptil, ave ou anfíbio.

Tabela 3. Exemplo de conjunto de dados dos vertebrados.

Nome	Temperatura Corporal	Cobertura	Ser Aquático	Ser Aéreo	Possui Pernas	Hiberna	Rotulo da Classe
Humano	Sangue Quente	Cabelo	Não	Não	Sim	Não	Mamífero
Salmão	Sangue Frio	Escamas	Sim	Não	Não	Não	Peixe
Sapo	Sangue Frio	Nenhuma	Sim	Não	Sim	Sim	Anfíbio
Pinguim	Sangue Quente	Penas	Semi	Não	Sim	Não	Ave
Enguia	Sangue Frio	Escamas	Sim	Não	Não	Não	Anfíbio

Fonte: TAN *et al.* (2009)

A modelagem preditiva possibilita prever rótulos de classe de registros desconhecidos. Para isso, é necessária a utilização de um modelo de classificação desenvolvido a partir de um conjunto de dados, conforme apresentado na Tabela 4, podendo assim determinar a classe à qual o objeto pertence.

Tabela 4. Exemplo de modelagem preditiva.

Nome	Temperatura Corporal	Cobertura	Ser Aquático	Ser Aéreo	Possui Pernas	Hiberna	Rotulo da Classe
Monstro de Gila	Sangue Frio	Escamas	Não	Não	Sim	Sim	?

Fonte: TAN *et al.* (2011)

A necessidade da utilização de modelos de classificação se dá pelo fato de existirem várias técnicas de classificação, sendo as mais conhecidas: árvores de decisão, baseado em regras, redes neurais, máquinas de vetor de

suporte e classificadores *bayes*. A partir da utilização de um modelo de classificação, é aplicado um algoritmo de aprendizagem, de modo a identificar o modelo mais adequado para relacionar o conjunto de atributos e o rótulo da classe de dados (TAN *et al.*, 2009).

O modelo definido pelo algoritmo de aprendizagem irá prever, de maneira adequada, os rótulos de classe de registros sem conhecê-lo. Sendo assim, um algoritmo de aprendizagem bem estruturado irá gerar modelos com boa capacidade de precisão em relação aos rótulos de classes de registros não conhecidos previamente (TAN *et al.*, 2009). Em sua grande maioria, os algoritmos de classificação usam modelos que atinjam a maior precisão ou a menor taxa de erro quando se faz uso do conjunto de testes.

3.3.2.1. Técnica de árvores de decisão

Considerada uma técnica de classificação simples, a árvore de decisão é muito utilizada em mineração de dados (GARCIA, 2000). Martins *et al.* (2012) a descreve como a estrutura em forma de uma árvore, sendo que cada nó interno infere um teste em um atributo; o resultado do teste é representado pelo ramo e cada folha nó contém um rótulo de classe. O nó zero (mais alto) é conhecido como raiz. Em suma, a árvore de decisão tem como objetivo classificar a entrada em uma classe específica. Sua utilização pode ser feita por meio das tarefas de classificação ou estimativa.

A aplicabilidade da árvore de decisão obtém melhores resultados quando o objetivo é categorizar dados de arquivos, assim como gerar regras que possam ser facilmente entendidas em linguagem natural.

Dias (2001) apresenta quatro exemplos de aplicações da técnica de árvores de decisão: o primeiro seria a classificação de pedidos de crédito como de baixo, médio e alto risco aos clientes que solicitam crédito; outra possibilidade seria para a obtenção de pedidos de seguros fraudulentos; também é possível identificar a melhor forma de tratamento de um determinado paciente; e, por fim,

prever a demanda de um determinado consumidor para um novo produto e/ou serviço. Sua estrutura é dividida em nós:

- Nó raiz: parte-se de zero, não possui arestas anteriores, podendo conter uma ou mais arestas;
- Nós internos: necessariamente chega uma aresta até ele e parte-se para duas ou mais;
- Nós folhas ou terminais: possui necessariamente uma aresta, sem saída.

Na figura 12, é apresentado um exemplo típico de árvore de decisão para a compra de um computador.

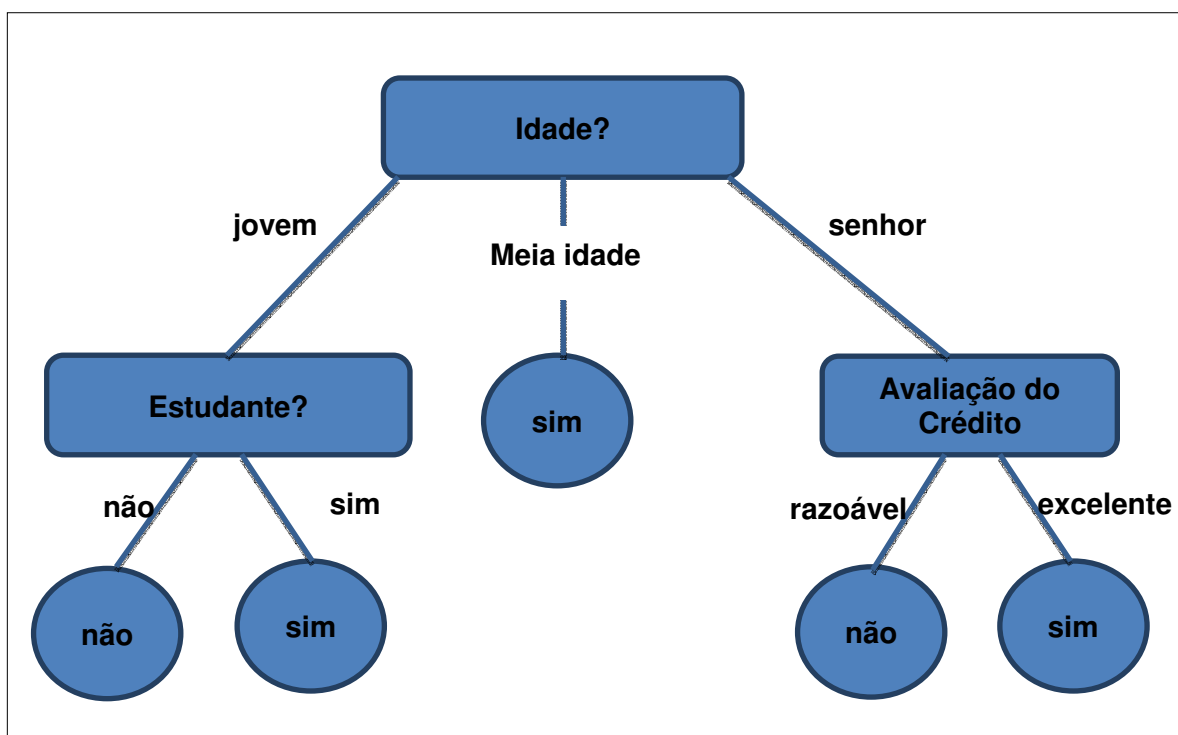


Figura 12. Árvore de decisão para a compra de um computador, HAN, *et al.* (2006).

Na figura acima pode-se observar que: jovens estudantes, pessoas de meia idade e senhores com crédito avaliado como excelente são clientes potenciais para a compra de computadores. Todavia, jovens que não estudam e senhores com crédito avaliado como razoável não são compradores de computadores em potencial.

Pela similaridade das árvores de decisão com árvores genealógicas, níveis hierárquicos ou até mesmo fluxos, sua assimilação por humanos é geralmente intuitiva e de fácil compreensão. Portanto, as etapas para classificação e predição de decisão em árvores de decisão costumam ser simples e rápidas e, de uma forma geral, resultam em uma boa precisão em seus resultados.

No que se refere ao algoritmo J48, trata-se de uma versão do tradicional algoritmo C4.5, sendo utilizado na técnica de árvores de decisão. Sua utilização é responsável pela geração de modelos de árvore de decisão baseando-se em um grupo de dados (RAMISCH, 2012).

A partir deste modelo, são realizadas a classificação e a análise por meio da estatística *Kappa*, usada em escalas nominais e que fornece uma ideia do quanto as observações se afastam daquelas esperadas, apontando a legitimidade das interpretações (PINTO, 2012).

Sua classificação é apresentada como: fraca, regular, moderada, boa ou excelente. Caso o resultado esteja entre fraco e moderado, os dados necessitam de ajustes.

Para obter melhores resultados, o J48 possibilita aplicar o uso de podas na árvore, número mínimo de instâncias por folha e também a construção de árvores binárias.

Martinhago (2005) aplicou o J48 em uma base de dados contendo informações dos candidatos inscritos no vestibular da Universidade Federal do Paraná (UFPR), dos onze cursos mais concorridos, e obteve resultados como:

- A nota da Redação tem grande influência na classificação do candidato;
- As notas de Matemática e Química, somadas às de Língua Portuguesa e Redação, influenciam na aprovação do candidato;
- Candidatos que tiraram notas acima de 42 pontos no ENEM indicam possível aprovação;

- Ter ou não feito cursos pré-vestibulares não exerce grande influência no resultado do concurso de vestibular.

3.3.3. Tarefa de Regressão/Estimativa

Conceitualmente, a tarefa de regressão é bem próxima à de classificação, tendo como diferença apenas o atributo a ser predito. Enquanto a tarefa de classificações é discreta, em regressão o predito é contínuo. Dias (2001) define a regressão como uma busca por funções, lineares ou não, que mapeiem os registros de um banco de dados em valores reais. Fayyad, *et al.* (1996) define a regressão como aprender uma função que mapeia um item de dado para uma variável de predição real estimada.

Com a aplicação da tarefa de regressão, pode-se estimar a quantidade de filhos de uma família; a probabilidade de um paciente sobreviver, baseando-se em um conjunto de diagnósticos de exames realizados; prever a demanda de um consumidor para um lançamento de produto, entre outras.

3.3.4. Tarefa de Sumarização

O conceito da tarefa de sumarização é identificar e apresentar, de maneira precisa e clara, as principais características dos dados compreendidos em um conjunto de dados. Uma aplicação de sumarização seria a tabulação dos resultados e desvios. Um exemplo de aplicação dado por Boente (2006) é a partir de um banco de dados contendo informações de clientes que assinam uma revista semanal, buscar características que sejam similares entre os clientes.

3.3.5. Tarefa de Agrupamento (*Clustering*)

Segundo Jain and Dubes (1988), agrupamento é a tarefa descritiva onde se procura identificar um conjunto finito de categorias ou 'clusters' para descrever uma informação. Ao contrário das tarefas de classificação que são

aprendizados supervisionados, onde o treinamento é formado de tuplas classificadas, o agrupamento é uma tarefa de aprendizado não supervisionado visto que os clusters representam classes que não estão definidas no início do processo de aprendizagem.

O algoritmo K-means é definido por Tan *et al.*, (2009) como uma técnica de partição para agrupamento baseado em protótipos, que busca encontrar um determinado número definido pelo usuário de grupos (K), para serem representados pelos seus centroides. A figura 13 apresenta um exemplo de agrupamento considerado por Tan *et al.*, (2009) como um resultado excelente, já que itens idênticos estão agrupados entre si.

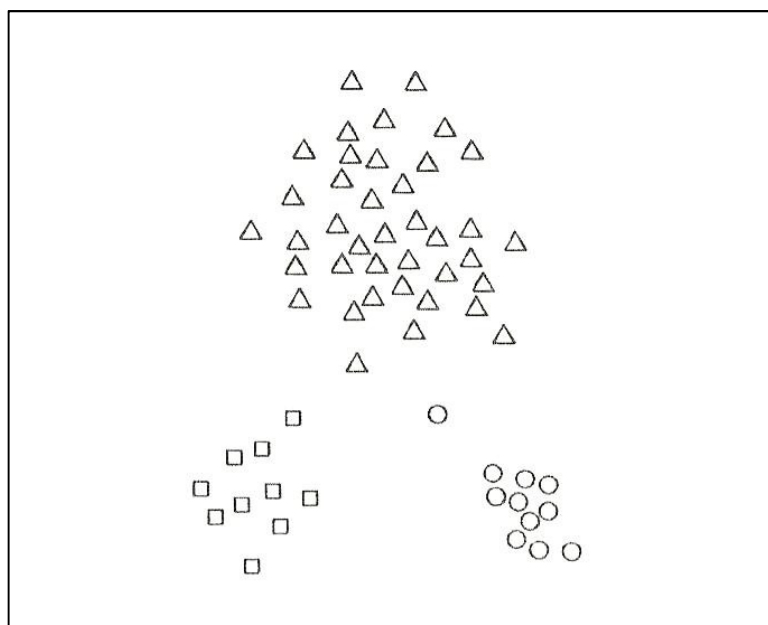


Figura 13. Exemplo de agrupamento em excelentes condições (TAN, *et al.* 2009).

Todavia, o algoritmo K-Means contém uma anomalia em utilizar grupos com formas não esféricas ou de tamanhos ou densidades muito diferentes TAN, *et al.* (2009) cita como exemplo a divisão em três grupos, sendo um muito maior do que os outros dois como consequência o grupo maior será dividido enquanto os dois outros grupos de menor tamanho serão combinados como parte do grupo maior. A figura 14 apresenta um exemplo de grupos com tamanhos diferentes, sendo o lado esquerdo (quadrados) menor em comparação ao grupo central e direito (losangos e elipses).

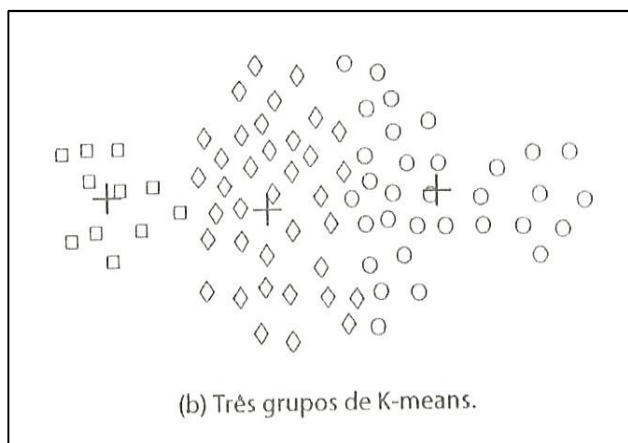


Figura 14. K-means com grupos de tamanhos diferentes (TAN, *et al.* 2009).

Apesar da dificuldade em executar grupos de tamanhos muito distintos, o algoritmo de agrupamento k-means é considerado simples e muito utilizado em varias áreas de conhecimento, além de ser eficiente na maioria dos bancos de dados que é aplicado. (TAN *et al.* 2009).

3.3.5.1. Estatística Kappa

A definição dada por Pinto (2012) é que a estatística Kappa é uma medida de concordância usada em escalas nominais que nos fornece uma base do quanto as observações se afastam daquelas esperadas, fruto do acaso, indicando-nos assim o quão legítimas as interpretações são.

A estatística Kappa é utilizada no algoritmo k-means por meio do software Weka, a fim de estabelecer o grau de concordância do agrupamento a ser gerado, tendo como valor máximo o '1'. Os valores abaixo de '0' indicam nenhuma concordância. Na tabela 5 apresentada por Landis (1977), são descritas as faixas dos resultados estatísticos Kappa esperados, contendo respectivamente o grau de importância.

Tabela 5. Tabela de valores da estatística Kappa, traduzido de Landis, 1977.

Valores de Kappa	Interpretação
<0	Nenhuma aceitação
0-0.19	Fraca aceitação
0.20-0.39	Pouca aceitação
0.40-0.59	Aceitação moderada
0.60-0.79	Aceitação substancial
0.80-1.00	Aceitação quase perfeita

4. SEGURANÇA E PRIVACIDADE

O SBTVD conta com diversos padrões já definidos pela ABNT. Um deles está definido na norma NBR 15605-1, que trata as questões de segurança, porém focada no controle e regras de cópia de conteúdo digital. Já a norma NBR 15605-2, ainda em aprovação, referencia o modelo de segurança de aplicações interativas para TV Digital, definindo que aplicações disponíveis via carrossel (Figura 15) sejam autenticadas através de assinaturas digitais adquiridas junto a órgãos reguladores, como Infraestrutura de Chaves Pública Brasileira (ICP-Brasil).

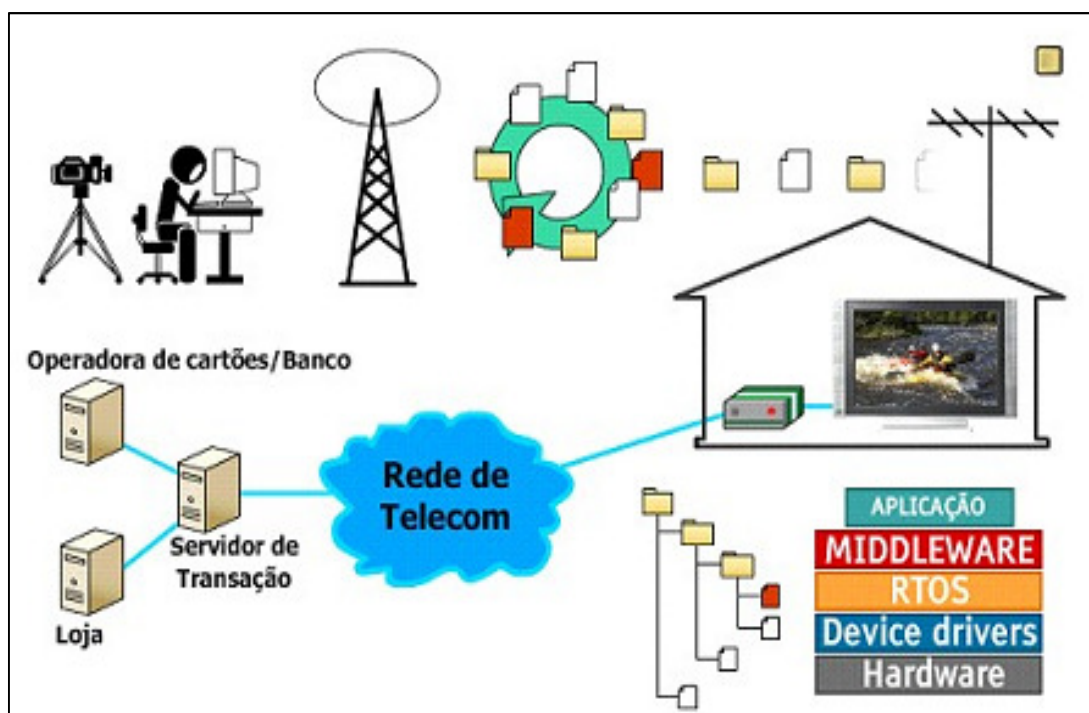


Figura 15. Exemplo de um ambiente de aplicações TCP/IP, Fonte: CPqD 2012.

A figura acima mostra o processo de uma compra online, feita por meio de um aplicativo da TV Digital. O processo parte da criação dos aplicativos interativos, seguindo-se de: transmissão do aplicativo por meio de uma antena através do carrossel de dados; captação do sinal nas residências; apresentação do aplicativo no aparelho de TV; acesso ao aplicativo pelo telespectador; envio dos dados capturados por meio do canal de retorno; recebimento da transação

por parte da loja responsável pela interação; e, por fim, a aprovação do crédito por parte da operadora do cartão/banco.

O processo de comunicação do canal de retorno da TV Digital é apresentado na NBR 15607, dividida em cinco fases: Conexão de Linha, Conexão de Enlace, Transferência de Dados, Desconexão de Enlace e Desconexão de Linha; ela também define o STV-B21 como protocolo de coleta de informações de telespectador, tornando a conexão entre os receptores fixo/portáteis e a rede de coleta possível. A conexão é feita via protocolo de rede TCP/IP, possibilitando que os servidores de coleta possam se encontrar em qualquer local com acesso à internet.

Uma vez conectado à internet, roubo de dados, transações fraudulentas e violação de dados podem ocorrer com o usuário final. A evolução desses ataques acompanha o crescimento da demanda de tais tecnologias. Na figura 16, é apresentado um paralelo entre a evolução das técnicas de segurança e as tecnologias existentes (CPQD, 2012)

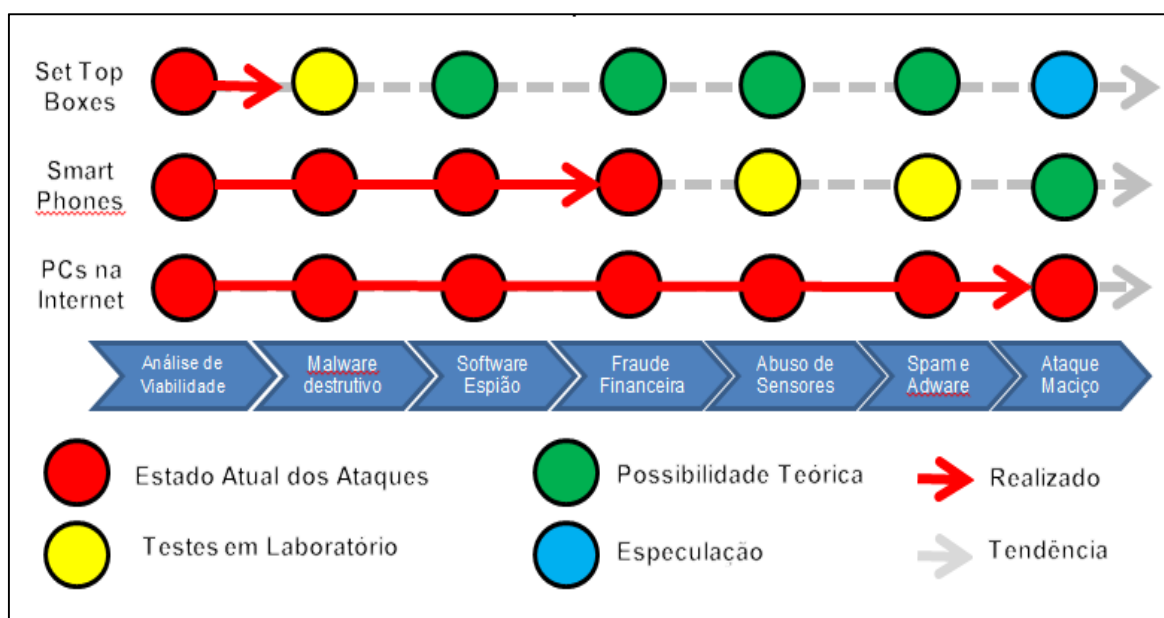


Figura 16. Tecnologias e evolução na segurança, adaptado de CPqD 2012.

A figura mostra que os computadores conectados à internet estão suscetíveis a todos os tipos de ataques, porém as soluções em segurança podem

ser encontradas em todos os níveis. Segundo Brentano (2012), 17% dos televisores no Brasil já podem realizar conexão com a internet. Nos EUA, esse número chega a 38%. Todavia, o ritmo de crescimento não está refletindo na evolução das técnicas de segurança, bem como no merecido interesse em definir normas para tal, como a NBR 15605-2.

Casos recentes como o ocorrido na rede *Playstation Network* (PSN) da empresa Sony, onde mais de 100 milhões de usuários de todo o mundo tiveram dados como nome completo, endereço e dados do cartão de crédito roubados (G1, 2012), deixam claro que não apenas computadores, mas quaisquer dispositivos conectados à internet necessitam de padrões definidos e técnicas de segurança eficazes.

