

## Seleção de Técnicas de Mineração de Dados para Segmentação de Mercado

Gabriel G.M.S. de Magalhães<sup>1</sup>, Fernanda Lima<sup>2</sup>, Eduardo F. Santos<sup>1</sup>, Paulo Junior<sup>1</sup>,  
Lucas Rosa<sup>1</sup>

<sup>1</sup>Agência Espacial Brasileira – AEB  
SPO Quadra 5, bloco Q – Brasília – DF – Brazil

<sup>2</sup>Universidade de Brasília – UnB  
Departamento de Ciência da Computação – CIC  
Campus Universitário Darcy Ribeiro – Brasília/DF – Brasil, CEP 70910-900

`gabrielgmsm@gmail.com, ferlima@cic.unb.br`

`{eduardo.santos, paulo.junior, lucas.rosa}@aeb.gov.br`

**Abstract.** *Market segmentation process is used to gather information about consumer to help specific marketing strategies. As the Brazilian Space Agency is a public organization which depends on investments, it is necessary to know the interested citizen profile, making possible to find out if the investments on the Space Program are getting to their knowledge. There are works about data mining on market segmentation, but they are spread around different sources. This research seeks a possible classification for data mining techniques in market segmentation, using systematic review approach. This review analysis presents a classification proposal for market segmentation data mining techniques to be applied in the profiles of the Brazilian Space Program interest population.*

**Resumo.** *O processo de segmentação de mercado é utilizado para obter conhecimento sobre os consumidores e ajudar na construção de estratégias de marketing mais específicas. Sendo a Agência Espacial Brasileira uma organização pública dependente de investimento, é preciso conhecer o perfil do cidadão interessado para validar se os investimentos realizados no Programa Espacial estão chegando a seu conhecimento. Existem trabalhos que tratam da aplicação de mineração de dados para processos de segmentação de mercado, porém, estão espalhados pela literatura. O trabalho busca, de maneira sistemática, encontrar uma possível classificação das técnicas de mineração de dados que vêm sido empregadas para processos de segmentação de mercado. A análise das técnicas encontradas apresenta como resultado uma proposta de classificação das técnicas de mineração de dados para segmentação de mercado a ser aplicada ao perfil da população de interesse do Programa Espacial Brasileiro.*

### 1. Introdução

A Agência Espacial Brasileira (AEB) administra a política espacial brasileira, como o desenvolvimento de planos de veículos lançadores, aplicações de satélites e centros de lançamento. De tudo o que é produzido, constata-se que a população brasileira, em geral, não conhece a atuação ou compreende a relevância para o país do que a AEB produz. É necessário, então, que a agência busque novas formas de aproximação com a

população, com um processo de divulgação eficaz e um fortalecimento de sua imagem [Rollemberg et al. 2010].

No intuito de avaliar qual o perfil da população de interesse para a AEB e validar se os investimentos realizados no Programa Espacial estão chegando a seu conhecimento, um processo de segmentação de mercado pode ser explorado. Tal processo pode aproveitar-se de dados disponíveis na Internet, com soluções de mineração de dados para a extração e análise destes dados, de modo a revelar quais são características comuns de cada grupo encontrado.

Segmentação de mercado trata da busca por novas oportunidades ao se conhecer com mais propriedade os clientes de determinado negócio [Beane and Ennis 1994]. Tanto em órgãos privados quanto em órgãos públicos essa estratégia tem sido utilizada e, apesar da diferença entre os objetivos, para ambos existem aplicações de segmentação de mercado que são benéficas [Capon 1981].

Para se realizar o processo de segmentação de mercado é possível aplicar técnicas de mineração de dados, que tratam de extrair conhecimento de uma base de dados, por meio de algoritmos e análises probabilísticas [Liao et al. 2012]. Se faz necessário, então, entender o processo e conhecer quais são as técnicas e os principais trabalhos existentes na literatura que darão base para a realização de um processo de segmentação de mercado.

O presente estudo busca identificar, de maneira sistemática, a partir de revisões publicadas nos últimos seis anos, quais são as principais técnicas de mineração de dados que podem ser aplicadas à segmentação de mercado. Na segunda seção, apresenta uma revisão dos principais conceitos de segmentação de mercado e mineração de dados. Na terceira seção, propõe uma metodologia para a busca de trabalhos e classificação das técnicas; na quarta seção, apresenta o processo de pesquisa; na quinta seção os resultados alcançados e por fim conclui o trabalho com sugestões futuras.