4.1. Legislação e Aspectos Legais

Assim como a internet, a TV Digital Interativa é um meio de publicidade que está conquistando maior espaço a cada dia, e sistemas de recomendação são cada vez mais utilizados. Segundo a pesquisa da Bling Media, de fevereiro de 2011, os sistemas de recomendação potencializam as taxas de cliques, elevando-as em até 7,5 vezes se comparadas aos anúncios tradicionais (WEBINAR, 2012). Mas qual o limite para obtenção desses dados? Ao contrário do que alguns autores como (SIMÕES 2007), (JUNIOR, 2012) e (EXPM, 2012), dizem sobre não haver invasão de privacidade pelo fato de que os usuários deixam rastros ao navegar por conteúdos digitais, a coleta e utilização desses rastros, seja na web ou nas TVDI, somente devem ser manipuladas com o consentimento do usuário, visto que isso transcende à possibilidade de eficiência da comunicação e entra na questão do direito fundamental à privacidade e à intimidade.

Em países/blocos como a União Europeia, a Comissão Europeia apresentou uma lei específica aos Estados membros, que poderá entrar em vigor num prazo máximo de dois anos, reforçando as leis de proteção de dados de 1995 já existentes. Nestas novas regras, destacam-se itens como:

- Todas as organizações deverão notificar no prazo de 24 horas à autoridade nacional de supervisão e aos cidadãos afetados sobre violações de proteção de dados;
- Nomear um responsável de proteção de dados para cada empresa;
- As organizações devem explicar suas políticas de privacidade em linguagem clara, simples e acessível a todos os cidadãos. Estes devem conseguir perceber como os seus dados são utilizados e poder dar o seu consentimento explícito às organizações sobre como a sua informação deverá ser utilizada;
- Deverão, ainda, ter o direito de eliminar os seus dados e movê-los de um fornecedor para outro sem restrições.

Já nos EUA, um projeto de lei que está em tramitação no congresso pretende garantir o direito individual do consumidor, controlando qual tipo de informação poderá ser coletada e exigindo transparência quanto ao uso da informação obtida.

Na legislação brasileira, não há leis de privacidade online muito claras, porém com a unificação dos termos de compromisso do Google, que entrou em vigor em 01 de Março de 2012, uma audiência pública na comissão de defesa do consumidor da Câmara dos Deputados dará início a essa discussão – ainda sem data definida.

Segundo Varella (2012), no Brasil há uma lacuna normativa e regulatória com relação à proteção de dados: não existe o estabelecimento de princípios, normas e responsabilidades, tampouco uma autoridade competente para fiscalizá-lo e evitar abusos.

Aos usuários, resta a dependência das políticas de privacidade das empresas, ora inexistentes, ora em completo descompasso com outros diplomas legais, como o Código de Defesa do Consumidor.

4.2. Considerações

Uma pesquisa intitulada *Winning over the Empowered Consumers* (IBM, 2012) realizada pela IBM com 28 mil usuários em 15 países, inclusive 1,8 mil brasileiros mostra que os usuários estão mais dispostos a compartilhar informações pessoais a seus grupos favoritos de varejo, buscando uma experiência mais personalizada. No Brasil, 55% dos usuários estariam dispostos a fornecer dados demográficos e 41% divulgariam informações sobre estilo de vida e preferências (IBM, 2012). No entanto, o censo americano mostra que 75% dos usuários daquele país se preocupam com a possível divulgação de dados que fornecem às empresas (TORRES, 2004).

Como no Brasil não há uma legislação em vigor nesta área, a proposta desse trabalho é instigar a comunidade brasileira a refletir sobre a questão, bem como propor que os desenvolvedores de aplicativos que optam por utilizar sistemas de recomendação para TV Digital interativa sigam algumas recomendações básicas, como:

- Obrigatoriedade na utilização dos termos de uso, utilizando uma linguagem clara, simples e objetiva, podendo o usuário aceitá-la completamente ou em partes, por exemplo, liberando seus dados para recomendação e propagandas, mas recusando mailings ou venda de tais dados;
- Na descrição dos termos de uso, informar que os dados serão utilizados para recomendação e não serão vendidos, caso contrário, solicitar a autorização para tal procedimento;
- A qualquer momento, o telespectador poderá, de uma maneira simples e objetiva, cancelar a autorização de captura de seus dados;
- O controle deve estar sempre nas mãos do usuário, sendo possível a ele estabelecer o quanto e com quem irá querer compartilhar, de acordo com seu desejo e/ou necessidade;
- O usuário tem o direito de saber que tipo de informação é coletado, e também de que forma (quais ferramentas) são capturados esses dados do serviço no momento em que lhes for conveniente.

No que se refere ao desenvolvimento de aplicações interativas para TV Digital Interativa, sejam elas para fins de transações e-commerce, portais de notícias, jogos etc., pontos importantes devem ser considerados pelos desenvolvedores de aplicativos e difusoras:

- Utilização de Certificados Digitais em transações que envolvam comércio eletrônico, além da utilização de canais seguros através de protocolos como Secure Sockets Layer (SSL), Internet Protocol Security (IPSec), etc.;
- Aplicação de Assinaturas Digitais e verificação de autenticidade tanto em e-commerce quanto disponibilização de aplicativos;
- A exclusão de aplicativos interativos feita com segurança e eliminação total do mesmo sem que haja resíduos ou exposição de dados privativos;
- Segurança na retenção de dados bancários para aplicações e-commerce, passando por vistorias e monitoramento a todo tempo, a fim de evitar possíveis vazamentos;
- Para aplicações e-commerce, a utilização de normas como ISO/IEC 27001 e a *Payment Card industry (PCI)* são totalmente aplicáveis.

Com isso, acredita-se que os aplicativos para a TV Digital interativa se tornarão ainda mais atraentes, seguros e claros a seus usuários, aumentando a relação de satisfação tanto de quem os utiliza quanto de quem anuncia.

5. METODOLOGIA

Dentro do cenário da TV digital Interativa até aqui apresentado, é possível perceber que se faz necessário um aplicativo que ofereça a integração entre tecnologias de forma independente e flexível, além de realizar a captura de dados da forma implícita e explícita, armazenando e tratando esses dados para a apresentação de publicidade.

Assim se propõe, para atender o objetivo deste trabalho, o aplicativo denominado *Smart Marketing*, que tem como proposta a apresentação de notícias e recomendação de publicidades personalizadas baseadas na navegação do telespectador pelo aplicativo.

Para isso, este trabalho utilizou a pesquisa descritiva, apresentando uma nova visão sobre a utilização de aplicativos pelas difusoras e pelo mercado publicitário, propondo uma forma viável, além da possibilidade de torná-la rentável no que diz respeito ao cenário atual.

5.1. Descrição de Cenários de Uso

Para ilustrar a aplicabilidade do *Smart Marketing*, e ajudar na compreensão deste trabalho, é apresentada a descrição de possíveis cenários de uso por meio de casos.

(i) *"Luiz, um trabalhador de 50 anos de idade, casado, com filhos, veículo próprio, classificado como pessoa da classe econômica C, acessa por meio da TV Digital Interativa o aplicativo de Notícias, e as notícias que mais acessa são relacionadas a veículos. Após alguns acessos, Luiz começa a visualizar publicidades de forma personalizada, um veículo popular zero quilômetro, que provavelmente ele gostará de visualizar."*

(ii) *William, um jovem de 23 anos sem filhos, solteiro e classificado como pessoa da classe B, acessa por meio da TV Digital Interativa o aplicativo de Notícias, e as notícias que mais acessa são relacionadas a viagens e esportes. Após alguns acessos, Willian começa a visualizar publicidades de forma*

personalizada, como um pacote de viagens para esqui no Chile, pacote de viagem com trilhas, rafting³ e acampamento, como também oferta de passagens aéreas e itens recomendados de acordo com seu perfil e interesse.

Por meio dos cenários de uso acima, é possível identificar algumas preferências do telespectador de forma implícita, como a categoria de notícia acessada; no caso de Luiz (ator do cenário de uso (i)), seus acessos são, em sua maioria, para veículos; já Willian (ator do cenário de uso (ii)), busca notícias da(s) categoria(s): viagem, esportes, esportes radicais. O tipo de publicidade a ser apresentado dependerá dos interesses da marca a ser mostrada, visto que também são obtidos dados do telespectador por meio direto através do formulário.

5.2. Requisitos Funcionais e Não Funcionais

Para a construção da solução proposta, a seguir são apresentados os Requisitos Funcionais (RF), classificados em dois tipos: o RF-DT refere-se a Requisito Funcional do dispositivo do telespectador e o RF-PC, Requisito Funcional do provedor de conteúdo, bem como os requisitos não funcionais.

5.2.1. Requisitos Funcionais

- RF-DT-1 - Prover um serviço responsável pela aquisição de forma implícita, armazenamento e acesso de informações da navegação do telespectador;
- RF-DT-2 - Prover um serviço responsável pela aquisição e manipulação do perfil, de acordo com as informações inseridas de forma explícita pelo telespectador;
- RF-DT-3 - Prover um serviço responsável pela integração e controle dos principais serviços do aplicativo proposto;

³ Rafting é um esporte que utiliza botes infláveis para a descida de rios (BUENO, 2012).

- RF-DT-4 - Criar um mecanismo que permita obter perfis dos telespectadores de forma explícita (formulário) e implícita (captura da navegação pelo aplicativo), com o intuito de gerar dinamicamente publicidades personalizadas;
- RF-PC-1 - Prover um serviço capaz de criar modelos de conhecimento a partir dos dados implícitos e explícitos adquiridos.
- RF-PC-2 - Prover um serviço de armazenagem de informações acessadas predizendo que o telespectador poderá ter interesse em forma de propaganda.

5.2.2. Requisitos Não Funcionais

- RNF-PC-1 - Definir uma arquitetura que ofereça componentes para suporte ao desenvolvimento e execução do sistema de recomendação, ocultando a complexidade do baixo nível da programação existente;
- RNF-PC-2 – O aplicativo proposto deve basear-se nas especificações de padrões de metadados voltados para a descrição e personalização de conteúdos publicitários de domínio da TV Digital como XML que estruturam os metadados;
- RNF-PC-3 – A solução proposta deverá ser independente do sistema específico para a mineração de dados, portanto é proposta a utilização de um banco de dados independente;
- RNF-PC-4 - Devido a limitações de hardware, a aplicação deverá permitir que o processamento da recomendação personalizada seja executado de forma distribuída, sendo 'p' dispositivo do usuário apenas o meio de captura e retorno de dados, e deixando o provedor de serviço responsável pelo maior processamento.

5.3. Diagrama de Casos de Uso

Após a definição dos requisitos funcionais e não funcionais, a seguir apresenta-se o diagrama geral de casos de uso (Figura 17), baseando-se na linguagem UML (*Unified Modeling Language*). Os atores identificados foram:

- Telespectador (usuário): simboliza a entidade que interage com o aplicativo para obter recomendação personalizada de propaganda.
- Administrador Provedor de Conteúdo: caracteriza a entidade que interage com o sistema como provedor de conteúdo.

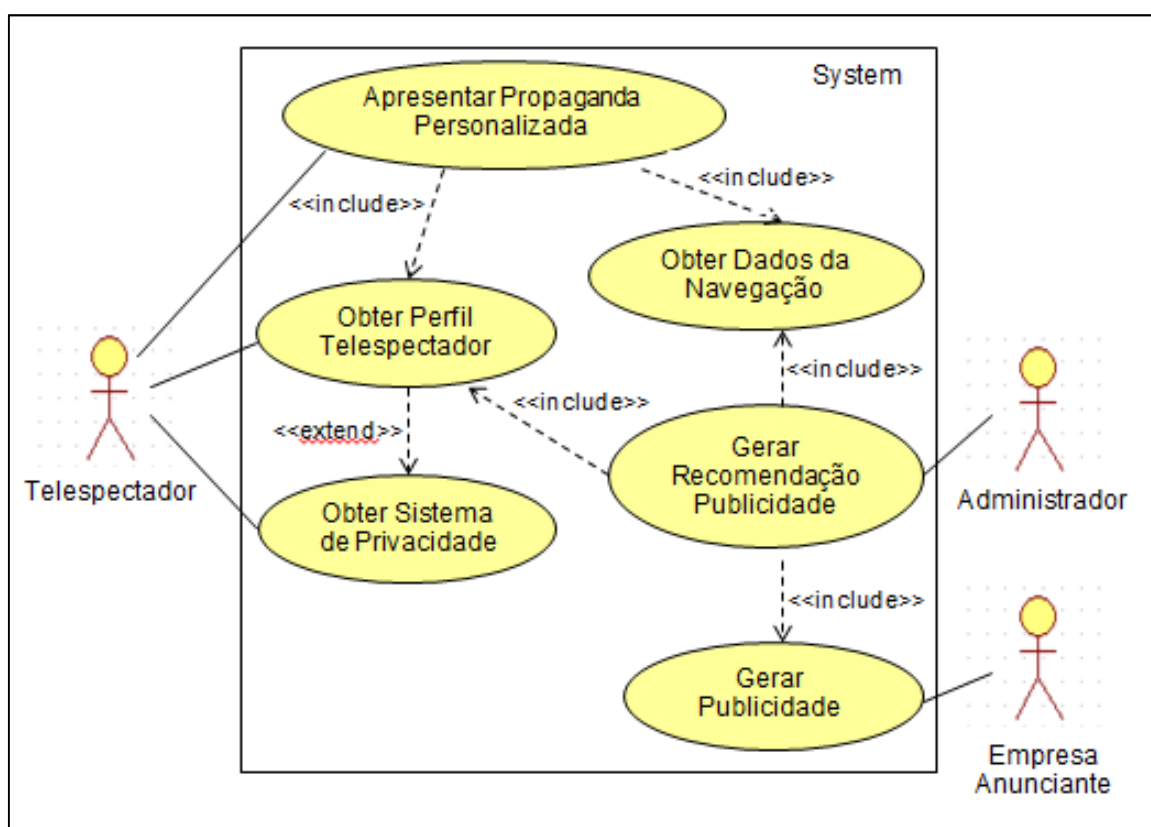


Figura 17. Diagrama de Caso de Uso.

De acordo com a figura apresentada acima, a seguir são descritos os seis caso de uso.

Caso de Uso: 01	Obter Perfil do Telespectador
Ator	Telespectador
Descrição	Por meio do sistema de recomendação, o telespectador especifica, de forma explícita, os seus dados pessoais: idade, sexo, se possui filhos e veículo, estado civil, faixa salarial e o tipo de residência que possui.
Evento iniciador	Acesso ao Smart Marketing na aba 2- Cadastro pelo telespectador.
Pré-condição	O Smart Marketing deve estar em execução.
Pós-condição	Informações do usuário armazenadas no banco de dados.
Extensões	Não há extensões
Inclusões	Não há inclusões

Caso de Uso: 02	Apresentar Publicidade Personalizada
Ator	Telespectador
Descrição	É apresentada ao telespectador, baseando-se na captura de dados implícita e explícita, uma peça publicitária personalizada.
Evento iniciador	O telespectador acessa a aba 4- Recomendação.
Pré-condição	Inserção da publicidade no perfil do telespectador pelo administrador.
Pós-condição	Apresentação da publicidade.
Extensões	Não há extensões
Inclusões	Caso de Uso “Obter Perfil do Telespectador” e “Obter Dados da Navegação”.

Caso de Uso: 03	Obter Dados da Navegação
Ator	Telespectador
Descrição	Os ids das categorias de notícias acessadas pelo telespectador são registrados e associados ao mesmo.
Evento iniciador	O telespectador clica sobre a manchete da notícia que deseja ler.
Pré-condição	O telespectador deve ter feito login no Smart Marketing.
Pós-condição	Os dados são armazenados na base de dados do telespectador.
Extensões	Não há extensões
Inclusões	Não há inclusões

Caso de Uso: 04	Obter Sistema de Privacidade
Ator	Telespectador
Descrição	Permite ao telespectador visualizar perguntas frequentes sobre o Smart Marketing, bem como o que deve fazer caso não queira compartilhar seus dados.
Evento iniciador	O telespectador acessa o Smart Marketing na aba 5- Segurança e Privacidade
Pré-condição	O telespectador deve ter feito login no Smart Marketing.

Pós-condição	É armazenado no seu perfil o aceite total ou parcial das condições de uso.
Extensões	Caso de Uso “Obter Perfil do Telespectador”
Inclusões	Não há inclusões

Caso de Uso: 05	Gerar Recomendação da Publicidade
Ator	Administrador
Descrição	Permite ao administrador inserir a publicidade no perfil do usuário.
Evento iniciador	Análise das publicidades cadastradas.
Pré-condição	Análise dos dados dos telespectadores, capturados implícita e explicitamente.
Pós-condição	Publicidade pronta para ser acessada pelo telespectador.
Extensões	Não há extensões
Inclusões	Casos de Uso “Obter Perfil do Telespectador”, “Obter Dados da Navegação” e “Gerar Publicidade”.

Caso de Uso: 06	Gerar Publicidades
Ator	Empresa Anunciante
Descrição	Permite à empresa inserir a publicidade, descrevendo o público alvo ao qual a mesma se destina.
Evento iniciador	Acesso ao link de cadastro da publicidade.
Pré-condição	A empresa anunciante ter criado a publicidade.
Pós-condição	Armazenamento da publicidade no banco de dados da publicidade.
Extensões	Não há extensões
Inclusões	Não há inclusões

A modelagem do usuário mostra as classes de informações do perfil do telespectador, que são necessárias para viabilizar o sistema de recomendação proposto. A seguir é apresentada a figura 18, contendo o diagrama de classes na linguagem UML utilizado neste trabalho.

5.4. Modelagem de Usuário

A modelagem do usuário mostra as classes de informações do perfil do telespectador, que são necessárias para viabilizar o sistema de recomendação proposto. A seguir é apresentada a figura 18, contendo o diagrama de classes na linguagem UML utilizado neste trabalho.

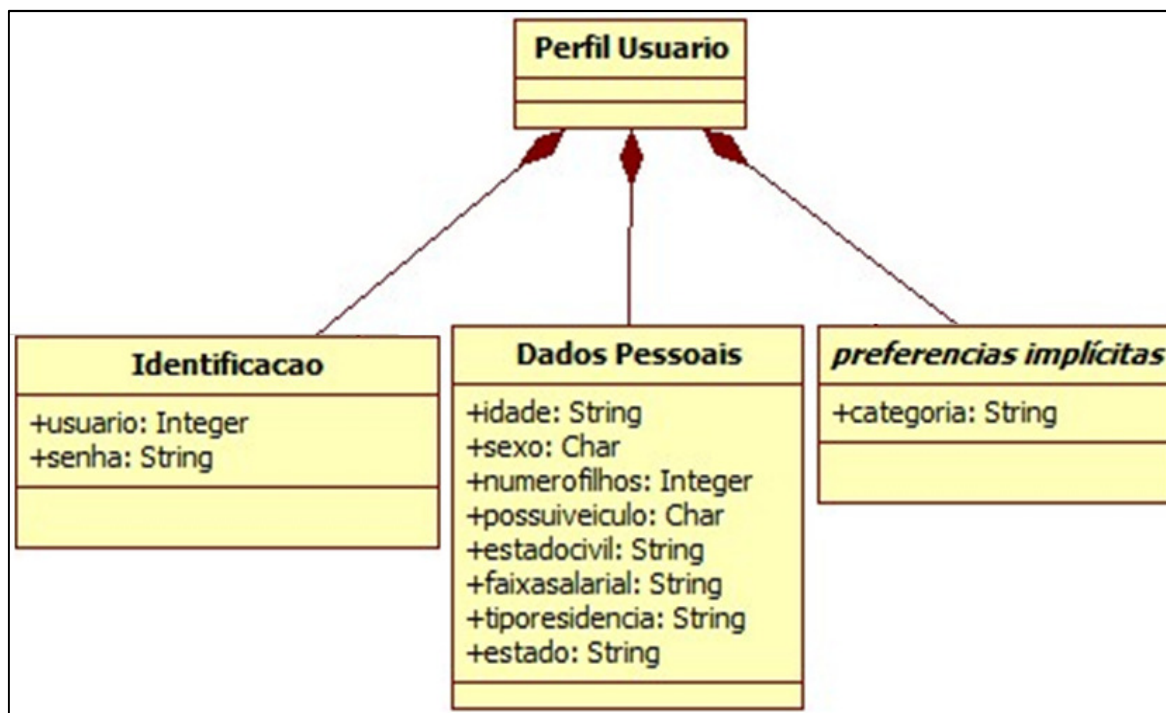


Figura 18. Modelo do Usuário utilizado no *Smart Marketing*.

A figura acima apresenta o modelo do usuário, contendo classes de informações que constituem o perfil do telespectador, tais como Identificação, Dados Pessoais (preferências explícitas) e Preferências Implícitas.

A identificação é formada por informações que permitem discernir um determinado perfil de telespectador em relação aos demais, e também permite o acesso ao sistema. Os dados pessoais, também chamados de captura explícita, apresentam as características dos telespectadores. Já as preferências implícitas se caracterizam por capturar os dados da navegação de maneira transparente.

5.5. Tecnologias e Ferramentas Utilizadas

O requisito não funcional RNF-PC-3 (seção 5.2.2), apresentado neste trabalho, diz que o *Smart Marketing* deve ser independente de plataforma de software. Portanto, em sua implementação utilizou-se apenas de tecnologias de código aberto sob a licença *Creative Commons* (CC, 2012).

Para torná-lo aderente às diversas tecnologias disponíveis, o projeto *Smart Marketing* foi concebido da forma modularizada, flexível e adepto do

conceito *Cloud Computing* (Computação na Nuvem). Os dados encontram-se armazenados na web, podendo ser utilizado o banco de dados que for mais conveniente.

5.5.1. Definição do Middleware para aplicações

Nesta seção, são apresentados os itens escolhidos para a escolha do *Middleware* para desenvolvimento de aplicações⁴ de TV Digital Interativa para ser utilizada no desenvolvimento do aplicativo *Smart Marketing*. Os itens definidos foram:

- Solução de negócio;
- Disponibilização do aplicativo pela web;
- Disponibilização do aplicativo ao telespectador pela TV;
- Fórum de apoio ao desenvolvimento;
- Manuais e apostilas para auxiliar no desenvolvimento;
- Interface gráfica padrão definida;
- Dispor de ambiente para testes;
- Possibilidade de utilização em aparelhos móveis.

A partir da aderência dos itens citados, pode-se considerar o quão maduro o projeto está, partindo do pressuposto que a plataforma de negócios esteja definida, contemplado uma infraestrutura de desenvolvimento de aplicativo, suporte ao desenvolvedor assim como disponibilização do aplicativo por meio da web e também pela própria TV Interativa ao telespectador em geral.

⁴ O middleware para desenvolvimento de aplicações consiste de máquinas de execução das linguagens oferecidas e bibliotecas de funções, permitindo o desenvolvimento rápido e fácil de aplicações para a TV Digital Interativa (GINGA, 2012).

Outro ponto relevante, definido como requisito mínimo, é a questão de suporte ao desenvolvimento, analisando a disponibilização de fórum de apoio ao desenvolvedor, manuais e apostilas. Também se tomou o cuidado para analisar se há ou não a definição de interface gráfica padrão e também um ambiente para testes. Também foram analisadas apenas ferramentas com licenças de uso sob licença GPL ou similares. Com isso, mantém-se a liberdade de personalização e uso do mesmo.

Os Middlewares para desenvolvimento de aplicações analisados foram: Ginga-NCL, Ginga-J, Projeto Serviços Multiplataforma de TV Interativa (PSMTVI) e Astro TV. Os Middlewares PSMTVI e AstroTV são implementações realizadas a partir da versão pura do Ginga. A tabela 6 apresenta os itens de descrição relacionados aos Middlewares analisados.

Tabela 6. Comparação de Middlewares para o projeto Smart Marketing.

Descrição	Ginga NCL Puro	Ginga J Puro	PSMTVI	AstroTV
Solução de negócios Definida	Não	Não	Não	Sim
Portal web para disponibilização dos aplicativos	Sim	Sim	Não	Sim
Disponibilidade do aplicativo ao Telespectador pela TV	Não	Não	Não	Sim*
Fórum de Apoio ao Desenvolvimento.	Sim	Sim	Não	Sim
Manuais e Apostilas para auxiliar no desenvolvimento	Sim	Sim	Sim	Sim
Interface Gráfica Padrão Definida	Não	Não	Sim	Sim
Ambiente para testes	Sim	sim	Não	Sim
Funcionamento em aparelhos móveis	Sim	Sim	Sim	Sim
Licença de uso	Livre GPL	Livre GPL	Livre GPL	Livre GPL

* Compatível com o STB Visiontec VT7200E e com os televisores Toshiba (modelos 65WL800i3D, 55WL800i3D, 46WL800i3D). Fonte: Autor

Atendendo a todos os itens descritos na tabela 6, o AstroTV com auxílio da tecnologia SOAP (*Simple Object Access Protocol*), tornou-se a opção para utilização no projeto *Smart Marketing*.

O *middleware* AstroTV oferece aos desenvolvedores o portal Astro DevNet, que conta com kits de desenvolvimento, fóruns de discussão, documentação e tutoriais, além de testar o aplicativo submetido de forma gratuita pela equipe da mantenedora da solução AstroTV.

Após aprovado pelo portal astroDevNet, o aplicativo é disponibilizado no portal Stickercenter. A disponibilização na própria TV Digital do telespectador é realizada através do portal Stickershop, possibilitando o download de aplicativos na TV Digital Interativa do telespectador (ASTRODEVNET, 2012).

O *middleware* AstroTV utiliza como plataforma de emulação da TV Digital o Astrobox, customizado a partir da distribuição Linux Ubuntu 10.04 versão r81460 compatível com a especificação brasileira.

A obtenção das notícias é feita por meio de notícias via RSS (*Really Simple Syndication*) 2.0, sendo o mesmo um subconjunto XML.

5.5.2. Definição do Banco de Dados

O banco de dados a ser definido no projeto necessitava estar sob licença livre, ser compatível com o software *Weka* e se comunicar com o SOAP NCLua aqui já citados.

Para atender a estes requisitos, optou-se por utilizar o MySQL 5.1.46 por meio do PHPmyAdmin 3.5.2 e Apache 2.2.17. A figura 19 apresenta a estrutura de banco de dados utilizada no Smart Marketing.

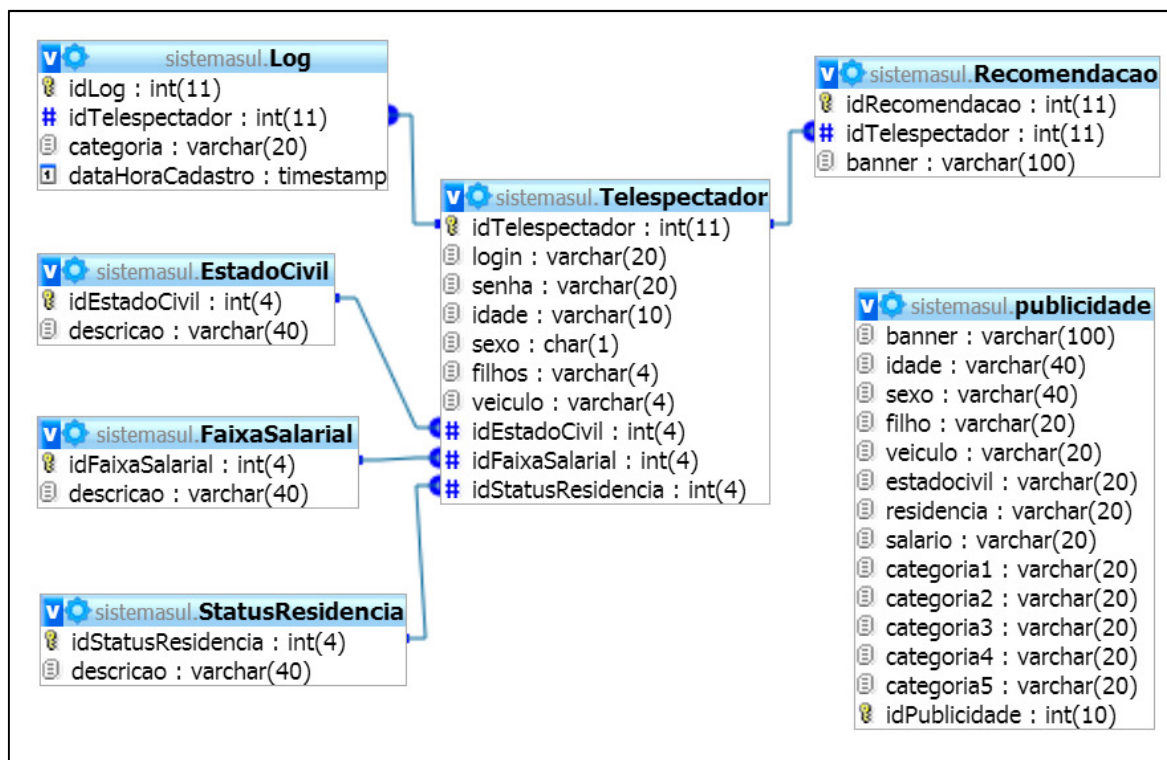


Figura 19. Diagrama do Banco de Dados do Smart Marketing.

5.5.3. Definição do software de descoberta de conhecimento

A ferramenta adotada neste projeto para a descoberta de conhecimento deve contemplar alguns requisitos mínimos, sendo:

- Estar sob licença GPL (*General Public License*);
- Oferecida de forma Gratuita;
- Contar com material de apoio como livros, fóruns, artigos, etc.;
- Ser utilizado em pesquisas na área acadêmica;
- Realizar integração com os principais SGBD (Sistema de Gerenciamento de Banco de Dados) via web;
- Gerar gráficos e relatórios, como por exemplo: uma árvore de decisão em forma de imagem;

- Possibilitar a inserção de algoritmos próprios;

Existem inúmeras ferramentas para descoberta de conhecimento, em sua maioria pagas, como Enterprise Miner (SAS, 2012), KXEN (KXEN, 2012), Knime, (KNIME, 2012), entre outras. Todavia, softwares de código aberto (*Open Source*) e gratuitos também são encontrados como: IlliMine (IlliMine, 2012) e Pimiento (ERABAKI, 2012). A tabela 7 apresenta os itens de descrição relacionados aos softwares de descoberta de conhecimento analisados.

Tabela 7. Comparação de softwares para descoberta de conhecimento.

Descrição	Enterprise Miner	Kxen	Knime	Pimiento	IlliMine	Weka
Estar sob licença GPL;	Não	Não	Sim	Não	Sim	Sim
Oferecida de forma Gratuita;	Não	Sim	Sim	Sim*	Sim	Sim
Contar com material de apoio como livros, fóruns, artigos, etc.;	Sim	Sim	Sim	Não	Não	Sim
Ser utilizado em pesquisas na área acadêmica;	Não	Não	Não	Sim	Sim	Sim
Realizar integração com os principais SGBD via web;	Sim	Sim	Sim	Não	Não	Sim
Gerar gráficos e relatórios	Sim	Sim	Sim	Sim	Sim	Sim
Possibilitar a inserção de algoritmos próprios;	Sim	Não	Sim	Sim	Sim	Sim

* Oferecido de forma gratuita para pesquisas acadêmicas, pesquisa sem fins lucrativos ou instituição acadêmica.

Atendendo todos os requisitos mínimos citados, este projeto utilizou o pacote de software *Weka* versão 7.7.7 feito em linguagem Java, disponível no endereço eletrônico < <http://www.cs.waikato.ac.nz/ml/weka/>>.

5.6. Arquitetura do Aplicativo

A proposta de arquitetura deste trabalho foi desenvolvida com o objetivo de oferecer flexibilidade e (liberdade de escolha) para o administrador de

conteúdo e novas aplicações do gênero, além de sanar o problema de limitação de hardware existente nos STB atuais. Portanto, o processo de recomendação será realizado de forma distribuída, conforme apresentado na figura 20 a seguir.

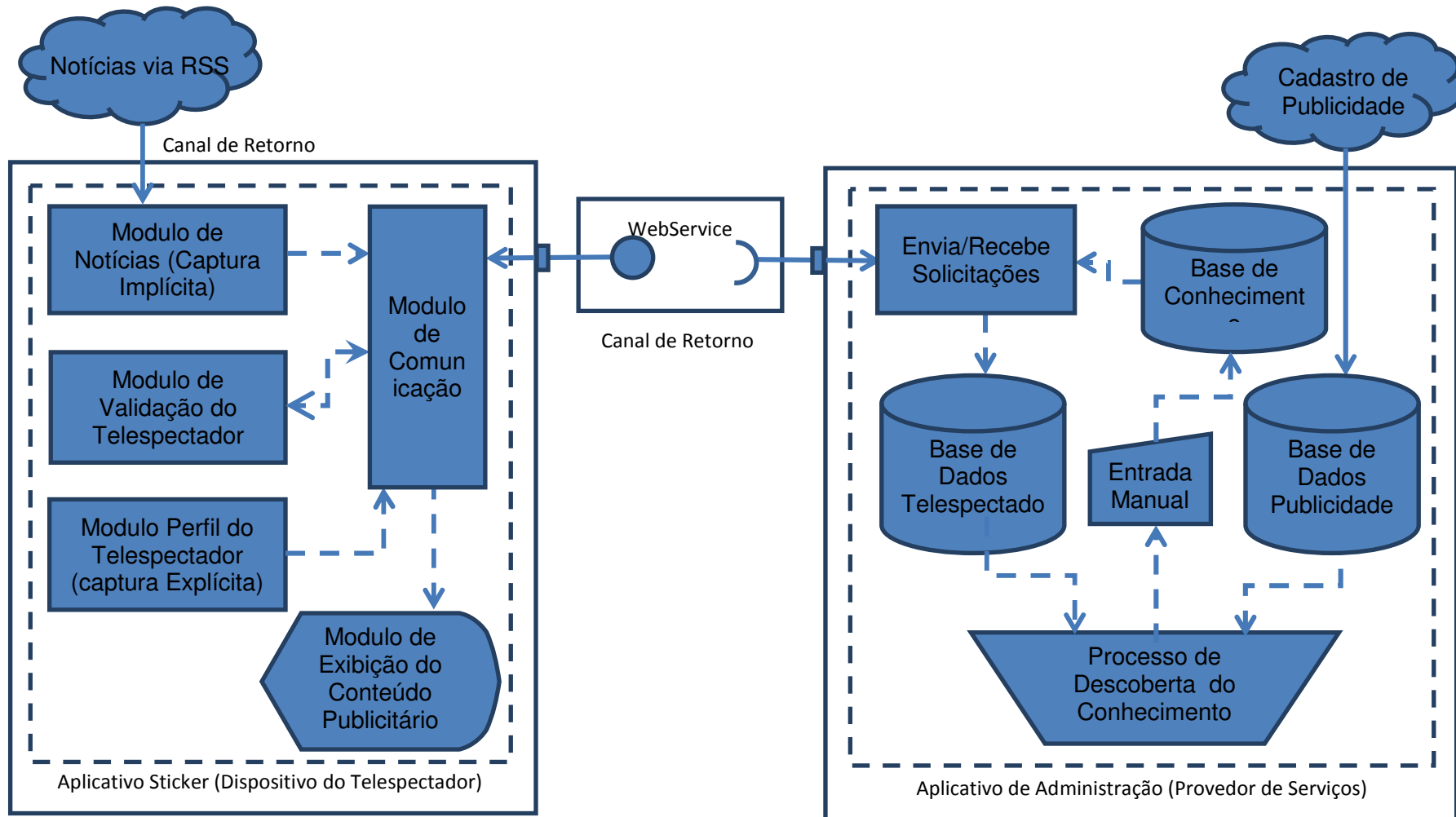


Figura 20. Arquitetura do Sistema Smart Marketing.

A figura 20 ilustra a arquitetura do sistema proposto dividido em duas estruturas: o dispositivo do telespectador e o provedor de serviços. O dispositivo do telespectador representa o lado de quem acessa o aplicativo na TV Digital Interativa, local em que se encontram os módulos do aplicativo responsáveis por exibir as notícias, propaganda e captura dos dados do telespectador. Já o provedor de serviços é independente do hardware da TV digital, sua localização é em um servidor de banco de dados web em MySQL, servidor o qual armazena os dados do telespectador, publicidade cadastrada via um website em PHP e a base de conhecimento.

O processo de descoberta do conhecimento é feito via desktop através de uma ferramenta de Mineração de Dados, sendo que, neste trabalho, optou-se por utilizar o Weka, conforme citado anteriormente.

5.6.1. Módulo Perfil do Telespectador (Captura Explícita)

Localizado no dispositivo do telespectador, o módulo 'perfil do telespectador' gerencia o acesso e aquisição de forma explícita das informações que compõem o perfil.

Por meio de seu formulário, o telespectador poderá criar seu login e senha para acesso ao aplicativo, além de inserir dados pessoais como: idade, sexo, se tem filhos, veículo, seu estado civil, faixa salarial e, por fim, o status da residência, conforme apresentado na figura 21, sendo todos os campos de preenchimento obrigatório.

PUC-Campinas

1 Login

2 Cadastro

Login

Senha

Idade

Sexo < Masculino >

Filhos < Sim >

Veiculo < Sim >

Estado Civil < Solteiro >

Faixa Salarial < Menor que 1 >

Status Residencia < Propria >

Salvar

3 Noticias

4 Recomendacao

5 Seguranca e Privacidade

X ?

Figura 21. Tela de cadastro Smart Marketing.

O campo de preenchimento login pode ser composto por até 20 caracteres entre números e letras e outros caracteres. Ao pressionar o botão ENTER no controle remoto, é apresentado o teclado virtual (Figura 22), facilitando a acessibilidade do telespectador.

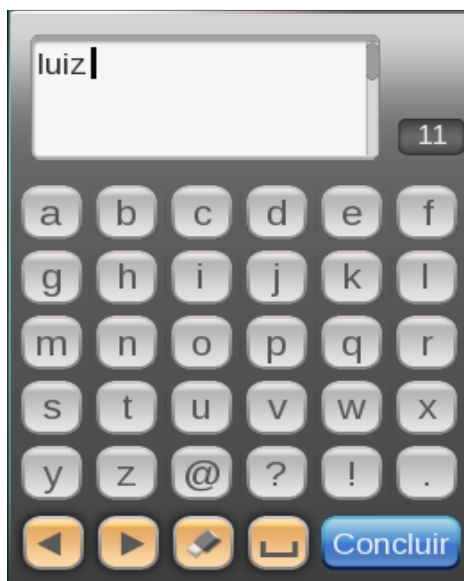


Figura 22. Teclado apresentado no campo login quando pressionador ENTER.

O campo senha deve ser composto por até 6 números, inseridos a partir do teclado existente no controle remoto. Para o preenchimento dos campos: Sexo, Filhos, Veículo, Estado Civil, Faixa Salarial, Status Residência é necessária apenas a utilização das setas para o lado direito ou esquerdo, pois se trata de campos pré-definidos.

Após o preenchimento dos campos, deve-se selecionar o botão 'salvar'. Com isso, os dados serão enviados via SOAP e gravados na base de dados. A figura 23 apresenta o log do envio dos dados inseridos no aplicativo Smart Marketing.

```

                                Telespectador.login soap response:
idStatusResidencia      2
idEstadoCivil    1
veiculo 1
senha  333
filhos 2
sexo   M
idTelespectador 49
idade  20
idFaixaSalarial 3
login  c

```

Figura 23. Trecho do Perfil do Telespectador enviado por NCLua SOAP.

O log gerado após o cadastro do telespectador no aplicativo *Smart Marketing* é composto por:

- IdStatusResidencia, composto por: 1=Própria, 2=Alugada e 3=Financiada;
- IdEstadoCivil, filhos e veículos podem ser: 1=Sim e 2=Não;
- IdTelespectador trata-se de uma chave primaria única gerada de forma sequencial;
- IdFaixaSalarial pode ser: 1= Classe E, 2=Classe D, 3=Classe C, 4=Classe B e 5=Classe A;
- E, para finalizar o log apresenta o login cadastrado.

5.6.2. Módulo de Validação do Telespectador

Os dados login e senha são validados diretamente no banco de dados por meio do Modulo de comunicação e SOAP Web Service de integração. Depois de validados, toda a navegação pelo modulo de notícias será capturada. A figura 24 apresenta a tela de validação do telespectador.



Figura 24. Tela de validação do telespectador.

Da mesma forma que é apresentado o teclado virtual para o preenchimento do perfil do telespectador, ele também é exibido ao pressionar ENTER no campo login. Já o campo senha, por questões de segurança é preenchido utilizando o teclado do controle remoto composto por números, sendo os mesmo exibidos por meio de uma máscara composta por '*', conforme a figura 25.

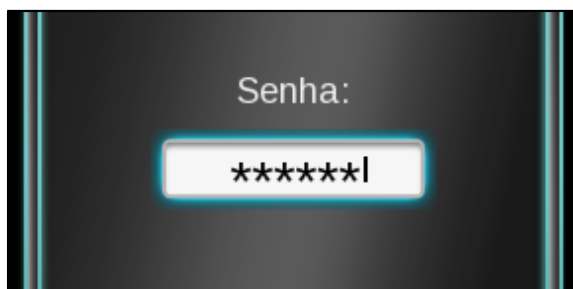


Figura 25. Campo senha ocultam os números digitados por meio de uma máscara.

5.6.3. Módulo de Notícias (Captura Implícita)

O módulo de notícias coordena o acesso, aquisição e armazenamento das notícias acessadas através de logs pela via implícita de forma automática. O aplicativo *Smart Marketing* apresenta as notícias em tempo real via canal RSS. A figura 26 mostra a estrutura de exibição do aplicativo, ao clicar em um título de uma notícia, é aberta uma extensão da janela (*popup*) no lado direito.

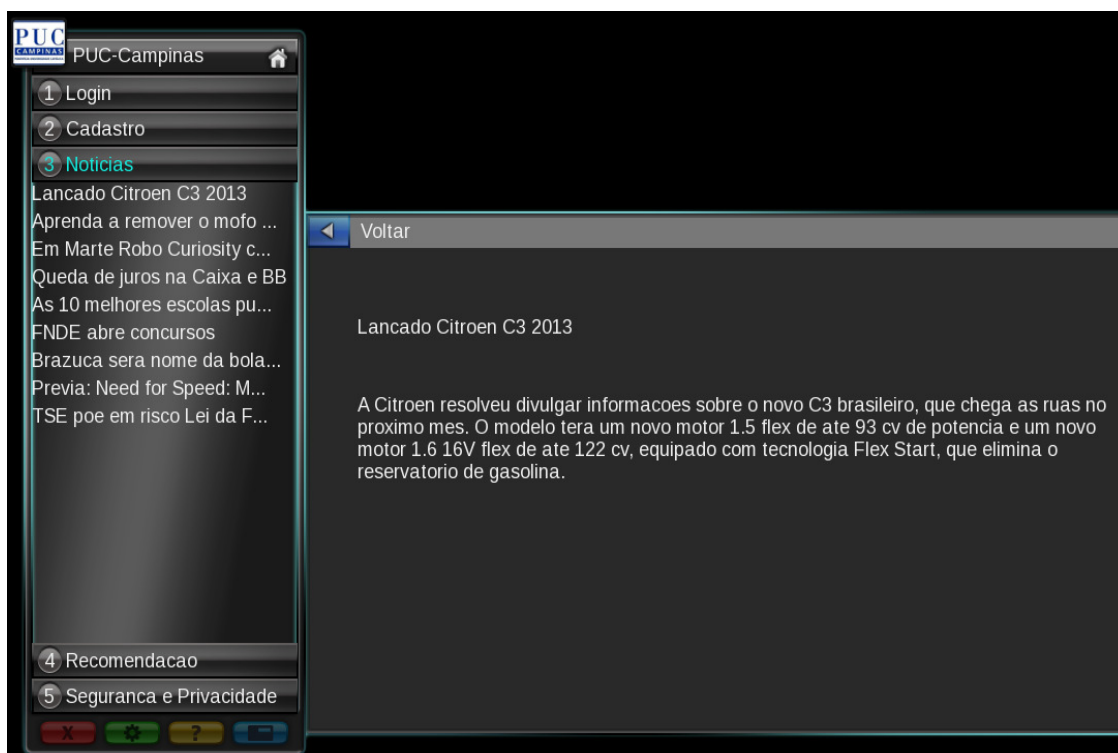


Figura 26. Apresentação da tela de notícias no Smart Marketing.

Com o clique na notícia, também é acionada a captura por meio implícito. Conforme apresentado na seção 3.1.1, a forma de captura é feita através da aquisição do campo *category* da estrutura XML, conforme exibido em um trecho de notícias em XML (Figura 27).

The image shows a snippet of XML data representing news items. Three red boxes highlight specific parts of the XML: the first box highlights the description of the first item, the second box highlights the category of the second item, and the third box highlights the description of the third item. A red callout box on the right contains the text "Dados capturados e armazenados no perfil do telespectador". Red arrows point from the highlighted XML fields to this callout box.

```

do Instituto de Metabolismo e Nutricao de Sao Paulo. </description>
<category>Saude</category>
<pubDate>Sat, 01 Sep 2012 21:40:12 -0300</pubDate>
</item>
- <item>
<title>Oi testa 4G e alcanca 88 Mbps de velocidade</title>
<link>http://idgnow.uol.com.br/internet/2012/08/31/em-testes-da-oi-rede-4g-alcanc
<description> A Oi concluiu esta semana degustacao da tecnologia 4G. A demonstracao f
telecomunicacoes, que aconteceu nos dias 29 e 30 de agosto, em Brasilia. A rede mo
velocidade de 88 Mbps com rajadas de ate 100 Mbps, realizados co
de 100 pessoas, entre representantes do M
participantes do evento, puderam experim
conexao a internet por meio de 4G. </descri
<category>Tecnologia</category>
<pubDate>Sat, 01 Sep 2012 21:38:43 -0300</pubDate>
</item>
- <item>
<title>Conheca as Cataratas do Iguacu oferece atividades para todas as idades</title>
<link>http://viagem.uol.com.br/guia/cidade/foz-do-iguacu.jhtm</link>
<description> Brasil e Argentina compartilham uma paisagem de beleza intensa no oeste
medidas. As cataratas do rio Iguacu formam o maior conjunto de quedas d'agua da T
m. Elas estao la, se exibindo, despidoras, ha milhares de anos, mas somente nas
mirantes ousados, que permitem tocar e ser tocado pelo aguaceiro. Ouvir o estrondo
memoria registra em Iguacu nao sairao mais com agua e sabao. O Parque Nacional d
1.154.000 visitantes no ano anterior, ou cerca de 3.160 pessoas por dia o parque ab
estrangeiros, se concentra na atracao principal que a visao dos borbotoes de espuma
tambem a encontra em esportes radicais como tirolesa, rapel e rafting no rio Iguacu
aventura: banho de catarata. </description>
<category>Viagem</category>
<pubDate>Sat, 01 Sep 2012 21:36:44 -0300</pubDate>

```

Figura 27. Dados Capturados na estrutura de Notícias em XML

A categoria é capturada e vinculada ao telespectador de forma individual, após o mesmo ser validado por meio de login e senha, realizando o vínculo; na figura 28 é apresentado um exemplo de captura de categorias de um telespectador.

```

Usuario alan logado
Conectado a www.sistemasul.com.br pela porta 80
Log.insert soap response: idLog gerado= 187
Conectado a www.sistemasul.com.br pela porta 80
Log.insert soap response: idLog gerado= 188

```

Figura 28. Log Gerado após o acesso a uma notícia.