## **2. Referencial Teórico**

A seção apresenta as técnicas de segmentação de mercado e mineração de dados a serem aplicadas no processo de identificação do perfil da população que o Programa Espacial Brasileiro tenta atingir.

### **2.1. Segmentação de Mercado**

A segmentação de mercado trata do estudo e posterior agrupamento de indivíduos de acordo com características em comum para uma abordagem de marketing direcionada a cada grupo identificado [Smith 1956]. Segundo Beane e Ennis [Beane and Ennis 1994], que fizeram uma revisão dos principais trabalhos a respeito de segmentação de mercado, as principais motivações para uma segmentação de mercado são:

1. Procurar por novas oportunidades para produtos ou áreas que poderão ser receptivas a um reposicionamento de produtos;
2. Criar mensagens publicitárias aperfeiçoadas ao se obter um melhor entendimento de seus clientes.

A segmentação ocorre de acordo com uma escolha de variáveis prévias, que podem ser de cunho (1) geográfico; (2) demográfico, como classe social, raça, gênero e outros; (3) psicográfico, em que se abordam características intrapessoais do consumidor ou

(4) comportamental, relacionadas a atitudes do consumidor em relação ao produto. Existem autores que trazem também segmentação por imagem, em que a variável é na verdade uma imagem que o consumidor possui do produto ou da marca. A segmentação comportamental pode ainda ser dividida como: (1) ocasião de compra, a segmentação pelas razões que estimulam o consumidor a utilizar o produto; (2) benefício que o consumidor enxerga em utilizar determinado produto, como prestígio, economia ou conveniência; (3) incidência de uso, trata das razões ou ocasiões que levam o consumidor a escolher determinado produto, pode ser considerada como uma extensão da segmentação por ocasião de compra e benefício; (4) status do usuário em relação ao produto, por exemplo: não usuário, usuário de primeira viagem ou usuário assíduo; (5) taxa de uso, similarmente à segmentação por status, esta se difere ao não incluir usuários que não utilizam o produto; (6) lealdade, segmentação que busca identificar o grau de lealdade de consumidores à marca/produto; (7) prontidão para compra, o quão inclinado determinado cliente está para realizar a compra, alguns clientes conhecem o produto outros estão ansiosos por comprá-lo e (8) fatores de mercado, como os clientes reagem a estímulos do mercado, como diminuição do preço, ofertas e ações de marketing [Beane and Ennis 1994].

Um processo de segmentação de mercado busca, então, agrupar os consumidores em uma ou mais dessas formas apresentadas, podendo ser *a priori*, em que os segmentos já são escolhidos previamente, ou *post hoc*, em que os segmentos são identificados após análise dos dados [Beane and Ennis 1994].

## 2.2. Mineração de Dados

O processo de mineração de dados envolve a extração de conhecimento a partir de uma grande quantidade de dados. Esse tipo de segmentação pode ser realizada sobre dados estruturados, não-estruturados ou semi-estruturados na busca por padrões, a partir da utilização de conhecimentos interdisciplinares nas áreas de estatística, computação e demais áreas ligadas à aplicação do processo de mineração, como Marketing [Liao et al. 2012, Mukhopadhyay et al. 2014, Ziafat and Shakeri 2014].

O modelo chamado CRISP-DM, traz as fases fundamentais de um processo de mineração de dados, independente da indústria em que é aplicada. Essas fases são: o entendimento do negócio; o entendimento dos dados; a preparação dos dados; a criação de um modelo; a avaliação do modelo e por fim a implementação do que foi construído [Wirth 2000]. Ziafat [Ziafat and Shakeri 2014] cita uma metodologia para aplicar à Segmentação de Mercado o modelo (CRISP-DM):

1. **Entender o negócio e desenhar o processo de segmentação.** Nesta etapa, são realizadas reuniões entre profissionais de marketing e cientistas de dados para um levantamento de requisitos e posterior definição de estratégias para a segmentação. Esse passo envolve, então: definir objetivos de negócio, selecionar os critérios apropriados para a segmentação, determinar a população e o nível de segmentação.
2. **Entender, preparar e enriquecer os dados.** Nesta etapa, as diversas fontes são investigadas, definindo-se quais dados serão utilizados, realiza-se a integração, agregação, validação e limpeza dos dados, assim como a transformação, redução e enriquecimento dos mesmos.
3. **Identificação dos segmentos com modelagem por *cluster*.** Nesta etapa, diversos modelos de clusterização devem ser testados na busca por distintos *clusters* que

separem os clientes. Vale lembrar que cada modelo utilizado produzirá segmentos distintos, o que não significa que estejam errados. Os resultados não serão idênticos, mas similares, convergindo para alguns segmentos. A análise envolve então a avaliação do nível de concordância entre os modelos.