5.6.4. Módulo de Comunicação

O módulo de comunicação torna possível a comunicação entre o dispositivo do telespectador e o provedor de serviços, por meio do *Web Service*, que envia requisições via protocolo SOAP, apresentado por Filho, (2011). Este módulo foi desenvolvido na linguagem Lua, e permite o acesso ao *Web Service* a partir da aplicação da TV Digital Interativa.

5.6.5. Módulo de Exibição de Conteúdo (Recomendação)

Após o processo de descoberta de conhecimento e inserção na tabela Base de Conhecimento, o módulo de comunicação, de maneira cíclica, busca o banner (Figura 29) disponível na tabela Publicidade e o apresenta no módulo de exibição do conteúdo publicitário.



Figura 29. Exemplo de Retorno de publicidade⁵ no Smart Marketing

5.6.6. Segurança e Privacidade

O capítulo 4 apresentou questões relacionadas à privacidade e segurança dos dados. Neste quesito, o aplicativo *Smart Marketing* exhibe na quinta aba uma lista de perguntas e respostas, em formato de “Dúvidas frequentes”, sobre como é feita a captura dos dados, armazenamento, funcionamento do processo de recomendação, além de mostrar o que fazer caso o usuário não

⁵ A(s) publicidade(s) apresentada(s) ao longo deste projeto são veiculada(s) na mídia em geral, não sendo da criadas pelo autor deste projeto.

queira compartilhar seus dados. A figura 30 apresenta a tela segurança e privacidade com auxílio do recurso de uma barra de rolagem.

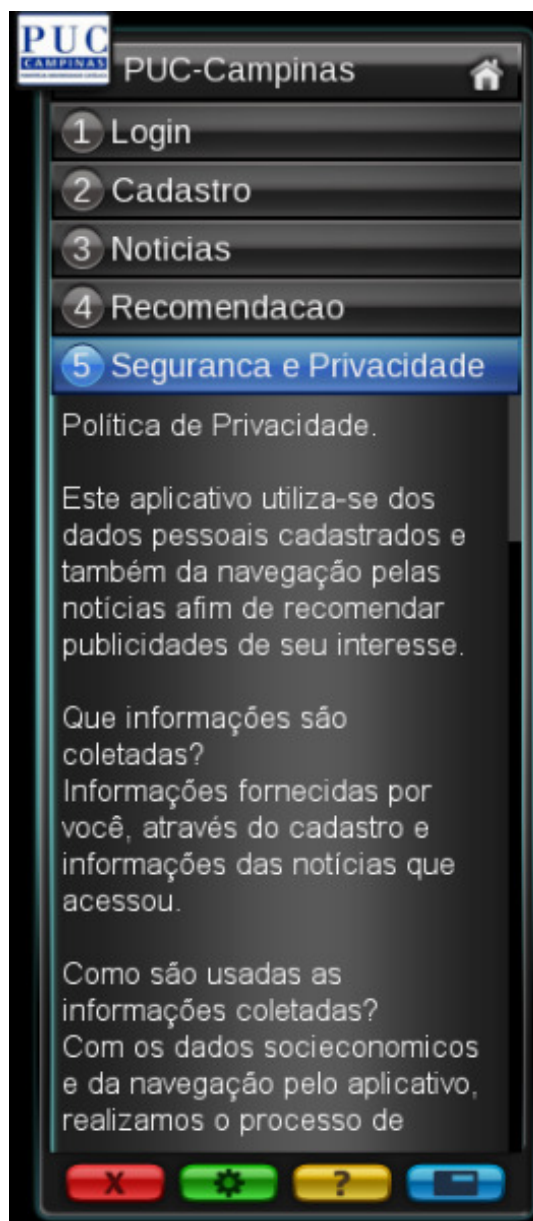


Figura 30. Tela Segurança e Privacidade

5.6.7. Direitos de Uso

O botão representado pela cor verde apresenta os direitos de uso do aplicativo, desenvolvedor etc.(Figura 31).



Figura 31. Tela Direitos de Uso

5.6.8. Ajuda (help)

O atual trabalho utiliza o *Middleware* da TOTVS⁶. Com isso, a padronização do aplicativo se faz necessária. Na documentação disponível no site da AstroTV, pode-se encontrar o meio de estruturação do aplicativo e as regras para a mesma (ASTROTV, 2012). O padrão definido optou pela utilização das cores dos botões do controle remoto do telespectador, sendo: vermelho – fechar o aplicativo; verde – acessar as configurações e direitos de uso; amarelo – abrir ou fechar a janela ajuda; e azul – alternar entre o vídeo em tela cheia e o aplicativo.

⁶ TOTVS, é empresa mantenedora do middleware para a TV Digital Brasileira AstroTV.

O módulo de Ajuda exibe também uma ilustração da forma que se realiza a navegação pelo controle remoto (Figura 32).

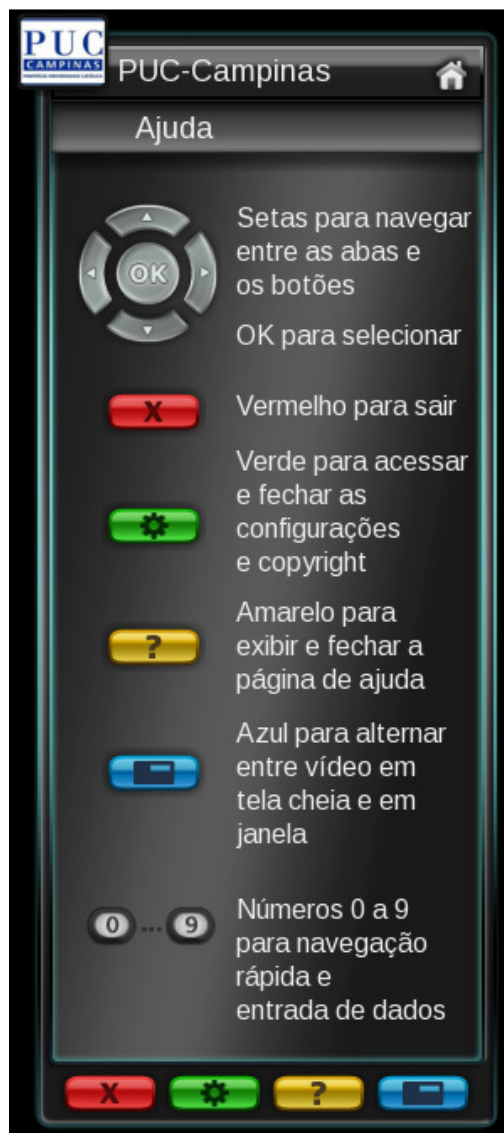


Figura 32. Estrutura de ajuda definida pela TOTVS.

5.6.9. Módulo Cadastro e Armazenamento da Publicidade

O cadastro e o armazenamento da publicidade são feitos via web, a partir do preenchimento de um formulário (Figura 33), sendo que a empresa anunciante realiza a inserção do perfil do público alvo de seu produto. Os campos obrigatórios solicitados são: estrutura etária, definição do sexo, se possui filhos,

veículo, estado civil, o tipo de residência do público alvo, a classe social à qual pertence.

Além das opções comuns de classificação, todos os campos possuem a opção indiferente. Além disso, devem ser cadastrados cinco categorias relacionadas à publicidade inserida, sendo no mínimo uma obrigatória.

CADASTRO PUBLICIDADE TV DIGITAL INTERATIVA - PUC-CAMPINAS

Idade:

Sexo:

Filhos:

Possui Veículo:

Estado Civil:

Tipo Residência:

Faixa Salarial:

Categoria 1:

Categoria 2:

Categoria 3:

Categoria 4:

Categoria 5:

Link Banner:

Figura 33. Formulário de cadastro da publicidade.

A publicidade cadastrada é armazenada na base de dados da publicidade, sendo utilizadas no relacionamento manual entre os interesses do telespectador as notícias disponíveis.

5.7. Processo de Recomendação da Publicidade Personalizada

Levando em consideração que o telespectador tenha preenchido o formulário contendo os dados de seu perfil, e que o sistema também já tenha sido treinado para a geração de modelos de conhecimento a partir da navegação pelo modulo de notícias, o processo de recomendação inicia-se a partir da aplicação do processo de descoberta de conhecimento. A entrada do resultado é feita de forma manual, na base de conhecimento na tabela recomendação.

A figura 34 apresenta um diagrama de atividades no processo de recomendação, iniciando-se pela conexão do telespectador por meio do login e senha; em seguida os dados da navegação pelo aplicativo de notícias são capturados e armazenados. Para a geração do perfil do telespectador, são necessários, além dos dados capturados de forma implícita, os dados pessoais já inseridos.

Com o perfil gerado, são feitas a predição das preferências do telespectador, a atualização do perfil com a publicidade definida e, por fim, a exibição da publicidade personalizada.

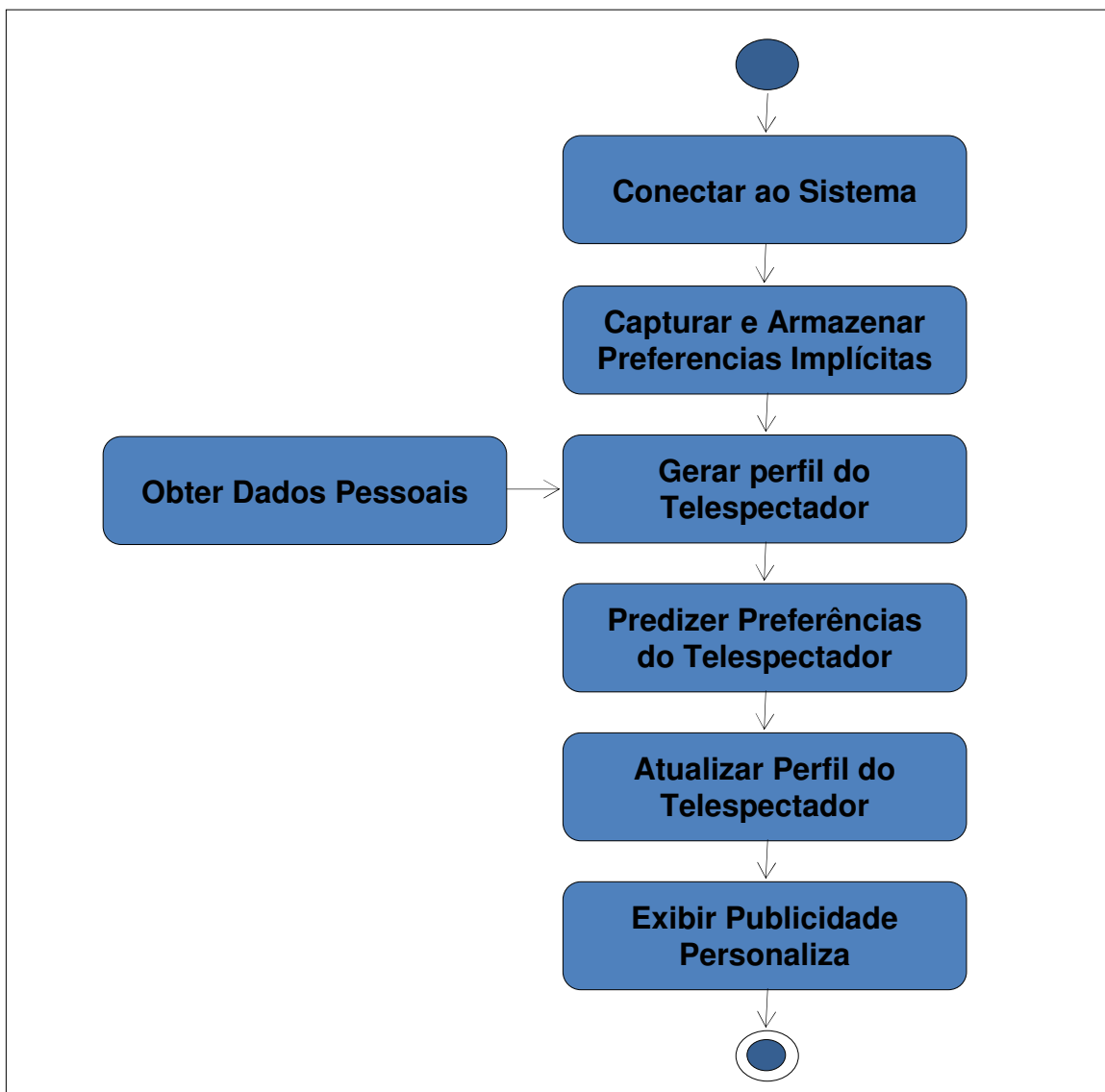


Figura 34: Diagrama das atividades no processo de recomendação

6. APLICAÇÃO DO SMART MARKETING – RESULTADOS E ANÁLISES

A realização do experimento, neste trabalho, visou alcançar os seguintes objetivos: analisar a qualidade do algoritmo aplicado e verificar a qualidade das recomendações a partir de um determinado modelo de recomendação. A seção 6.1 apresenta a definição dos dados utilizados no experimento, já a seção 6.2 mostra de forma prática a aplicação do experimento aos telespectadores, os quais serão chamados, a partir desse ponto, de usuários, visto que os mesmos participaram do experimento.

6.1. Definição dos Dados

A realização dos experimentos foi feita com base nos dois meios de captura de dados, o explícito e o implícito, conforme apresentado na seção 3.1. A seguir é apresentado os dados do formulário para captura explícita, bem como a definição, classificação e agrupamento dos dados do mesmo. Foram utilizados os dados de Idade, Filhos, Veículo, Estado Civil, Residência e Salário, os quais foram fundamentais e garantiram uma melhor qualidade do perfil do usuário e compatibilidade entre os conteúdos publicitários apresentados na seção 6.2.2.

O primeiro campo a ser classificado foi a idade, para a qual se aplicou a classificação do IBGE (2011), conforme apresentado na tabela 7. A classificação foi realizada por estrutura etária, sendo: Criança, Adolescente, Jovem, Adulto e Idoso.

Tabela 8. Classificação por estrutura etária.

Estrutura Etária	Faixa Etária
Criança	0 a 9 Anos
Adolescente	10 a 14 anos
Jovem	15 a 24 Anos
Adulto	25 a 59 Anos
Idoso	Maior que 60 Anos

Fonte: IBGE, 2011

Para o campo Filhos, a definição foi feita em forma da seguinte pergunta: “Tem filhos?”. O usuário preencheu S para Sim ou N para Não. A mesma regra se aplicou para o campo veículo, onde se fez a pergunta: “Tem veículo?”. Já o campo Estado civil foi preenchido de acordo com os principais itens, segundo IBGE (2011), sendo: Casado(a), Solteiro(a), Divorciado(a) e Viúvo(a). O Campo Residência foi preenchido com uma das opções: Própria, Alugada ou Financiada. E, para finalizar, foi solicitado o preenchimento do campo Salário, inserindo a quantidade de salário(s) mínimos(s) do usuário. Esse dado permitiu obter a classe social do telespectador, segundo classificação do IBGE (DATOS, 2011), conforme apresentado na Tabela 8.

Tabela 9. Classificação social.

Classe Social	Renda Familiar
Classe E	Até 1 Salário Mínimo
Classe D	De 2 a 3 Salários Mínimos
Classe C	De 4 a 5 Salários Mínimos
Classe B	De 5 a 14 Salários Mínimos
Classe A	Acima de 15 Salários Mínimos

Fonte: Dados Marketing, 2011

O meio de captura implícito registrou quais categorias de notícias foram acessadas pelo usuário. A definição dessas categorias baseou-se nos três maiores portais de notícias online do Brasil: UOL, Terra e Globo.com (COMSCORE, 2010). Para tornar o projeto mensurável, foram definidas doze categorias baseadas nas principais dos portais de notícias citados.

Cada categoria é composta por várias subcategorias, em sua maioria já mapeadas pelos portais. Na tabela 9, é apresentado um resumo das categorias e os conteúdos relacionados que as compõem.

Tabela 10. Categorias e suas composições

Categoria	Conteúdos relacionados
Carros	Test Drives, Comparação, Motos, Marcas, Avaliações, Lançamentos, Salão do Automóvel, Segredos Automotivos, entre outras notícias relacionadas.
Casa e Decoração	Casa Cor, Cômodos, Construção e Reforma, Decoração, Design, Jardinagem, Projetos, passo a passo etc.
Ciência	Astronomia, Cérebro e Mente, Meio Ambiente etc.
Economia	Bolsa de Valores, Indicadores, Câmbio, Fundos, Índices Econômicos, Cotações, Finanças pessoais, Crise Econômica, etc.
Educação	Onde Estudar, Pais e Professores, Pesquisa Escolar, Testes e Simuladores, Vestibular, Enem, Fies, Prouni, Ensino à Distância, Intercâmbio, Ideb, Ortografia, entre outros temas relacionados.
Empregos	Carreira, Currículos, Profissões, Concursos, Vagas, Salários, Pesquisas, Estágios, Trainee etc.
Esportes	Futebol, Fórmula 1, Basquete, Vôlei, Tênis, Lutas, Atletismo, Beisebol, Canoagem, Ciclismo, Copa do mundo, Olimpíadas, entre outros.
Games	Jogos, Análises, Prévias, Galerias, Eventos etc.
Política	Julgamentos, Escândalos, Pesquisas, Políticos, Notícias, Eleições etc.
Saúde	Alimentação, Bem Estar, Calculadoras, Dengue, Dieta, Boa Forma, Mitos e Verdades.
Tecnologia	Apple, Curiosidades, Dicas, Internet, Segurança, Microsoft, Lançamentos, Testes, Produtos.
Viagem	Destinos Nacionais, Destinos Internacionais, Ecoturismo, Aventura, Mochileiros, Lugares, Vistos, Passaporte, Mapas, Campanhas Aéreas, Consulados, Turismo.

Adaptado de UOL, Terra e Globo.com, 2012.

6.2. Realização do Experimento

Para alcançar os objetivos citados anteriormente, foi aplicado o sistema *Smart Marketing* a um grupo de vinte e dois usuários. Posteriormente, foi possível avaliar o grau de eficiência do conteúdo publicitário recomendado.

Para a realização do experimento, utilizou-se o software de emulação de aplicativos para TV Digital Astrobox, apresentado na seção 5.5.1. O conteúdo apresentado aos usuários foi definido baseando-se nos três maiores portais de notícias citados na definição dos dados (Seção 6.1), sendo que foi apresentada

aos usuários a principal manchete do dia da aplicação do experimento, de cada categoria de notícias.

A escolha para participação do experimento foi feita de forma totalmente aleatória sem que houvesse qualquer elemento determinante para a região e/ou instituição no País, já que o aplicativo *Smart Marketing* pode ser utilizado em qualquer lugar do Brasil, sem restrições.

O experimento procedeu-se a partir de uma solicitação formal (APÊNDICE A) à direção da instituição de ensino para efetuar o convite aos alunos, professores e funcionários, a fim de obter um público misto, contemplado perfis socioeconômicos variados, e sempre deixando claros os objetivos científicos do experimento.

Dado o grande interesse de participação, foi utilizado um critério de seleção, onde se buscou o número de registro dos interessados e foi realizado um sorteio, estabelecendo, assim, os participantes do experimento.

Em seguida ao sorteio, foi passada a orientação quanto ao procedimento desejado: o preenchimento dos dados, a navegação pelo aplicativo e como funcionaria o processo de recomendação. Tomou-se o cuidado de informá-los das questões de segurança, frisando que não haveria qualquer compartilhamento de dados pessoais ou mesmo a identificação do usuário, visto que este não é o objetivo desse estudo.

Após a definição dos 22 usuários e breve apresentação do aplicativo, houve a aplicação do experimento propriamente dita, realizada de forma individual durante o período de 5 (cinco) dias. Na tabela 10, pode-se observar o cronograma de execução. A apresentação do experimento foi realizada em duas turmas da instituição de ensino superior na cidade de Maringá, sendo uma do primeiro ano do curso de Análise e Desenvolvimento de Sistemas e outra do segundo ano de Análise e Desenvolvimento de Sistemas, além do grupo de funcionários da instituição de diferentes setores.

Tabela 11. Descrição das atividades realizadas por dia

Dia	Atividade 01	Atividade 02
1º Dia	Cadastro Smart Marketing	1º Acesso às Notícias
2º Dia	2º Acesso às Notícias	
3º Dia	3º Acesso às Notícias	
4º Dia	4º Acesso às Notícias	
5º Dia	5º Acesso às Notícias	
6º Dia	Visualização das Publicidades	Avaliação das Publicidades

Observou-se que 71% dos usuários participantes eram do sexo masculino (Figura 35); isso se deve ao forte interesse dos alunos do sexo masculino em cursos ligados à área de tecnologia. Também foi verificado um grande interesse dessas turmas sobre o funcionamento, tecnologia e oportunidades ligadas à área de TV Digital. Todavia, o desconhecimento sobre o assunto foi nítido, demonstrando que investimento para fomentar a área de TV Digital se faz necessário.

**Figura 35.** Gráfico de usuários por sexo.

Outro dado relevante na pesquisa trata-se da idade do público participante (Figura 36), sendo predominantemente Adultos (57%) e Jovens (38%), contendo apenas 5% de adolescentes; não participaram crianças e idosos.

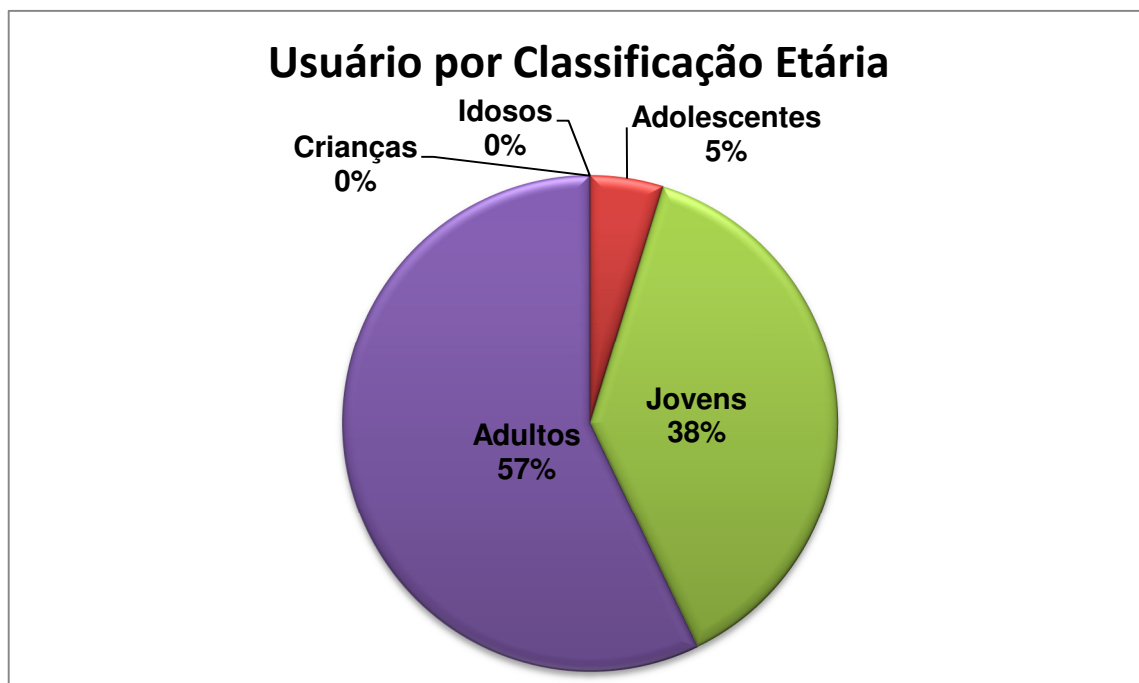


Figura 36. Gráfico de usuários por classificação etária, de acordo com IBGE.