4. **Avaliação e definição do perfil de cada segmento revelado.** Após a modelagem, os resultados que mais se adequam às necessidades do negócio são escolhidos para implementação. As tarefas envolvem avaliação técnica das soluções de clusterização, desenho de perfil dos *clusters* através de modelos supervisionados, utilização de pesquisas de mercado para enriquecer os segmentos comportamentais e etiquetá-los adequadamente.
5. **Implementação da solução de segmentação, projeto e aplicação das estratégias.** Neste passo, as soluções construídas são implementadas dentro da empresa para a construção de estratégias de marketing diferenciadas. Três tarefas são fundamentais neste processo: construir um modelo de pontuação para atualizar os segmentos, construir uma árvore de decisão para pontuar os segmentos e distribuir as informações dos segmentos.

### 3. Metodologia

A seguir é apresentada a metodologia utilizada para selecionar os trabalhos de aplicação de mineração de dados para segmentação de mercado. Esta metodologia envolve uma inicial seleção de técnicas de mineração de dados a partir da escolha de revisões sistemáticas e uma posterior classificação de trabalhos de aplicação destas técnicas em segmentação de mercado.

#### 3.1. Seleção de Técnicas de Mineração de Dados

Para selecionar as técnicas de mineração de dados, buscando cumprir os itens 3 e 4 da lista apresentada na seção 2.2, foi realizada uma pesquisa de revisões sistemáticas de mineração de dados que pudessem estar relacionadas às fases de um processo de segmentação de mercado.

Na busca por revisões sistemáticas, já havia uma revisão que cobria o período entre 2000 e 2011 [Liao et al. 2012]. Assim, iniciou-se o trabalho com uma pesquisa manual, a partir dos termos: “Data Mining Review” que cobrisse o período de 2011 a 2017. As pesquisas são feitas pelo Google Scholar, que atravessa diversas bases de dados científicos e possui a funcionalidade de classificação por relevância. Os trabalhos são escolhidos a partir de uma análise do título e abstract, selecionando-se trabalhos que indicam se tratar de uma revisão sistemática sobre as técnicas de Mineração de Dados. Após uma seleção inicial, seguindo a metodologia apresentada por Kitchenham et. al. [Kitchenham et al. 2009], foram elaboradas cinco perguntas para garantia de qualidade dos trabalhos selecionados em alcançar os objetivos de pesquisa:

- Q1. É uma revisão que trata de técnicas para Mineração de Dados?
- Q2. É classificada como A1 ou A2?
- Q3. Os critérios de inclusão e exclusão utilizados foram descritos e são apropriados?
- Q4. Os trabalhos/estudos básicos foram adequadamente descritos?
- Q5. O trabalho pode ter aplicação em alguma parte do processo de segmentação de mercado?

A resposta para a Q2 indica qual a classificação do trabalho. Para as outras questões, a resposta pode ser 'S', quando o pesquisador acredita que a resposta é sim, 'P', quando acredita-se que responde parcialmente (50%) e 'N', quando a resposta é negativa.

Uma vez escolhidas as revisões, o modelo de classificação das técnicas é definido de modo a encontrar uma classificação que englobe grande parte dos trabalhos e onde cada trabalho é categorizado em apenas um tópico.

### 3.2. Seleção de Trabalhos de Segmentação de Mercado

Ao todo, 7 técnicas foram extraídas da etapa anterior, que variam desde aplicações de redes neurais a clusterização. Para cada técnica, foram avaliados quais são os trabalhos de segmentação de mercado que utilizam mineração de dados, no período de 2011 a 2017. Para cada técnica é feita uma pesquisa manual, no Google Scholar, com os termos: “nome da técnica” + “market segmentation” e seguindo os seguintes critérios de categorização dos trabalhos:

- Q6. Vem de uma fonte A1 ou A2?
- Q7. Está relacionado à Segmentação de Mercado?
- Q8. Qual é a área de aplicação do processo de segmentação de mercado?
- Q9. Qual(is) técnica(s) o trabalho aborda?
- Q10. Qual(is) algoritmo(s) o trabalho apresenta?

A resposta para a pergunta Q6 é a classificação Qualis. Para a Q7 a resposta deve ser: Geográfica, Demográfica, Psicográfica, Comportamental ou por Imagem, quando o trabalho envolve um caso de uso de segmentação de mercado; 'P' quando não possui um estudo de caso específico, mas apresenta uma técnica de mineração de dados passível de ser utilizada para segmentação de mercado ou 'N' quando não se relaciona a segmentação de mercado. A Q8 é respondida identificando a área de aplicação do processo de segmentação, *e.g.* Turismo, Alimentação, Varejo ou outros. A Q9 é respondida tendo em vista as técnicas que foram selecionadas anteriormente para a classificação e, por fim, a Q10 com os algoritmos que o autor identificou no trabalho.

## 4. Desenvolvimento

Da pesquisa manual realizada na primeira fase, 300 trabalhos foram avaliados. Após análise inicial, 37 artigos foram escolhidos, dos quais 6 eram de fonte A1; 10 eram B ou C e o restante não possui qualificação. Logo em seguida, foram selecionadas 4 revisões sistemáticas que estavam de acordo com os critérios estabelecidos.

### 4.1. Análise das Revisões Sistemáticas

O trabalho de Park et. al. [Park et al. 2012] traz uma revisão de sistemas de recomendação, que tratam de verificar a probabilidade de um cliente realizar uma compra para recomendá-lo outros produtos. Os autores escolheram artigos publicados em periódicos de 2000 a 2010 e os classificaram de acordo com os campos de atuação e as técnicas de mineração de dados. Ao todo, oito técnicas de mineração de dados foram escolhidas para a classificação: (1) Regra de Associação; (2) *Clustering*; (3) Árvore de decisão; (4) K-vizinhos próximos; (5) Redes neurais; (6) Análise de Links; (7) Regressão e (8) Outros métodos heurísticos.

O artigo de Liao et. al. [Liao et al. 2012] apresenta uma revisão de 2000 a 2011 das principais técnicas e suas aplicações de mineração de dados, dividindo em tipo de conhecimento, análise e arquitetura. As categorias são: (1) Redes Neurais; (2) Arquitetura de Algoritmos; (3) *Dynamic Prediction-Based*; (4) Análise de Arquitetura de Sistema; (5) *Intelligence Agents Systems*; (6) Modelagem; (7) Sistemas baseados em conhecimento; (8) Otimização de Sistemas e (9) Sistemas da Informação. O autor utilizou índices de palavras chave e dos *abstracts* para categorizar as técnicas, de acordo com seu conhecimento no assunto. Nenhum trabalho posterior com este formato foi encontrado.

O trabalho de Tsytsarau e Palpanas [Tsytsarau and Palpanas 2012] apresenta uma revisão voltada à análise de sentimento e mineração de Opiniões. Tal tema relaciona-se indiretamente com segmentação de mercado uma vez que a segmentação de mercado pode ser realizada através da análise do comportamento de clientes em relação a um produto. Os autores trazem, em uma tabela, uma classificação de tendências no ramo de mineração de opiniões divididos em: algoritmo utilizado, *dataset* para os testes e nos domínios alvo. Os temas identificados foram: (1) *Rule Based*; (2) Dicionário; (3) Estatística; (4) *Naive Bayes*; (5) Máxima Entropia; (6) *Support Vector Machine*; (7) *Natural Language Processing*; (8) *Latent Semantic Analysis*; (9) *Pointwise Mutual Information*; (10) Semântico; (11) *Conditional Markov Models*; (12) Classificador Linear; (13) ML; (14) SVM OVA; (15) Multi SVM; (16) SVR; (17) Grafos; (18) HMM; (19) DT; (20) ADT; (21) Estatísticos SO-LSA e (22) Corpus.

Outro trabalho que trata de uma revisão relacionada à mineração de opinião é o apresentado por Nassirtoussi et. al. [Khadjeh Nassirtoussi et al. 2014] em que é apresentada uma revisão dos trabalhos que tratam de mineração textual para previsões de mercado. Os autores dividem os algoritmos em seis classes: (1) Support Vector Machine; (2) Algoritmos de Regressão; (3) Naïve Bayes; (4) Árvores ou Regras de Decisão; (5) Algoritmos Combinatórios e (6) Experimentos de Multi-Algoritmos.

## 4.2. Escolha das Técnicas

Tendo em vista os trabalhos escolhidos, a escolha das técnicas foi realizada buscando-se uma classificação que fosse ao mesmo tempo única, no sentido de que uma aplicação só se encaixa em uma classificação, e abrangente, ao se considerar o maior número de técnicas de mineração de dados. O trabalho de Park et. al. [Park et al. 2012], que trouxe uma classificação para sistemas de recomendação, foi então o escolhido, excluindo-se a classificação “Outros métodos heurísticos”, que não é específica. O trabalho de Liao et. al. [Liao et al. 2012] é bastante abrangente, porém possui ambiguidades ao se escolher em qual classificação encaixar determinada técnica e os outros dois trabalhos trazem uma classificação das técnicas específica para mineração textual.