Os usuários também foram seccionados por classe social (Figura 37), sendo em sua maioria pertencentes à classe C, 57%; também houve representantes da classe D, 25%, e Classe B, 19%. Não foram identificados usuários pertencentes às classes sociais A e E.

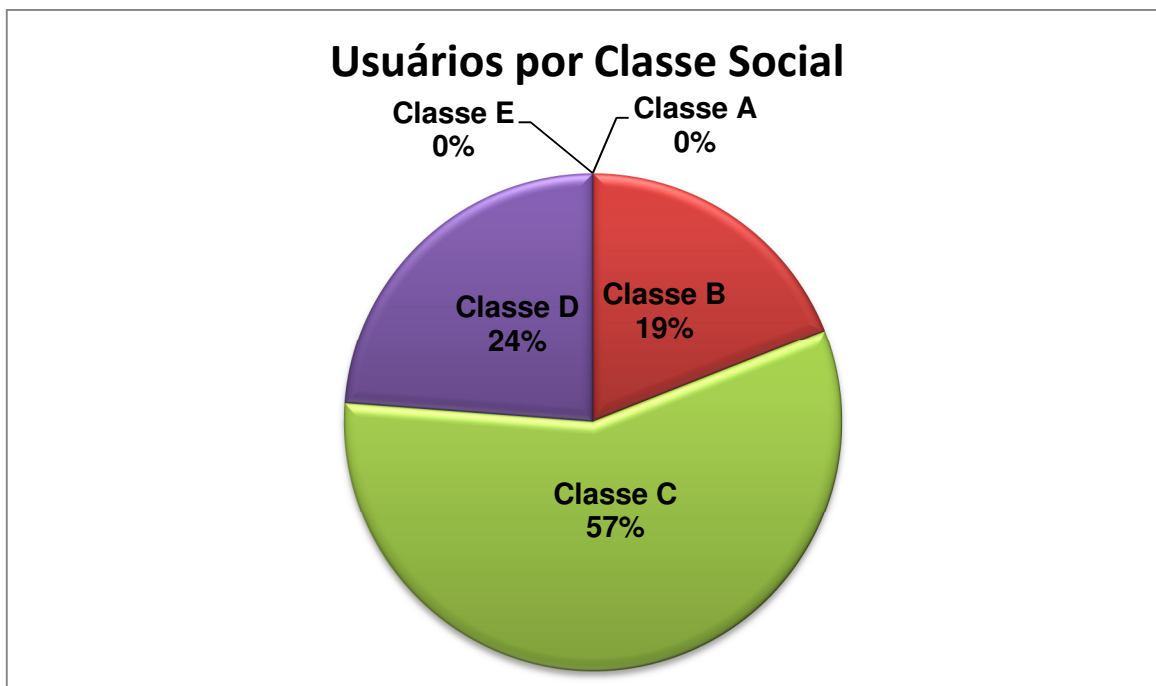


Figura 37. Gráfico de usuários por classe social.

No que diz respeito ao estado civil dos usuários participantes, houve representantes de todos os classificadores. É possível observar que a sua maioria está entre casados e solteiros, cada um contabilizando 43%; a parcela de divorciados e viúvos é pequena, sendo 9% e 5% respectivamente (Figura 38).

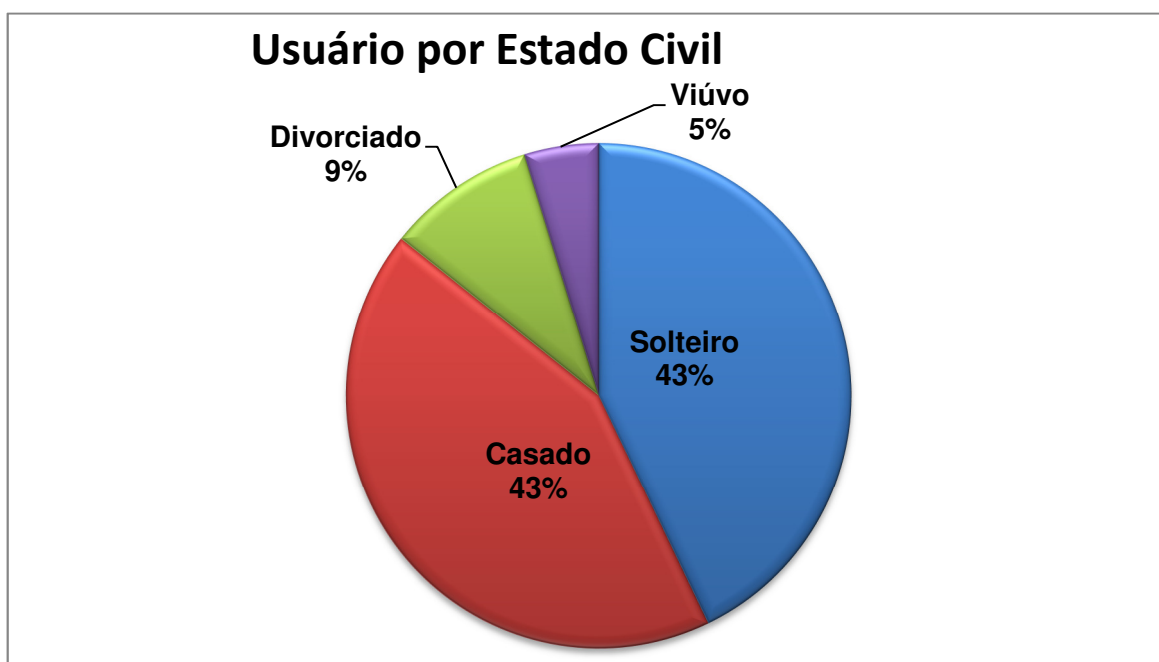


Figura 38. Gráfico de usuários por estado civil.

A Figura 39 apresenta os usuários que possuem ou não filhos, sendo 24% e 76%, respectivamente.



Figura 39. Gráfico de usuários com e sem filhos.

A Figura 40 mostra uma tendência da cidade que é a terceira no país com maior quantidade de veículos por habitante, segundo Guedes (2012). Maringá conta com 68,5% da população como proprietária de veículo(s); nos usuários pesquisados, o resultado ficou ligeiramente maior, com 76% dos usuários possuindo veículo e 24%, não.



Figura 40. Gráfico de usuários com ou sem veículos.

Outro dado relevante na pesquisa refere-se ao tipo de residência dos usuários (Figura 41), onde predominam residências próprias com (76%); em seguida, casas alugadas, com 19% e uma pequena parte, 5%, financiada.

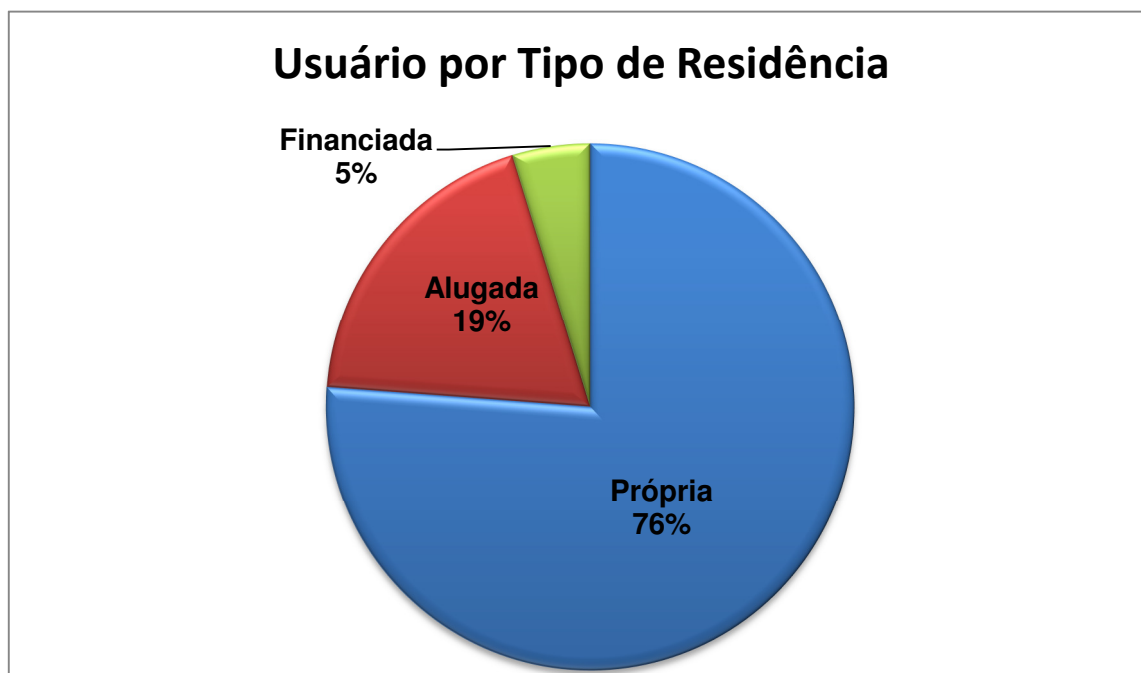


Figura 41. Gráfico de usuários por tipo de residência.

Conforme já mencionado anteriormente, após a definição dos usuários participantes no experimento, iniciou-se a aplicação do *Smart Marketing*. No primeiro dia, os usuários efetuaram o cadastro no aplicativo com seus dados pessoais (captura explícita) e, em seguida, foi feito o primeiro acesso às notícias do *Smart Marketing* (captura implícita). Ao final dessas atividades, também lhes foi entregue o termo de consentimento da participação no experimento, sendo uma via para o usuário e outra para o pesquisador (Apêndice B).

Para o controle de frequência dos 22 usuários que participaram do experimento, fez-se uma planilha de participação, contendo o login do usuário e os dias de participação do mesmo (Apêndice C). O tempo de realização das atividades no primeiro dia, para cada usuário, foi, em média, de 12 (doze) minutos.

No segundo dia, seguiu-se o mesmo procedimento feito no dia anterior, com a diferença de que o usuário apenas se conectou no sistema, ao invés de se cadastrar. Foram apresentadas aos usuários (todos compareceram) as notícias de destaque do dia em cada categoria. O processo de utilização do aplicativo foi mais rápido comparado ao primeiro acesso, pelo fato de seu cadastro já ter sido feito no primeiro dia, e também por já estarem familiarizados com a sua interface, levando em torno de 6 (seis) minutos por usuário. Houve algumas perguntas por parte deles com relação à TV Digital, como os seus aplicativos são acessados, a necessidade de haver uma conexão com a internet, entre outras dúvidas simples.

Do terceiro até o último dia do experimento, foi observada uma padronização no tempo de acesso (aproximadamente 4 minutos), e também não houve usuários ausentes nesses dias.

6.2.1. Aplicação da Descoberta do Conhecimento

Após concluir a aquisição e armazenamento dos dados, iniciou-se o processo de descoberta do conhecimento, onde foi realizada posteriormente a montagem do modelo do sistema de recomendação utilizado no *Smart Marketing*, já apresentado na seção 3.3.

A primeira etapa foi a seleção dos dados para o processo de descoberta do conhecimento. A etapa de pré-processamento se encarregou da limpeza dos dados, como registros duplicados, vazios, erros de acentuação etc. Foram desconsideradas as notícias acessadas durante um tempo igual ou inferior a 10 segundos, sendo consideradas como acessos indesejados ou acidentais. Isso tornou os dados mais confiáveis.

Na etapa de transformação, com o objetivo de obter uma melhor representação dos dados e maior eficiência dos algoritmos, utilizou-se uma amostragem da base original, contendo aproximadamente 70% dos dados. O processo de escolha foi a partir da categoria de notícias mais acessada por cada usuário, e posterior à seleção dos dados, foi gerado um arquivo ARFF (*Attribute-Relation File Format*) com os resultados (Figura 42).

```
@relation Resultado
@attribute Descricao {Carros, CasaDecoracao, Ciencia, Economia,
Educacao, Empregos, Esportes, Games, Política, Saude, Tecnologia, Viagem}
@attribute Sexo {M, F}
@attribute Filhos {Sim, Nao}
@attribute Veiculo {Sim, Nao}
@attribute Idade {Adolescente, Jovem, Adulto}
@attribute EstadoCivil {Solteiro, Casado, Divorciado, Viuvo}
@attribute ClasseSocial {ClasseB, ClasseE, ClasseD, ClasseC}
@attribute Residencia {Propria, Alugada, Financiada}

@data
Esportes,M,Sim,Sim,Adulto,Casado,ClasseD,Propria
Games,M,Sim,Sim,Adulto,Casado,ClasseD,Propria
Economia,M,Nao,Sim,Jovem,Divorciado,ClasseC,Alugada
Games,M,Nao,Sim,Jovem,Solteiro,ClasseD,Propria
Empregos,M,Nao,Sim,Adulto,Casado,ClasseD,Propria
Esportes,M,Nao,Nao,Jovem,Solteiro,ClasseE,Propria
Games,M,Nao,Nao,Jovem,Solteiro,ClasseE,Propria
Educacao,M,Sim,Sim,Adulto,Casado,ClasseC,Propria
Empregos,M,Sim,Sim,Adulto,Casado,ClasseC,Propria
Política.M.Sim.Sim,Adulto,Casado,ClasseC,Propria
```

Figura 42. Amostra do arquivo ARFF gerado

Com os dados selecionados, pré-processados e transformados, iniciou-se a etapa de mineração de dados com o auxílio do software Weka, apresentado na seção 5.5. Os dados foram submetidos à tarefa de classificação por meio do

algoritmo J48, sendo 80% dos dados para o treinamento e 20% para validar o modelo gerado.

O resultado foi uma árvore com total de 26 nós, constituídos por 18 nós folhas e 8 nós de decisão. A parcela de treinamento obteve 64,9% de instâncias classificadas corretamente, contra 57,1% de forma incorreta. A Estatística Kappa atingiu 0.58, valor considerado moderado de acordo com a tabela de valores apresentada na seção 3.3.5.1 (LANDIS, 1977). Já a parcela de validação atingiu resultados parecidos com o treinamento, obtendo 63,4% de instâncias classificadas corretamente e 0.61 na Estatística Kappa, valor considerado substancial. A figura 43 apresenta a árvore gerada.

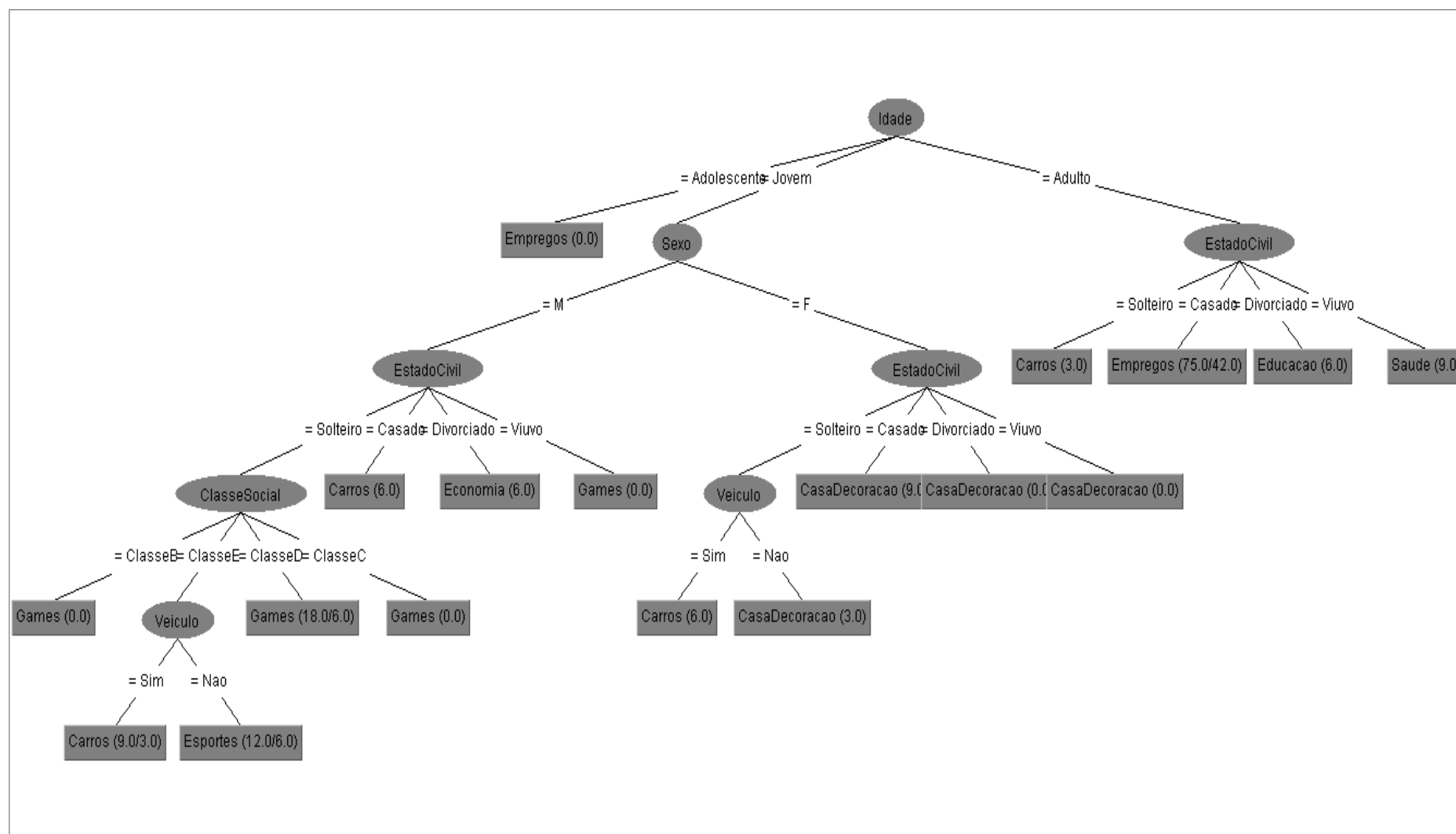


Figura 43. Árvore de decisão gerada através do algoritmo J48.

Com a árvore de decisão gerada, foi realizada uma análise para definição dos perfis, chegando a quatorze afirmações:

- Perfil1 – Adolescentes, independente do sexo, leem notícias de empregos.
- Perfil2 – Adultos e viúvos, independente do sexo, acessam notícias sobre saúde.
- Perfil3 – Jovens do sexo masculino e casados acessam notícias relacionadas a carros.
- Perfil4 – Jovens, do sexo masculino, solteiros e que possuem veículo acessam notícias relacionadas a carros.
- Perfil5 - Jovens, do sexo feminino, solteiras e que têm veículo acessam notícias relacionadas a carros.
- Perfil6 – Adultos, divorciados, independente do sexo, leem notícias sobre educação.
- Perfil7 – Homens, jovens e viúvos acessam notícias relacionadas a games.
- Perfil8 – Homens, jovens, solteiros pertencentes às classes sociais B, C e D acessam notícias relacionadas a games.
- Perfil9 – Homens, jovens e divorciados leem notícias sobre economia.
- Perfil10 – Mulheres, jovens, sejam elas casadas, divorciadas ou viúvas, leem notícias sobre casa e decoração.
- Perfil11 – Mulheres, jovens, solteiras e sem veículo também leem notícias sobre casa e decoração.
- Perfil12 - Pessoas adultas e divorciadas, independente do sexo, acessam notícias sobre educação.

- Perfil13 – Adultos, solteiros, independente do sexo, acessam notícias sobre veículos.
- Perfil14 – Jovens, do sexo masculino, solteiros, pertencentes à classe E, sem veículo acessam notícias sobre esportes.

Com o intuito de efetuar uma comparação para analisar a eficiência do algoritmo J48, além da técnica de classificação este projeto utilizou-se também da técnica de agrupamento, por meio do algoritmo KMeans; na parametrização do algoritmo, foram testadas as funções *Euclidean Distance*, *Manhattan Distance* e *Chebyshev Distance*, entretanto ficou definida a primeira, por ter obtido um melhor resultado.

Foram usados 80% dos dados para treinamento e 20% para validar o modelo de grupo (cluster) gerado, assim como na técnica de classificação. Levando-se em consideração a soma dos erros quadráticos entre clusters apresentados pelo Weka, o menor índice de erro foi obtido utilizando-se 10 clusters, conforme apresentado na figura 44.

Attribute	Cluster#										
	Full Data (162)	0 (9)	1 (39)	2 (18)	3 (15)	4 (27)	5 (18)	6 (12)	7 (6)	8 (9)	9 (9)
Descricao	Empregos	Saude	Empregos	Carros	Esportes	Games	Empregos	Carros	Educacao	Empregos	Saude
Sexo	M	F	M	M	M	M	M	M	F	F	F
Filhos	Nao	Sim	Nao	Nao	Nao	Nao	Sim	Nao	Sim	Nao	Nao
Veiculo	Sim	Nao	Sim	Sim	Nao	Sim	Sim	Sim	Nao	Sim	Sim
Idade	Adulto	Adulto	Adulto	Jovem	Jovem	Jovem	Adulto	Jovem	Adulto	Adulto	Adulto
EstadoCivil	Casado	Viuvo	Casado	Solteiro	Solteiro	Solteiro	Casado	Casado	Divorciado	Casado	Casado
ClasseSocial	ClasseD	ClasseC	ClasseC	ClasseE	ClasseE	ClasseD	ClasseD	ClasseD	ClasseD	ClasseD	ClasseD
Residencia	Propria	Financiada	Propria	Propria	Propria	Propria	Propria	Alugada	Propria	Alugada	Alugada

Figura 44. Clusters gerados através do algoritmo Kmeans.

A partir da análise da figura 44, pôde-se verificar que o agrupamento aglutina interesses em comum dos usuários, sendo:

- Cluster0 - Mulheres adultas, viúvas, com filhos, sem veículo, com casa financiada e pertencente à classe C, leem notícias sobre saúde.
- Cluster1 – Homens, adultos, casados, sem filhos, com veículo, pertencentes à classe C e com residência própria, acessam notícias sobre empregos.
- Cluster2 - Jovens, solteiros, do sexo masculino, pertencentes à classe E, com casa própria e veículo e sem filhos, acessam notícias relacionadas a veículo.
- Cluster3 - As notícias relacionadas a esportes são acessadas por homens, jovens, pertencentes à classe E, com casa própria, sem filhos e sem veículo.
- Cluster4 - Os usuários que leem notícias sobre games são homens, jovens, solteiros, sem filhos, com veículo, pertencentes à classe D e com residência própria.
- Cluster5 – Homens, adultos, casados, com veículo, filhos e casa própria acessam notícias sobre empregos.
- Cluster6 – Homens, jovens, casados, sem filhos, com veículo e residentes em casa alugada, pertencentes à classe D, acessam notícias relacionadas a carros.
- Cluster7 – Mulheres, adultas, divorciadas, com filhos e sem veículo, pertencentes à classe D e que possuem casa própria, acessam notícias sobre educação.
- Cluster8 – Mulheres, adultas, casadas, sem filhos, com veículo, da classe D e que moram em casa alugada leem notícias sobre empregos.

- Cluster9 – Mulheres, adultas, casadas, sem filhos, com veículo, pertencentes à classe D e que residem em casa alugada, leem notícias relacionadas à saúde.

A ferramenta Weka possibilita, além da análise direta apresentada por meio do agrupamento, a extração de conhecimento de acordo com o interesse do especialista. Por conseguinte, foram realizadas combinações diferentes entre os eixos X (linha), Y (coluna) e a classe de cores, permitindo assim diferentes inferências sobre os dados coletados.

Na figura 45 é mostrado um exemplo onde o eixo X contém as categorias de notícias acessadas, o eixo Y apresenta o sexo dos usuários, e a classe de cores secciona o seu estado civil, sendo: azul = solteiro, vermelho = casado, verde = divorciado e azul claro = viúvo.

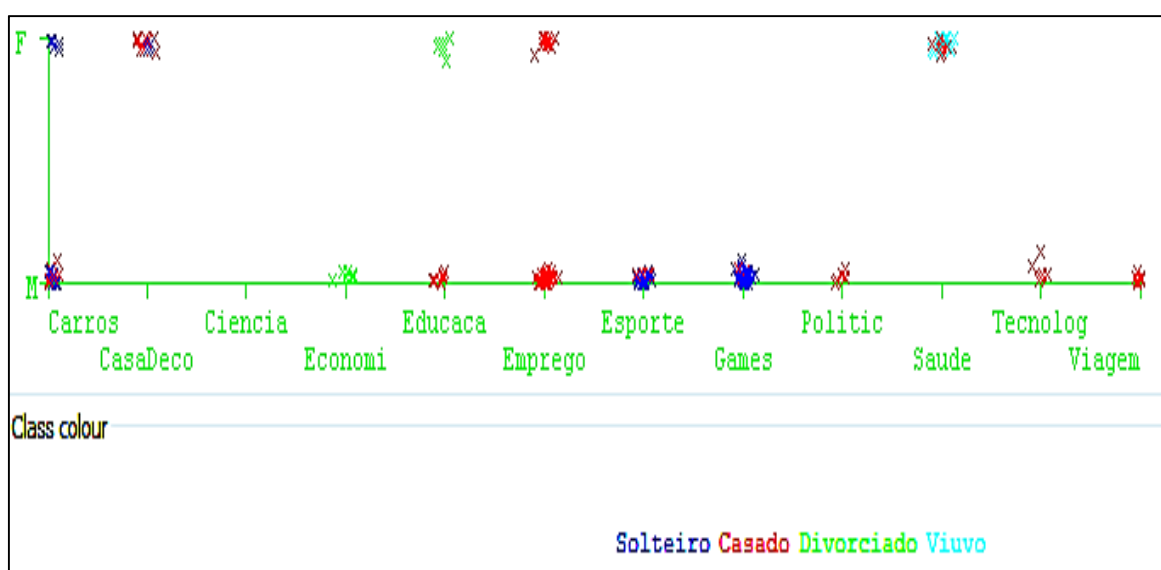


Figura 45. Clusters gerados através do algoritmo Kmeans.

A partir da análise da figura 45, pode-se inferir que:

- Usuários do sexo feminino divorciados tendem a acessar notícias relacionadas à educação.
- Tanto mulheres quanto homens casados, em sua maioria, acessam notícias relacionadas a empregos.

- Mulheres viúvas leem notícias relacionadas à saúde.
- Jovens do sexo masculino acessam notícias relacionadas a jogos e esportes.

Outra combinação de dados pode ser observada na figura 46, que mescla as categorias de notícias mais acessadas (linha X) e a classe social dos usuários (coluna Y e classe de cores); as cores equivalem a: azul = classe B, vermelho = classe E, verde = classe D e azul claro = classe C.

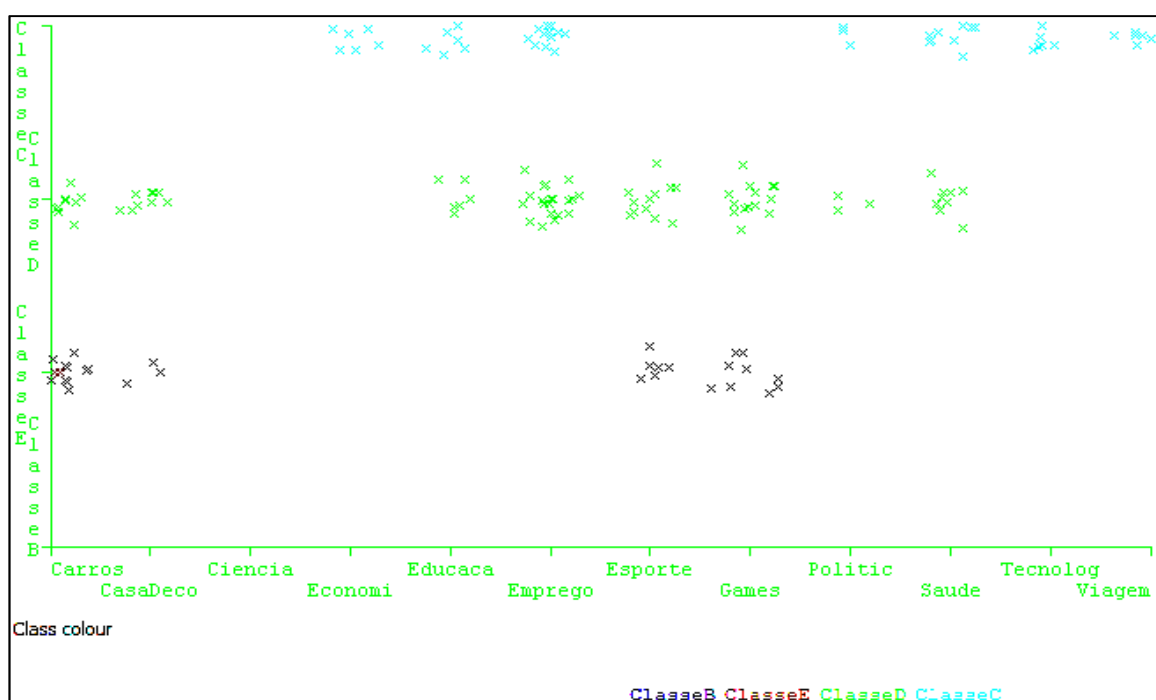


Figura 46. Clusters gerados através do algoritmo Kmeans.

Pode-se concluir que:

- Pessoas da classe E acessam notícias de carros, casa e decoração, esportes e games.
- Pessoas da classe D acessam notícias sobre carros, casa e decoração, educação, emprego, esportes, games, política e saúde.
- Os pertencentes à classe C acessam notícias relacionadas a economia, educação, emprego, política, saúde, tecnologia e viagens.

Analisando o resultado acima, conclui-se que, quanto maior a classe social pertencente, maior o interesse pelo consumo relacionado a tecnologia, viagens e saúde, além do aumento do interesse por política e saúde.

Após a análise dos resultados obtidos através do processo de descoberta do conhecimento, foram criados os códigos das consultas em SQL, para associar cada usuário ao respectivo perfil gerado pela tarefa de Classificação (Apêndice E), e também ao respectivo cluster, gerado a partir da tarefa de Agrupamento (Apêndice F). Tal associação permitiu a inserção dos banners de publicidade correspondentes a cada perfil/cluster do usuário.

6.2.2. Definição da apresentação das publicidades

A escolha das publicidades a serem apresentadas foi embasada em artigos e pesquisas que identificaram o público alvo de determinados produtos e/ou serviços. A seguir são apresentados dois exemplos de *targets* (alvos) publicitários para veículos.

O primeiro exemplo, citado por Camacho (2012), apresenta o Volkswagen Gol (Figura 47) como sendo um carro voltado para o público das classes B e C, do sexo masculino, casado e adulto (de 35 a 49 anos).

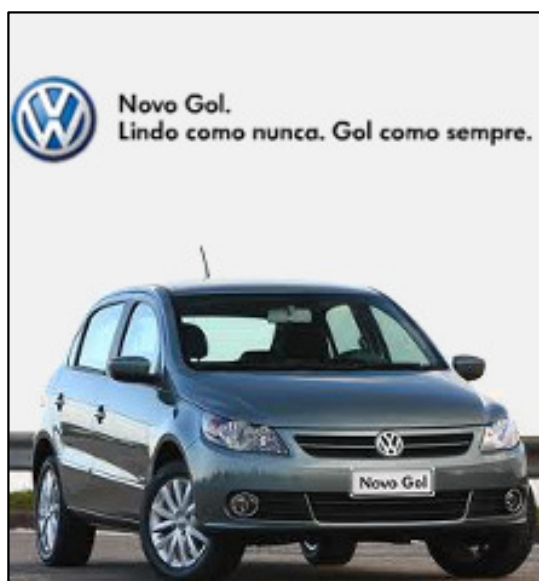


Figura 47. Publicidade apresentada para o publico alvo.

Outro exemplo pode ser visto na figura 48 a revista WebMotors (2012), após realizar uma pesquisa, inferiu que o público alvo do veículo Citroën C3 é o feminino, pois se trata de um automóvel de fácil condução e estacionamento.

Segundo Polizei (2012), as mulheres se preocupam com design, praticidade, dão atenção ao estilo mais harmônico e ao acabamento, observando o interior do veículo, se o mesmo possui porta-trecos, espelhos, entre outros acessórios. Sendo assim, a partir da análise do resultado da Árvore de Decisão, foi exibida a publicidade da figura 48 aos usuários do perfil definido como sendo: mulheres, jovens, solteiras, que acessam notícias relacionadas a veículos.



Figura 48. Publicidade apresentada para o público alvo.

Da mesma forma, a seleção das demais publicidades foi feita com base em pesquisa nos sites de fabricantes de produtos ou serviços, análises de mercado e artigos. Vale ressaltar que o estudo sobre *targets* publicitários não é o objetivo dessa pesquisa; para um maior detalhamento se faz necessário um estudo mais aprofundado.

A exibição das peças publicitárias aos usuários foi constituída de três momentos. Primeiro, realizou-se a apresentação da peça publicitária com o

sistema de recomendação desabilitado; em seguida foi apresentada a peça publicitária baseada na tarefa de Classificação e, por fim, a apresentação da publicidade baseada no agrupamento.

Com um questionário em mãos (APÊNDICE G), o usuário, após a visualização de cada publicidade, assinalava uma das opções apresentadas abaixo, de acordo com a relevância da mesma. Foi realizada a seguinte pergunta:

“Considere o quanto a indicação apresentada se adequa à sua pessoa. Considera-se aqui que, quanto mais adequada a recomendação, maior será a relevância para você. Assinale com X.”

- Extremamente Relevante
- Relevante
- Pouco Relevante
- Irrelevante
- Inadequada

O objetivo deste experimento foi avaliar a qualidade do sistema de recomendação proposto, além de validar as duas tarefas apresentadas neste projeto (Classificação e Agrupamento). Outro ponto importante é a observação de um significativo aumento do interesse, por parte dos usuários, na publicidade apresentada utilizando os resultados da recomendação. Com a conclusão dos experimentos, realizou-se a avaliação dos resultados obtidos em forma de gráficos, os quais são apresentados a seguir.

6.3. Resultados

A qualidade das recomendações foi mensurada por meio da opinião de cada usuário participante do experimento, a qual foi expressa através de um formulário em que o usuário assinalou o nível de adequação da recomendação

apresentada, por meio das alternativas pré-definidas conforme apresentado no Apêndice G.

A forma de avaliar o grau de relevância das publicidades apresentadas fundamentou-se em pesquisas de interação do usuário, que estão sendo cada vez mais utilizadas. Tais pesquisas apontam que o nível de satisfação do usuário não representa em sua totalidade a precisão ou revogação das recomendações (ZIEGLER *et al.*, 2005). Com isso, trabalhos nesta área ganham cada vez mais importância. Pu e Chen (2010) propõem um *framework* focado no usuário para a avaliação de sistemas de recomendação, chamado de ResQue (*Recommender System's Quality of user experience*), que apresenta sessenta (60) questões divididas em treze categorias de critérios de avaliação, abordando aspectos relacionados à interface do usuário, facilidade de uso, exploração do contexto e também à qualidade/relevância dos itens recomendados.

Nesse contexto, o ResQue foi aplicado neste trabalho a fim de avaliar a eficiência das técnicas de Classificação e Agrupamento, utilizando especialmente os itens relacionados ao grau de satisfação do usuário em relação à publicidade apresentada, a qual baseou-se na captura de dados.

Para efeito de análise comparativa, as três próximas figuras apresentam, respectivamente, a avaliação dos usuários em relação a cada publicidade visualizada. A figura 49 mostra que a publicidade exibida sem o sistema de recomendação ativo foi assinalada como Relevante por 17% dos usuários, não sendo considerada Extremamente Relevante para nenhum dos participantes do experimento. Somando os resultados Pouco Relevante, Irrelevante e Inadequada, chega-se ao 87%.



Figura 49. Gráfico de relevância da publicidade com o sistema de recomendação desabilitado.

Depois de ativado o sistema de recomendação baseado na tarefa de Classificação, utilizando o algoritmo J48, foi observado um aumento satisfatório em relação à publicidade não personalizada, contabilizando 72% de relevância e 28% como sendo Extremamente Relevante (Figura 50). Os itens: Inadequada, Irrelevante e Pouco Relevante não foram assinalados.

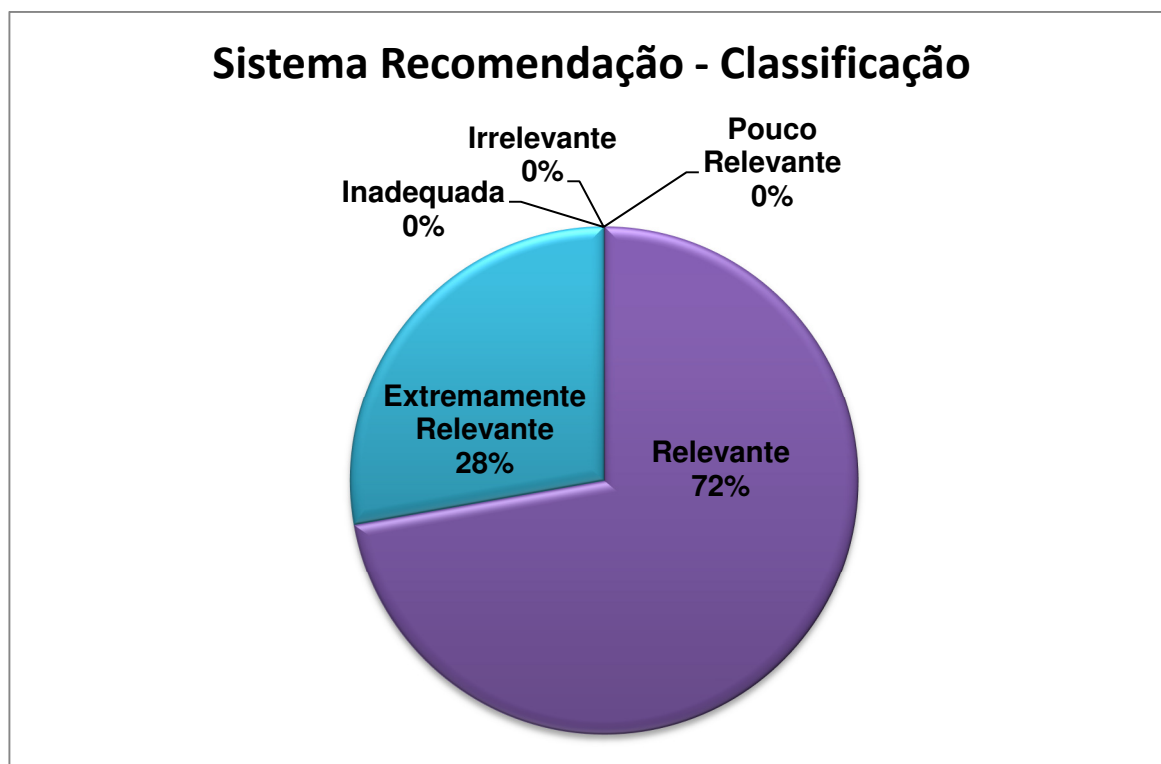


Figura 50. Gráfico de relevância da publicidade baseada na tarefa de Classificação.

O resultado da recomendação utilizando a tarefa de Agrupamento pode ser visto na figura 51, onde o grau de satisfação dos usuários se concentrou em Extremamente Relevante e Relevante, somando 88%.

Comparando esses valores com os obtidos na tarefa de Classificação, percebeu-se um recuo. Isso porque esta contabilizou 12% de publicidades consideradas como Inadequada (6%) e Pouco Relevante (6%). Todavia, quando analisado apenas o item Totalmente Relevante, nota-se que a tarefa de Agrupamento foi mais eficiente, visto que obteve 44% contra 28% assinalados na tarefa de Classificação.



Figura 51. Gráfico de relevância da publicidade baseada na tarefa de Agrupamento.

A comparação entre a publicidade apresentada de maneira genérica e as outras duas exibidas de acordo com o perfil gerado a partir dos sistemas de recomendação pode ser analisada na figura 52, ilustrando de maneira consistente que o ganho em relação à recomendação é significativa, independente da tarefa utilizada.

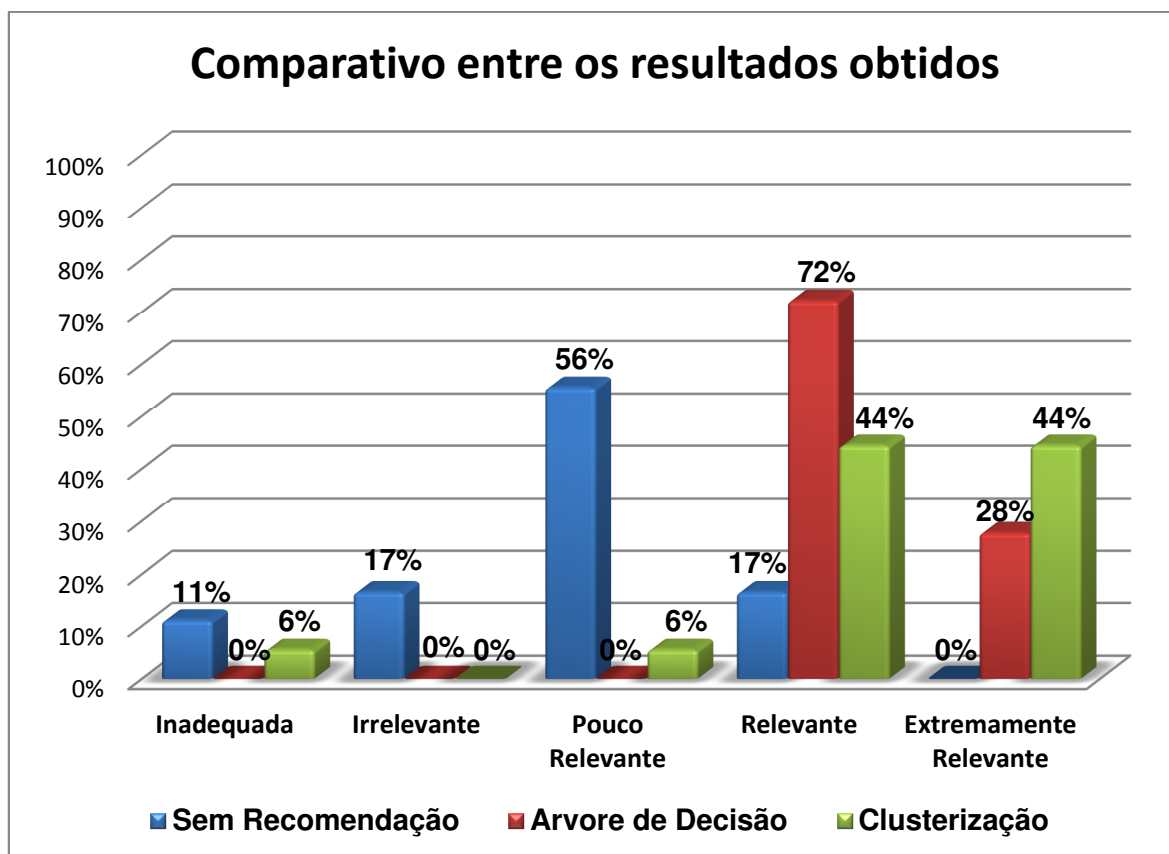


Figura 52. Comparativo entre os resultados obtidos.

Tais resultados confirmam a hipótese deste trabalho, assegurando que sistemas de recomendação de diferentes tarefas, utilizando dados do perfil do usuário e também sua navegação pelo aplicativo, são capazes de realizar recomendação de publicidade de forma satisfatória, aumentando significativamente o interesse do usuário em relação à publicidade inserida no contexto do Sistema Brasileiro de TV Digital.

Pode-se afirmar que, neste trabalho, a árvore de decisão obteve melhores resultados em comparação com a clusterização, isso se deve pelo fato que a árvore de decisão possibilitar um número de combinações muito grande, podendo abstrair conhecimento em apenas 2 níveis da árvore; já a clusterização realiza a criação de grupos bem definidos, limitando o especialista na geração de conhecimento.

Todavia, devem-se considerar a qualidade e a eficiência da peça publicitária ou campanha de marketing. Segundo Burrowes (2005), a sua criação

exige planejamento de comunicação, definido a partir de dados obtidos por meio de pesquisas, como a descoberta do público alvo (idade, sexo, escolaridade, renda, perfil psicológico, localização geográfica, nível cultural), assim como a linguagem, aspectos do produto etc. Sendo assim, o sistema de recomendação não compensará uma possível peça publicitária de baixa qualidade.

6.4. Trabalhos Relacionados

Na literatura, pode-se encontrar uma grande quantidade de trabalhos acadêmicos nas mais diversas áreas do conhecimento, que utilizam sistemas de recomendação com o objetivo de solucionar o problema de sobrecarga de informação, como também para recomendar produtos e/ou serviços, tendo uma maior aplicação no comércio eletrônico de uma forma em geral.

Em relação à TV Digital Brasileira, propostas têm surgido para atender o telespectador na sugestão de programação que se adapte às preferências do mesmo, dentro da imensidão de canais que estão disponíveis.

A recomendação proposta neste trabalho possui um foco diferenciado dos demais trabalhos, realizando a recomendação de publicidade nos aplicativos da TV Digital Brasileira. Até a elaboração deste trabalho, não foram encontradas pesquisas de recomendação de publicidade em aplicativos na TV Digital.

7. CONSIDERAÇÕES FINAIS

Este trabalho desenvolveu um aplicativo para a TV Digital Interativa intitulado *Smart Marketing*, capaz de capturar e armazenar dados explícitos e implícitos, apresentando publicidades personalizadas armazenadas remotamente, de acordo com o interesse do telespectador.

De forma geral, este trabalho apresentou uma nova proposta para tornar mais atraente o mercado de aplicativos para a TV Digital por meio de publicidades inteligentes, fazendo com que:

- As empresas difusoras encontrem estímulo para realizar investimentos nesta área;
- Os telespectadores tenham uma experiência de interatividade mais prazerosa; e
- Por fim, as empresas de publicidade ganhem um novo meio de exibição de seus produtos diretamente ao seu público alvo, aumentando assim a eficácia da publicidade.

Para desenvolver o *Smart Marketing*, optou-se pelo middleware AstroBox da TOTVS, por se tratar de uma ferramenta madura, além de possuir uma interface do usuário definida, oferecer toda a infraestrutura necessária para o desenvolvimento de aplicativos e, principalmente, por apresentar uma solução de negócios.

Em relação à ferramenta de mineração de dados, optou-se pelo Weka, um software gratuito sob licença GPL, utilizado em vários trabalhos acadêmicos na área de TV Digital, e que oferece a integração com banco de dados de forma online.

Para alcançar o objetivo esperado nesse trabalho, foram cumpridas algumas etapas. A primeira etapa foi realizar um estudo teórico sobre tópicos relacionados ao Sistema Brasileiro de TV Digital, sistemas de recomendação, segurança e privacidade na captura dos dados, apresentando os principais itens que embasaram o desenvolvimento dessa dissertação.

Após realizadas as definições, desenvolveu-se o protótipo de sistema de recomendação de forma modular, permitindo variações de SGBD (Sistema de Gerenciamento de Banco de Dados), bem como variações de softwares de mineração de dados.

Finalizado o aplicativo, o experimento foi realizado utilizando-se da fundamentação do *framework* ResQue, a fim de apontar o nível de satisfação do usuário em relação à peça publicitária apresentada, visto que um bom algoritmo de recomendação por si só não garante a satisfação do telespectador, pois uma peça publicitária mal elaborada não será aderente ao seu público alvo. Conclui-se que avaliar o grau de satisfação do telespectador é crucial para mensurar a eficácia e a relevância das publicidades exibidas aos telespectadores.

7.1. Aprimoramentos

Um dos pontos no qual este trabalho pode ser melhorado trata-se da forma em efetuar a predição por tipos de notícias acessadas, devido à baixa quantidade de logs gerados, o que compromete a qualidade das recomendações. Este problema é conhecido como partida fria, em inglês *cold start problem* (LAM *et al.* 2008).

Para obter uma quantidade substancial de dados (logs de acessos às notícias pelos usuários) foi essencial a aplicação do experimento no período de seis dias, para a geração tanto dos perfis, quanto do log por meio de capturas implícitas e explícitas, seguindo a metodologia já descrita neste trabalho.

Outro fator que deve ser observado é a identificação do telespectador por meio de uma validação (login e senha), a qual pode ser considerada uma abordagem intrusiva e incômoda em termos de segurança e praticidade. Sendo assim, a pesquisa de outros métodos mais eficazes e menos custosos se faz necessária para o SBTVD.

7.2. Trabalhos Futuros

Por ser flexível e de código aberto, espera-se que o *Smart Marketing* receba novas funcionalidades ao longo do tempo. No decorrer desta pesquisa, algumas possibilidades foram identificadas como possíveis trabalhos futuros.

No que diz respeito à obtenção de dados, o *Smart Marketing* utilizou-se da captura implícita de categorias de notícias acessadas, contudo esta é apenas uma das possibilidades que pode ser utilizada. Outros aplicativos como: previsão do tempo, comércio eletrônico, jogos etc. podem oferecer dados valiosos para o processo de recomendação. Logo, pode ser realizada uma captura coletiva dos dados, melhorando ainda mais o processo de recomendação e satisfação do telespectador em relação à sua interação com os aplicativos da TV Digital Interativa Brasileira.

Outro ponto diz respeito ao meio de envio do aplicativo, o qual é disponibilizado nesse trabalho por meio do canal de retorno via conexão web; ele pode ser adaptado para disponibilizar os dados via carrossel de dados por meio da emissora de TV. Também é importante que sejam realizados estudos sobre melhorias no hardware em que a arquitetura Ginga é armazenada, já que o mesmo pode ser considerado obsoleto e limitado (SILVA *et al.*, 2010), inviabilizando novas tecnologias e suas respectivas funções.

Além disso, seria interessante a criação de um sistema que armazena no cache do STB os dados da navegação, sem a necessidade de uma conexão direta e em tempo real com a internet, enviando para um banco de dados externo somente quando a conexão com a internet for estabelecida.

Baseando-se neste trabalho e utilizando metodologia semelhante, poder-se-ão aplicar outras técnicas de recomendação, a fim de se obter melhores resultados.

Sugere-se também a adaptação da implementação do *Smart Marketing* para a utilização em forma de multiusuário, já que as TVs também são acessadas por grupos de pessoas, tornando a recomendação em grupo importante à sociedade.

Outro ponto é a migração do *Smart Marketing* para a TV Digital de aplicativos móveis, pois a tendência ao crescimento no uso de tal tecnologia é evidente nos dias atuais.

Para concluir, um tema a ser analisado em trabalho futuro, e que se configura como a continuidade deste trabalho, se refere à realização de estudos mais detalhados sobre a utilização de Inteligência Artificial e Computação Afetiva em sistemas de recomendação para aplicativos da TV Digital. Dessa vez, elaborar novas modelagens que considerem informações baseadas na psicologia, como fatores culturais, emocionais, sociais e regionais.

8. REFERÊNCIAS

ADOMAVICIUS, G.; TUZHILIN, A. *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions*. In IEEE Transactions on Knowledge and Data Engineering, v.17, n6, p.734-749, 2005.

ANATEL, *TV Digital alcança 46% da população brasileira*. Disponível em: <<http://www.anatel.gov.br/Portal/exibirPortalPaginaEspecialPesquisa.do?acao=&tipoConteudoHtml=1&codNoticia=22469>>. Acesso em 29 Jun. 2011.

ARANHA, Francisco. *Análise de redes em procedimentos de cooperação indireta: utilização no sistema de recomendações da Biblioteca*. Karl A. Boedecker. São Paulo: EAESP/FGV/NPP, 2000. 71p.

ASTROTV. *Astro DEvNet! Developers Network*. Disponível em: <<https://www.astrodevnet.com/AstroDevNet/home.html>>. Acesso em: 12 Jan. 2012.

BALTAR, T. Valeria. OKANO, Valdir. *Análise de Concordância – Kappa*. Disponível em: <<http://www.lee.dante.br/pesquisa/kappa/index.html>>. Acesso em 15 Out. 2012.

BECKER, Valdecir e MORAES, Áureo. *Do analógico ao Digital: uma proposta de comercial para TV interativa*. In: III SIMPÓSIO CATARINENSE DE PROCESSAMENTO DIGITAL DE IMAGENS. 2003, Florianópolis. Anais Florianópolis. 2003. p. 122-134.

BENNETON, Ricardo. *TV Digital no Brasil* In: III SIMPÓSIO CATARINENSE DE PROCESSAMENTO DIGITAL DE IMAGENS. Florianópolis, 2003. Anais Florianópolis: 2003.

BLOCH, Ethan. *Have We Reached a World of Infinite Information?*. Disponível em: <<http://www.flowtown.com/blog/have-we-reached-a-world-of-infinite-information>>. Acesso em: 12 Nov. 2011.

BOENTE, N. P Alfredo. GOLDSCHIMIDT, R. Ronaldo. ESTRELA, V. Vânia. *Uma metodologia para apoio à realização do processo de descoberta de conhecimento em bases de dados*. Banco de Dados e Engenharia de Software, Rio de Janeiro, v.3, p3-18, 2006.

BRASIL, Ministério do Desenvolvimento, Indústria e Comercio Exterior e da ciência, tecnologia e inovação. Portaria Interministerial Nº140 de 23 de Fevereiro de 2012. Estabelece o Processo Produtivo Básico para o produto TELEVISOR COM TELA DE CRISTAL LÍQUIDO. Diário Oficial da união – Seção 1. Pg. 2. Fev. 2012.

BRENTANO, Laura. *TVs conectadas à web atingem 17% do mercado e saem na frente das 3D*. Disponível em: <<http://g1.globo.com/tecnologia/noticia/2011/08/tvs-conectadas-web-atingem-17-do-mercado-e-saem-na-frente-das-3d.html>>. Acesso 14 Ago. 2012.

BUENO, Chris. Quer curtir o verão com aventura? Pratique Rafting com segurança. Disponível em: <<http://360graus.terra.com.br/rafting/default.asp?did=13555&action=geral>>. Acesso 12 Out. 2012.

BURKE, Robin. Hybrid Recommender Systems: Survey and Experiments. User Modeling and User - Adapted Interaction. Department of Information Systems and Decision Sciences. Massachusetts, Novembro, p.331-370, 2002.

BURROWES, Patrícia. Viagem ao território da publicidade. Comunicação, Mídia e Consumo. 2012. São Paulo, vol. 2 n.5 p205-219 Nov. 2005.

CAMACHO, Karen. *Volks aposta no novo Gol para superar vendas da Fiat*. Disponível: <<http://www1.folha.uol.com.br/folha/dinheiro/ult91u416621.shtml>>. Acesso em 24 Mar. 2012.

CASSIA, Fernando. *Padrão de TV Digital brasileiro Conquista a América e parte para o Mundo!*. Disponível em: <<http://itvbr.com.br/blog/sem-categoria/padrão-de-tv-digital-brasileiro-conquista-a-america-e-parte-para-o-mundo>>. Acesso em 18 Mai. 2012.

CC. As licenças. Disponível em: <<http://creativecommons.org.br/as-licencas/>>. Acesso 10 Nov. 2012.

CESAR, P.; CHORIANOPOULOS, K.; JENSEN, J.F. Computers in Entertainment (CIE) - Social television and user interaction. ACM Digital Library. New York, NY, USA, v.4, p1-33, 2008.

COSTA, Cido. *Ginga será instalado em 75% dos televisores de plasma produzidos a partir de 2013*. Disponível em: <<http://www.douradosagora.com.br/tecnologia/ginga-sera-instalado-em-75-dos-televisores-de-plasma>>. Acesso em 30, Mar. 2012.

CPQD. Cartilha de Recomendações de Usabilidade para aplicações em TVDi. Campinas, v1, p1-28, 2012.

CPQD. *CPqD disponibiliza para download novas aplicações para TV digital interativa*. Disponível em: <<http://www.cpqd.com.br/imprensa-e-eventos/fatos/348-fatos-195/6229-cpqd-disponibiliza-para-download-novas-aplicacoes-para-tv-digital-interativa.html>>. Acesso em 28 Jul. 2012.

CRAIDE, Sabrina. *TV digital interativa pode virar política de governo para promover a inclusão digital e social*. Disponível em: <<http://agenciabrasil.ebc.com.br/noticia/2012-03-11/tv-digital-interativa-pode- virar-politica-de-governo-para-promover-inclusao-digital-e-social>>. Acesso em 12 Mar. 2012.

DATOS, Direto Marketing. *Lista de Classes Sociais IBGE*. Disponível: <<http://www.datosmarketing.com.br/listas-detelhes-classes-sociais.asp>>. Acesso em 14 Set. 2011.

DIAS, Maria M. Um Modelo de Formalização do processo de desenvolvimento de sistemas de descoberta de conhecimento em banco de dados. 2001. 197f. Tese (Doutorado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2001.

DIAS, Maria Madalena. *Mineração de Dados (Data Mining)*. Disponível em: <<http://www.des.uem.br/uploads/downloads/1027223849.pdf>>. Acesso em 14 Ago. 2011.

DIGITAL, Convergência. *TV Digital: CPqD amplia leque de aplicativos com Ginga*. Disponível em: <<http://www.cpqd.com.br/noticias-relacionadas/342-noticias-2012/6219-tv-digital-cpqd-amplia-leque-de-aplicativos-com-ginga.html>>. Acesso em 11 Jul. 2012.

DIGITAL, Olhar. *TVs conectadas já são realidade em 38% dos lares dos EUA. Mas a razão não está relacionada às Smart TVs*. Disponível em: <http://olhardigital.uol.com.br/produtos/digital_news/noticias/tvs-conectadas-estao-em-38-dos-lares-dos-eua-e,-nao-gracas-as-smarttvs>. Acesso 21 Abr. 2012.

DTV. Vantagens da TV digital. Disponível em: <<http://www.dtv.org.br/sobre-a-tv-digital/vantagens-da-tv-digital/>>. Acesso 10 Dez. 2011.

DTV, Cronograma de Implantação da TV Digital Brasileira. Disponível em: <<http://www.dtv.org.br/materias.asp?menuid=3&id=11>>. Acesso em 12 Fev. 2011.

POLIZEI, Eder. *Saiba quais são os dez carros preferidos pelas mulheres*. Disponível em: <<http://revista.webmotors.com.br/mercado/saiba-quais-sao-os-dez-carros-preferidos-pelas-mulheres/1334081149081>>. Acesso em: 12 Out. 2012.

EIRINAKI, M., Charalampos, STRATOS, L., VAZIRGIANNIS P. *Personalization Integrating Content Semantics and Navigational Patterns*. Proceedings of the 6th Annual. ACM International Workshop on Web Information and Data Management. 2004.

EMARKETER. *Targeting Boosts Low Facebook Click Rates*. Disponível em: <<http://www.emarketer.com/Article.aspx?R=1008238>>. Acesso 13 Fev. 2012.

ERABAKI. *Platform Independent Text Mining Engine Tool*. Disponível em: <<http://erabaki.ehu.es/jjga/pimiento/>>. Acesso 12 Ago. 2012.

EXPM, 2012. Behavioral targeting aumenta vendas. Fique ligado!. Disponível em: <<http://www.alumniespm.com.br/impressao/behavioral-targeting-aumenta-vendas-fique-ligado/>>. Acesso 14 Out. 2012.

FAYYAD, Usama, SHAPIRO-PIATETSKY, Gregory, SMYTH, Padhraic. *From Data Mining to Knowledge Discovery in Databases*. American Association for Artificial Intelligence. AI MAGAZINE. 0738-4602-1996. 37-54p. 1996.

FILHO, M. F. Fernando. GEUS, L. Paulo. ALBUQUERQUE, P. João. *Sistemas de Recomendação e Interação na Web Social*. Workshop de Aspectos da Interação Humano-Computador na Web Social, Porto Alegre, Brasil, 21 Outubro 2008, SBC. 24-27. ISBN 978-85-7669-213-3.

FILHO, C. S. Manoel. GONDIM, R. L. Paulo. *NCLua SOAP: Acesso a Web Services em aplicações de TVDi*. Coordenação de Informática. Palmas, v1, p1-10, 2011.

G1. *Entenda o ataque à rede on-line do PlayStation 3, a PSN*. Disponível em: <<http://g1.globo.com/tecnologia/noticia/2011/05/entenda-o-ataque-rede-line-do-playstation-3-psn.html>>. Acesso 12 Mai. 2012.

GARCIA, S.C. *O uso de árvores de decisão na descoberta de conhecimento na área da saúde*. In: SEMANA ACADÊMICA, 2000. Rio Grande do Sul: Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.

GINGA, 2012. *Sobre o Ginga*. Disponível em: <<http://www.ginga.org.br/pt-br/sobre>>. Acesso 15 Out. 2012.

GOOGLE, *Preferências*. Disponível em: <<http://www.google.pt/ads/preferences/html/intl/pt-PT/about.html>>. Acesso em 23 Jun. 2012.

GUEDES, Carla. *Maringá tem 3ª maior taxa veículos/habitantes do País*. Disponível em: <<http://www.odiariorio.com/maringa/noticia/320220/maringa-tem-3a-maior-taxa-veiculoshabitantes-do-pais/>>. Acesso em 12 Out. 2012.

HAMANN, Renan. *O que as empresas de internet sabem sobre você?*. Disponível em: <<http://www.tecmundo.com.br/privacidade/3776-o-que-as-empresas-de-internet-sabem-sobre-voce-.htm#ixzz2EPLLaEL>>. Acesso 10 Nov. 2012.

HAN, J., KAMBER, M., *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers, 2001.

HANSON. W. *Principles of Internet Marketing*. South-Western College Publishing. September, 1999.

HERLOCKER, J. L.; KONSTAN, J. L.; TERVEEN, L. G.; RIEDL, J. T. *Evaluating Collaborative Filtering Recommender Systems*. In: ACM Transactions on Information Systems, V22, p.5-53, 2004.

IBGE, *Pesquisa Nacional por Amostra de Domicílios*. Disponível em: <<http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad98/saude/analise.shtm>>. Acesso em 25 Set. 2012.

IBGE. Série Estudos e Pesquisas. Informação Demográfica e Socioeconômica. *População Jovem no Brasil*. Rio de Janeiro, v.1 n.1, p.55, 1999. Disponível em <http://www.ibge.gov.br/home/estatistica/populacao/populacao_jovem_brasil/populacaojovem.pdf>. Acesso em 14 Jul. 2012.

ILLIMINE, *Project Illimine*. Disponível em: <<http://illimine.cs.uiuc.edu/>>. Acesso 12 Ago. 2012.

ITVBR. Engenharia de Sistemas, Interatividade. Disponível em: <http://www.itvbr.com.br/index.php?option=com_content&view=article&id=86:interatividade&catid=49:tv-digital&Itemid=77&lang=pt>. Acesso em: 10 Nov. 2011.

ITVBR. O Sistema Brasileiro de Televisão Digital – SBTVD. Disponível em: <<http://itvbr.com.br/blog/daniel/o-sistema-brasileiro-de-televisao-digital-sbtvd>>. Acesso em: 10 Nov. 2012.

JAIN, A. K. and Dubes, R. C. 1988. Algorithms for clustering data, Inc., Upper Saddle River, NJ, USA.

JUNIOR, Cardozo, Marco, 2012. Algumas razões para utilizar Behavioral Targeting. Disponível em: <<http://www.ecommercebrasil.com.br/artigos/algumas-razoes-para-utilizar-behavioral-targeting/>>. Acesso 14 Out. 2012.

KNIME. *Konstanz Information Miner*. Disponível em: <<http://www.knime.org/>>. Acesso 12 Ago. 2012.

KXEN. *The Predictive Analytics Leader*. Disponível em: <<http://www.kxen.com/>>. Acesso 12 Ago. 2012.

LANDIS, J.; KOCH, G. G. *The measurements of agreement for categorical data*. *Biometrics*, v.33, n.3, p.159-179, 1977.

LAVID, 2012. *Desenvolvimento em Ginga*. Disponível em: <<http://gingacdn.lavid.ufpb.br/>>. Acesso em 16 Abr. 2012.

LEMES, Sara. *TV digital é a chance de inclusão em massa*. Disponível em: <http://www.brasilwiki.com.br/noticia.php?id_noticia=15776>. Acesso em 12 Dez. 2011.

MARTINHAGO, Sergio. Descoberta de conhecimento sobre o processo seletivo da UFPR. 2005. 114f. Dissertação (Mestrado em Ciências) - Programa de Pós-graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Curitiba, 2005.

MARTINS, A. Vagner, FONSECA, M. G. Leila. *Classificação de uso de solo baseada na análise orientada a objeto e mineração de dados utilizando imagens*

SPOT/HRG-5. Anais XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, Brasil, 25-30 Abril 2009, INPE, P. 7837-7844.

MEIRA, W. J., MURTA, D. C., CAMPOS, S., GUEDES D. *Sistemas de Comércio Eletrônico: Projeto e Desenvolvimento*. Rio de Janeiro, 2002, Editora Campus, ISBN 85-352-1012-1, 371p.

MENDONÇA, D. S. Análise Probabilística de Semântica Latente aplicada a Sistemas de Recomendação. 2008. 69f. Dissertação (Mestrado em Informática) – Programa de Pós-graduação em Informática, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

MONTEIRO, Aline. *Após 4 anos, TV Digital chega a 46% do Brasil*. Disponível em: <<http://info.abril.com.br/noticias/mercado/apos-4-anos-tv-digital-chega-a-46-do-brasil-15042011-32.shl>>. Acesso 15 Mai. 2012.

MURAD, Fernando. *Invasão de privacidade?*. Disponível em: <http://www.meioemensagem.com.br/home/marketing/em_perspectiva/2012/04/04/Invasao-de-privacidade.html>. Acesso 22 Mar. 2012.

NETO, Silveira, Antonio. A fragmentação da privacidade. Disponível em: <<http://www.ebah.com.br/content/ABAAAAbjQAG/a-fragmentacao-privacidade>>. Acesso: 10 Nov. 2012.

OLIVEIRA, Giulianna. Saiba tudo sobre televisão digital. Disponível em: <<http://www.tecmundo.com.br/lcd/2134-saiba-tudo-sobre-televisao-digital.htm>>. Acesso 10 Nov. 2012.

PINTO, Joaquim Souza, *et al.* Métodos para Estimação de Reprodutividade de Medidas, Estatísticas Kappa. Disponível em: <<http://users.med.up.pt/joakim/intromed/estatisticakappa.htm>>. Acesso 14 Out. 2012.

POSSEBON, Samuel. *Governo baixa portaria obrigando 75% das novas TVs a terem Giga em 2013*. Disponível em: <<http://www.telaviva.com.br/24/02/2012/governo-baixa-portaria-obrigando-75-das-novas-tvs-a-terem-giga-em-2013/tl/264298/news.aspx>>. Acesso 30 Abr. 2012.

PU, P; CHEN, L. *A User-Centric Evaluation Framework of Recommender Systems*. In: Proceedings of the ACM RecSys 2010 Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces (UCERSTI), Barcelona, Spain, 2010.

RABELO, Emerson. Avaliação de técnicas de visualização para mineração de dados. 2007. 204f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Ciência da Computação, Universidade Estadual de Maringá. Maringá, 2007.

RADFAHRER, Luli. *Sobrecarga de informação vai piorar e exige filtro*. Disponível em: <<http://www1.folha.uol.com.br/tec/1046616-sobrecarga-de-informacao-vai-piorar-e-exige-filtro-diz-luli-radfahrer.shtml>>. Acesso em: 12 Mar. 2012.

RAMISCH, Carlos. Trabalho pratico de mineração de dados. Disponível em: <http://www.inf.ufrgs.br/~ceramisch/download_files/courses/Undergraduate_BRAZIL/UFRGS_2009_1/Topicos_Especiais_em_Computacao_I_-_Mineracao_de_Dados_-_INF01179/Trabalho_1_-_Car_Evaluation/Relatorio.pdf>. Acesso 10 Nov. 2012.

RAVA, Ben-Hur. Internet e invasão de privacidade. Disponível em: <http://www.observatoriodaimprensa.com.br/news/view/internet_e_invasao_de_privacidade>. Acesso 10 Nov. 2012.

REATEGUI, B. Eliseo. CAZELLA, S. César. Um Agente de Inovação e Conhecimento. Sistemas de Recomendação. XXV Congresso da Sociedade Brasileira de Computação, São Leopoldo, Brasil 20 Abril 2005, V Enia, P. 306-348.

RESNICK, P.; VARIAN. H. R. *Recommender Systems. Communications of the ACM*, New York, v.40, n.3, p. 55-58. Mar. de 1997.

SANT'ANNA, Francisco, *et al. Desenvolvimento de Aplicações Declarativas para TV Digital no Middleware Ginga com Objetos Imperativos Lua*. Disponível em: <<http://www.telemidia.puc-rio.br/sites/telemidia.puc-rio.br/files/MCNCLua.pdf>>. Acesso em 12 Out. 2012.

SANT'IAGO, Marcelo. *A maior empresa de publicidade do mundo*. Disponível em: <<http://webinsider.uol.com.br/2011/07/18/a-maior-empresa-de-publicidade-do-mundo/>>. Acesso em: 09 Ago. 2012.

SAS. *Model Development and Deployment*. Disponível em: <<http://www.sas.com/technologies/analytics/datamining/miner/#section=4>>. Acesso 12 Ago. 2012.

SCHAEFER, Melissa. *Winning over the empowered consumer: Why trust matters*. Disponível em: <<http://public.dhe.ibm.com/common/ssi/ecm/en/gbe03483usen/GBE03483USEN.PDF>>. Acesso 12 Abr. 2012.

SILVA, S. Glauco; SEGUNDO, M. C. Ricardo; PEREIRA, R. C. Alisson;

SIMOES, Clecia. *Behavioral Targeting mapeia comportamentos*. Disponível em: <<http://webinsider.uol.com.br/2007/03/30/behavioral-targeting-mapeia-comportamentos-e-vende-mais/>>. Acesso 15 Mar. 2012.

TAN, Pang-Ning, VIPIN, Kumar, MICHAEL, Steinbach. *Introdução ao Data Mining - Mineração de Dados*. Ed. Ciencia Moderna. 1ed. 928f. 2009.

TAVARES, A. Tatiana. Proposta de animação de jogos 2D para TV Digital. SBC - Proceedings of SBGames. Florianopolis - SC. 8º Edição 10th. 316 - 319. 2010.

TOZETTO, Claudia. *TV pública quer liderar criação de conteúdo para Ginga*. Disponível em: <<http://tecnologia.ig.com.br/especial/tv-publica-quer-liderar-criacao-de-conteudo-para-ginga/n1597727279334.html>>. Acesso em 29 Abr. 2012.

TAVARES, M. L. Walkyria. Implantação da Televisão Digital no Brasil. Disponível em: <<http://www2.camara.leg.br/documentos-e-pesquisa/publicacoes/estnottec/pdf/108553.pdf>>. Acesso: 10 Nov. 2012.

WEBINAR, eMarketer. Targeting Boosts Low Facebook Click Rates. Disponível em: <<http://www.emarketer.com/Article.aspx?R=1008238>>. Acesso 10 Nov. 2012.

WEBMOTORS. *Saiba quais são os dez carros preferidos pelas mulheres*. Disponível em: <<http://revista.webmotors.com.br/mercado/saiba-quais-sao-os-dez-carros-preferidos-pelas-mulheres/1334081149081>> Acesso em 12 Out. 2012.

WEKA. *Data Mining Witch Open Source Machine Learning Software in Java*. Disponível em <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso 12 Ago. 2012.

ZIEGLER, C. N.; MCNEE. S. M; KONSTAN, J.A; LAUSEN, G. *Improving Recommendation Lists through Topic Diversification*. In: Proceedings of WWW 2005, ACM Press (2005). p-22-32, 2005.

9. APÊNDICES

APÊNDICE A – Termo de consentimento da instituição

Maringá 02 Outubro 2012.

À FCV (Faculdade Cidade Verde).

Prezada Professora Simone Regina Silva,
coordenadora do curso de graduação em Análise e Desenvolvimento de Sistemas.

Com a implantação do Sistema Brasileiro de TV Digital (SBTVD), inicia-se uma gama de novas oportunidades e possibilidades tanto para o telespectador quanto para as emissoras de TV. Para os telespectadores, uma imensa quantidade de canais, programas e propagandas interativas. Para as emissoras de TV, o aumento da possibilidade de propagandas em novos meios de comunicação. Nesse contexto, surge a oportunidade da utilização das técnicas de recomendação, com o intuito de personalizar o conteúdo a ser apresentado ao telespectador – seja ele propaganda, canais ou programas – utilizando uma abordagem baseada no perfil do usuário, histórico de uso ou sensível ao contexto.

A dissertação intitulada: *Smart Marketing* na TV Digital Interativa através de um sistema de recomendação de anúncios. Tem por objetivo a exibição de publicidade baseada nos dados obtidos de forma explícita e implícita, através da navegação em aplicativos Ginga. Com os dados armazenados é feito um processo de KDD (descoberta de conhecimento), utilizando o software de mineração de dados Weka.

Para o experimento, foi desenvolvido um aplicativo para a TV Digital, que será utilizado a fim de obter os dados explícitos e implícitos do usuário (telespectador). O aplicativo é dividido em abas, sendo: Cadastro do Telespectador; Login; Notícias; Recomendação; Segurança e Privacidade; Direitos de Uso; e Ajuda. O primeiro campo a ser preenchido pelo usuário é o Cadastro do Telespectador, iniciando pelo login e senha e seus respectivos dados pessoais. Será tomado o cuidado de informar aos participantes as questões de segurança, como o compartilhamento de dados pessoais ou mesmo a identificação do usuário, visto que este não é o objetivo deste estudo.

O objetivo deste experimento é avaliar a qualidade do sistema de recomendação proposto, além de validar o crescimento do interesse pela publicidade apresentada no aplicativo do sistema brasileiro de TV Digital.

Contando com o apoio da Coordenação do curso de Análise e Desenvolvimento de Sistemas e o apoio da Instituição por estarem cedendo o espaço físico e seus alunos, professores e funcionários, agradeço antecipadamente pela presteza e atenção.

Alan Menk

Mestrando em Gestão de Redes e Telecomunicação

PUC-Campinas

APÊNDICE B - Termo de consentimento do usuárioTERMO DE CONSENTIMENTO – PARTICIPAÇÃO VOLUNTÁRIA NA
AVALIAÇÃO DO SOFTWARE SMART MARKETING

Prezado Sr(a),

Convidamos o (a) Sr (a) para participar da Avaliação do Software *SMART MARKETING*, que é um produto resultante do desenvolvimento da dissertação de Mestrado intitulada “*Smart Marketing* na TV Digital Interativa”, através de um sistema de recomendação de anúncios que busca a exibição de publicidade baseada nos interesses pessoais do usuário (telespectador) por meio da navegação no aplicativo para a TV Digital Interativa.

Esta avaliação do software se faz sob a responsabilidade do pesquisador Alan Menk, do Curso de Mestrado em Gestão de Redes e Telecomunicações – Centro de Ciências e Exatas, Ambientais e de Tecnologias da PUCC - Pontifícia Universidade Católica de Campinas. O objetivo do trabalho é exibir publicidade baseada nos dados obtidos de forma explícita e implícita, através da navegação em aplicativos Ginga (notícias). A partir dos dados obtidos é feito um processo de descoberta de conhecimento – (KDD - Knowledge Discovery in Database) utilizando o software de mineração de dados Weka. Considera-se este estudo importante para a sociedade atual, pois permitirá tornar a interatividade com a TV Digital mais atraente aos telespectadores e rentável às difusoras e anunciantes.

Todos os procedimentos serão explicados detalhadamente antes da realização do mesmo, sendo o seu envolvimento nesse trabalho é voluntário, sendo-lhe garantido que os seus dados pessoais serão mantidos em sigilo e nunca serão divulgados. Os resultados obtidos na avaliação serão utilizados apenas para alcançar o objetivo do trabalho exposto acima, incluída sua publicação na literatura científica especializada.

Para o experimento, o primeiro campo a ser preenchido pelo usuário é o Cadastro do Telespectador, iniciando pelo login e senha e alguns dados pessoais necessários para o processo de recomendação, quais sejam: idade, sexo, se tem filhos, veículos, tipo de residência e salário.

A participação nessa pesquisa não lhe trará qualquer prejuízo ou benefício financeiro ou profissional e, se desejar, a sua exclusão do grupo de avaliação poderá ser solicitada em qualquer momento. Informo ainda que o termo será feito em duas vias, sendo uma para o participante e outra para o pesquisador.

Caso concorde dar o seu consentimento livre e esclarecido para participar do projeto de pesquisa supracitado, assine o seu nome abaixo e responda ao questionário.

Atenciosamente, Alan Menk

Data: ___/10/2012

Assinatura do Participante

Assinatura do Pesquisador

APÊNDICE C – Planilha de Frequência dos usuários

Controle de Frequência							
	Login	1º Dia	2º Dia	3º Dia	4º Dia	5º Dia	Recomend.
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							

APÊNDICE D – Selects da tarefa de classificação do algoritmo J48.

Perfil 01

```
select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Telespectador.idade='Adolescente'
and Log.Categoria='Empregos'
```

Perfil 02

```
select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Telespectador.idade='Adulto'
and EstadoCivil.Descricao='Viuvo'
and Log.Categoria='Saude'
```

Perfil 03

```
select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Telespectador.idade='Jovem'
and EstadoCivil.Descricao='Casado'
and Sexo='M'
and Log.Categoria='Carros'
```

Perfil 04

```
select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Telespectador.idade='Jovem'
and EstadoCivil.Descricao='Solteiro'
and Sexo='M'
and Telespectador.Veiculo='Sim'
and Log.Categoria='Carros'
```

Perfil 05

```
select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Telespectador.idade='Jovem'
and EstadoCivil.Descricao='Solteiro'
```

and Sexo='F'
 and Telespectador.Veiculo='Sim'
 and Log.Categoria='Carros'

Perfil 06

```
SELECT Telespectador.idTelespectador, Log.categoria FROM Log, Telespectador, EstadoCivil,
FaixaSalarial, StatusResidencia WHERE EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
AND FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
AND StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
AND Log.idTelespectador = Telespectador.idTelespectador
AND Telespectador.idade = 'Adulto'
AND EstadoCivil.Descricao = 'Divorciado'
AND Log.Categoria = 'Educacao'
```

Perfil 07

```
select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Telespectador.idade='Jovens'
and EstadoCivil.Descricao='Viuvos'
and Telespectador.sexo='M'
and Log.Categoria='Games'
```

Perfil 08

```
select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Telespectador.idade='Jovem'
and EstadoCivil.Descricao='Solteiro'
and Telespectador.sexo='M'
and FaixaSalarial.descricao<>'ClasseA'
and Log.Categoria='Games'
```

Perfil 09

```
select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Telespectador.idade='Jovem'
and EstadoCivil.Descricao='Divorciado'
and Telespectador.sexo='M'
and Log.Categoria='Economia'
```

Perfil 10

```
select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
```


and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
 and Log.idTelespectador = Telespectador.idTelespectador
 and Telespectador.idade='Jovem'
 and EstadoCivil.Descricao<>'Solteiro'
 and Telespectador.sexo='F'
 and Log.Categoria='CasaDecoracao'

Perfil 11

select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
 FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
 and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
 and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
 and Log.idTelespectador = Telespectador.idTelespectador
 and Telespectador.idade='Jovem'
 and EstadoCivil.Descricao='Solteiro'
 and Telespectador.sexo='F'
 and Telespectador.veiculo='Nao'
 and Log.Categoria='CasaDecoracao'

Perfil 12

select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
 FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
 and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
 and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
 and Log.idTelespectador = Telespectador.idTelespectador
 and Telespectador.idade='Adulto'
 and EstadoCivil.Descricao='Divorciado'
 and Log.Categoria='Educacao'

Perfil 13

select Telespectador.idTelespectador, Log.categoria from Log, Telespectador, EstadoCivil,
 FaixaSalarial, StatusResidencia where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
 and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
 and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
 and Log.idTelespectador = Telespectador.idTelespectador
 and Telespectador.idade='Adulto'
 and EstadoCivil.Descricao='Solteiro'
 and Log.Categoria='Carros'

APÊNDICE E – Resultado da aplicação do algoritmo J48

```

=== Run information ===
Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: Resultado
Instances: 162
Attributes: 8
    Descricao
    Sexo
    Filhos
    Veiculo
    Idade
    EstadoCivil
    ClasseSocial
    Residencia
Test mode:evaluate on training data

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

J48 pruned tree
-----
Idade = Adolescente: Empregos (0.0)
Idade = Jovem
  Sexo = M
    EstadoCivil = solteiro
      ClasseSocial = ClasseB: Games (0.0)
      ClasseSocial = ClasseE
        Veiculo = Sim: Carros (9.0/3.0)
        Veiculo = Nao: Esportes (12.0/6.0)
      ClasseSocial = ClasseD: Games (18.0/6.0)
      ClasseSocial = ClasseC: Games (0.0)
    EstadoCivil = Casado: Carros (6.0)
    EstadoCivil = Divorciado: Economia (6.0)
    EstadoCivil = Viuvo: Games (0.0)
  Sexo = F
    EstadoCivil = solteiro
      Veiculo = Sim: Carros (6.0)
      Veiculo = Nao: CasaDecoracao (3.0)
    EstadoCivil = Casado: CasaDecoracao (9.0)
    EstadoCivil = Divorciado: CasaDecoracao (0.0)
    EstadoCivil = Viuvo: CasaDecoracao (0.0)
Idade = Adulto
  EstadoCivil = solteiro: Carros (3.0)
  EstadoCivil = Casado: Empregos (75.0/42.0)
  EstadoCivil = Divorciado: Educacao (6.0)
  EstadoCivil = Viuvo: Saude (9.0)

Number of Leaves :    18
Size of the tree :    26
Time taken to build model: 0.02 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      105          64.8148 %
Incorrectly Classified Instances    57           35.1852 %
Kappa statistic                    0.5841
Mean absolute error                 0.077
Root mean squared error             0.1963
Relative absolute error              52.4522 %
Root relative squared error         72.5245 %
Total Number of Instances          162

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          1         0         0.875     1         0.933     0.997   Carros
          1         0         1         1         1         1       CasaDecoracao
          0         0         0         0         0         ?       Ciencia
          1         0         1         1         1         1       Economia
          0.5       0         1         0.5       0.667     0.885   Educacao
          1         0.326   0.44     1         0.611     0.837   Empregos
          0.333     0.042   0.5      0.333    0.4       0.844   Esportes
          0.5       0.043   0.667   0.5      0.571     0.91    Games
          0         0         0         0         0         0.779   Politica
          0.5       0         1         0.5       0.667     0.885   Saude
          0         0         0         0         0         0.779   Tecnologia
          0         0         0         0         0         0.779   Viagem
weighted Avg.  0.648   0.08    0.654    0.648    0.609    0.89

=== Confusion Matrix ===

 a b c d e f g h i j k l <-- classified as
21 0 0 0 0 0 0 0 0 0 0 0 | a = Carros
 0 12 0 0 0 0 0 0 0 0 0 0 | b = CasaDecoracao
 0 0 0 0 0 0 0 0 0 0 0 0 | c = Ciencia
 0 0 0 6 0 0 0 0 0 0 0 0 | d = Economia
 0 0 0 0 6 6 0 0 0 0 0 0 | e = Educacao
 0 0 0 0 0 33 0 0 0 0 0 0 | f = Empregos
 0 0 0 0 0 6 6 6 0 0 0 0 | g = Esportes
 3 0 0 0 0 3 6 12 0 0 0 0 | h = Games
 0 0 0 0 0 6 0 0 0 0 0 0 | i = Politica
 0 0 0 0 0 9 0 0 0 0 9 0 | j = Saude
 0 0 0 0 0 6 0 0 0 0 0 0 | k = Tecnologia
 0 0 0 0 0 6 0 0 0 0 0 0 | l = Viagem

```

APÊNDICE F – Selects da tarefa de agrupamento do algoritmo K-Means.

Cluster0

```
select Telespectador.idTelespectador, Log.categoria
from Log, Telespectador, EstadoCivil, FaixaSalarial, StatusResidencia
where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Log.Categoria='Empregos' and Telespectador.sexo='M'
and Telespectador.Filhos='Nao'
and Telespectador.Veiculo='Sim'
and Telespectador.idade='Adulto'
and EstadoCivil.Descricao='Casado'
and FaixaSalarial.Descricao='ClasseD'
and StatusResidencia.Descricao='Propria'
```

Cluster 1

```
select Telespectador.idTelespectador, Log.categoria
from Log, Telespectador, EstadoCivil, FaixaSalarial, StatusResidencia
where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Log.Categoria='Saude' and Telespectador.sexo='F'
and Telespectador.Filhos='Sim'
and Telespectador.Veiculo='Nao'
and Telespectador.idade='Adulto'
and EstadoCivil.Descricao='Viuvo'
and FaixaSalarial.Descricao='ClasseC'
and StatusResidencia.Descricao='Financiada'
```

Cluster 2

```
select Telespectador.idTelespectador, Log.categoria
from Log, Telespectador, EstadoCivil, FaixaSalarial, StatusResidencia
where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Log.Categoria='Carros' and Telespectador.sexo='M'
and Telespectador.Filhos='Nao'
and Telespectador.Veiculo='Sim'
and Telespectador.idade='Jovem'
and EstadoCivil.Descricao='Solteiro'
and FaixaSalarial.Descricao='ClasseE'
and StatusResidencia.Descricao='Propria'
```

Cluster 3

```
select Telespectador.idTelespectador, Log.categoria
from Log, Telespectador, EstadoCivil, FaixaSalarial, StatusResidencia
where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
```

and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
 and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
 and Log.idTelespectador = Telespectador.idTelespectador
 and Log.Categoria='Esportes' and Telespectador.sexo='M'
 and Telespectador.Filhos='Nao'
 and Telespectador.Veiculo='Nao'
 and Telespectador.idade='Jovem'
 and EstadoCivil.Descricao='Solteiro'
 and FaixaSalarial.Descricao='ClasseE'
 and StatusResidencia.Descricao='Propria'

Cluster 4

select Telespectador.idTelespectador, Log.categoria
 from Log, Telespectador, EstadoCivil, FaixaSalarial, StatusResidencia
 where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
 and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
 and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
 and Log.idTelespectador = Telespectador.idTelespectador
 and Log.Categoria='Games' and Telespectador.sexo='M'
 and Telespectador.Filhos='Nao'
 and Telespectador.Veiculo='Sim'
 and Telespectador.idade='Jovem'
 and EstadoCivil.Descricao='Solteiro'
 and FaixaSalarial.Descricao='ClasseD'
 and StatusResidencia.Descricao='Propria'

Cluster 5

select Telespectador.idTelespectador, Log.categoria
 from Log, Telespectador, EstadoCivil, FaixaSalarial, StatusResidencia
 where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
 and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
 and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
 and Log.idTelespectador = Telespectador.idTelespectador
 and Log.Categoria='Empregos' and Telespectador.sexo='M'
 and Telespectador.Filhos='Sim'
 and Telespectador.Veiculo='Sim'
 and Telespectador.idade='Adulto'
 and EstadoCivil.Descricao='Casado'
 and FaixaSalarial.Descricao='ClasseD'
 and StatusResidencia.Descricao='Propria'

Cluster 6

select Telespectador.idTelespectador, Log.categoria
 from Log, Telespectador, EstadoCivil, FaixaSalarial, StatusResidencia
 where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
 and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
 and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
 and Log.idTelespectador = Telespectador.idTelespectador
 and Log.Categoria='Carros' and Telespectador.sexo='M'
 and Telespectador.Filhos='Nao'
 and Telespectador.Veiculo='Sim'
 and Telespectador.idade='Jovem'

and EstadoCivil.Descricao='Casado'
 and FaixaSalarial.Descricao='ClasseD'
 and StatusResidencia.Descricao='Alugada'

Cluster 7

```
select Telespectador.idTelespectador, Log.categoria
from Log, Telespectador, EstadoCivil, FaixaSalarial, StatusResidencia
where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Log.Categoria='Educacao' and Telespectador.sexo='F'
and Telespectador.Filhos='Sim'
and Telespectador.Veiculo='Nao'
and Telespectador.idade='Adulto'
and EstadoCivil.Descricao='Divorciado'
and FaixaSalarial.Descricao='ClasseD'
and StatusResidencia.Descricao='Propria'
```

Cluster 8

```
select Telespectador.idTelespectador, Log.categoria
from Log, Telespectador, EstadoCivil, FaixaSalarial, StatusResidencia
where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Log.Categoria='Empregos' and Telespectador.sexo='F'
and Telespectador.Filhos='Nao'
and Telespectador.Veiculo='Sim'
and Telespectador.idade='Adulto'
and EstadoCivil.Descricao='Casado'
and FaixaSalarial.Descricao='ClasseD'
and StatusResidencia.Descricao='Alugada'
```

Cluster 9

```
select Telespectador.idTelespectador, Log.categoria
from Log, Telespectador, EstadoCivil, FaixaSalarial, StatusResidencia
where EstadoCivil.idEstadoCivil = Telespectador.idEstadoCivil
and FaixaSalarial.idFaixaSalarial = Telespectador.idFaixaSalarial
and StatusResidencia.idStatusResidencia = Telespectador.idStatusResidencia
and Log.idTelespectador = Telespectador.idTelespectador
and Log.Categoria='Saude' and Telespectador.sexo='F'
and Telespectador.Filhos='Nao'
and Telespectador.Veiculo='Sim'
and Telespectador.idade='Adulto'
and EstadoCivil.Descricao='Casado'
and FaixaSalarial.Descricao='ClasseD'
and StatusResidencia.Descricao='Alugada'
```

APÊNDICE G – Resultado da aplicação do algoritmo KMeans.

```

=== Model information ===
Filename:      ModeloCluster.model
Scheme:weka.clusterers.SimpleKMeans -N 10 -A "weka.core.ManhattanDistance -R first-last" -I 300 -S 5
Relation:Resultado
Attributes:8
      Descricao
      Sexo
      Filhos
      Veiculo
      Idade
      EstadoCivil
      ClasseSocial
      Residencia

=== Clustering model ===

kMeans
=====
Number of iterations: 3
Sum of within cluster distances: 135.0
Missing values globally replaced with mean/mode

Cluster centroids:
Attribute      Full Data      Cluster#
              (162)          0
              (9)           1
              (39)          2
              (18)          3
              (15)          4
              (27)          5
              (18)          6
              (12)          7
              (6)           8
              (9)           9
              (9)
=====
Descricao      Empregos      Saude      Empregos      Carros      Esportes      Games      Empregos      Carros      Educacao      Empregos      Saude
Sexo           M             F          M             M           M             M           M             M           F             F             F
Filhos        Nao          Sim        Nao          Nao         Nao          Nao         Sim          Nao         Sim          Nao          Nao
Veiculo       Sim          Nao       Sim          Sim         Sim          Sim         Sim          Sim          Nao         Sim          Sim
Idade         Adulto       Adulto    Adulto       Jovem      Jovem      Jovem      Adulto       Jovem      Adulto       Adulto       Adulto
EstadoCivil   Casado      Viuvo     Casado      Solteiro   Solteiro   Solteiro   Casado      Casado   Divorciado   Casado       Casado
ClasseSocial  ClasseD     ClasseC   ClasseC     ClasseE    ClasseE    ClasseD   ClasseD     ClasseD   ClasseD     ClasseD     ClasseD
Residencia    Propria     Financiada Propria     Propria    Propria    Propria    Propria     Propria   Alugada     Propria     Alugada   Alugada

```

APÊNDICE H - Formulário de avaliação da publicidade.**FORMULÁRIO DE AVALIAÇÃO DA PUBLICIDADE APRESENTADA –
APLICATIVO SMART MARKETING**

Login: _____

**Considere o quanto a indicação apresentada se adéqua a sua pessoa.
Considera-se aqui que quanto mais adequado a recomendação maior será a
relevância para você. Assinale com X.**

Publicidade 01

	Extremamente Relevante
	Relevante
	Pouco Relevante
	Irrelevante
	Inadequada

Publicidade 02

	Extremamente Relevante
	Relevante
	Pouco Relevante
	Irrelevante
	Inadequada

Publicidade 03

	Extremamente Relevante
	Relevante
	Pouco Relevante
	Irrelevante
	Inadequada