## 4.3. Pesquisa de Trabalhos de Segmentação de Mercado

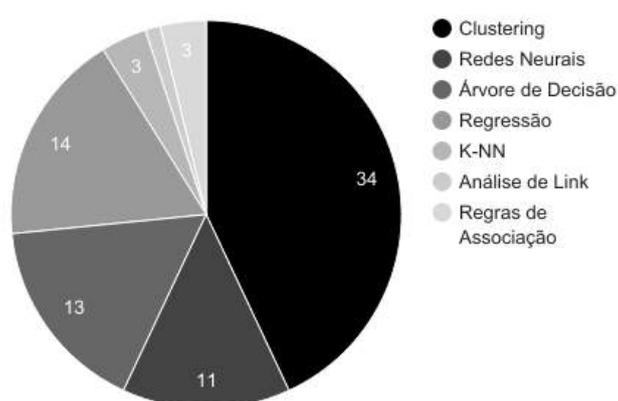
Ao todo, 350 resultados foram avaliados, sendo 50 para cada busca manual da técnica. Desses, 230 foram selecionados para posterior avaliação. Ao avaliar o Qualis, foram identificados 97 resultados como A1 ou A2; 15 como B; 18 como C e o restante não possuía classificação Qualis.

## 5. Resultados

O trabalho se iniciou com uma pesquisa por revisões sistemáticas, das quais apenas 4 estavam de acordo com os critérios estabelecidos. Percebe-se que, apesar da diversidade do tema de Mineração de Dados, poucos trabalhos atuais e de fontes renomadas trazem um agrupamento sistemático das técnicas de mineração sendo pesquisadas. Escolher um modelo de classificação envolve um profundo e vasto conhecimento na área, dada a diversidade de aplicações [Liao et al. 2012].

O trabalho de escolha dos artigos da etapa seguinte está sumarizado na tabela do Anexo I. Esta tabela apresenta os principais trabalhos de segmentação de mercado que utilizam de técnicas de mineração de dados. Cada artigo foi avaliado tendo em vista a aplicação da técnica e quais algoritmos ou modelos de mineração de dados foram empregados. Ao todo, 57 artigos foram selecionados que estavam de acordo com os critérios estipulados na metodologia. Desses, 59.65% foram identificados como relacionados a técnicas de Clusterização; 28.07% de Redes Neurais; 22.81% de Árvore de Decisão, 5.26% de K-NN; 5.26% Regras de Associação e 1.75% de Análise de Links, como mostra o gráfico na figura 1.

As técnicas de K-NN, Análise de Link e Regras de Associação foram as que apresentaram menos trabalhos relacionados à segmentação de mercado propriamente dita. Os trabalhos nessas técnicas se referem em sua maioria a sistemas de recomendação, que podem ser utilizados para processos relacionados a segmentação de mercado, por exemplo, ao avaliar os resultados de segmentação para recomendar produtos aos segmentos identificados. Logo, essas três categorias não devem ser incluídas em uma classificação específica para segmentação de mercado, mas podem ser utilizadas em conjunto a depender do contexto.



**Figura 1. Distribuição por Técnica**

Foram identificados, nestes trabalhos, 67 algoritmos ou modelos, sendo que 26.32% dos trabalhos tratavam do algoritmo de K-Means; 14.04% do algoritmo de SOM;

7.02% de Apriori; 7.02% de CHAID; 5.26% de K-NN e o restante de demais algoritmos ou modelos. Em relação às áreas de aplicação, 17.54% dos trabalhos traziam segmentação na área de Turismo; 12.28% na área de Telecomunicações; 8.77% de e-commerce; 10.53% de trabalhos com técnicas que podem ser aplicadas em diversas áreas; 5.26% de segmentação de mercados internacionais e o restante em outras áreas.

Tendo em vista que as técnicas de Clusterização e aplicação de Redes Neurais a partir dos algoritmos de K-Means e Self Organizing Maps (SOM) foram as com maior número de trabalhos com resultados concretos encontradas, é possível escolher tais técnicas como base para o projeto de identificação do perfil de interesse para a AEB. A segmentação de mercado será tanto demográfica, ao se buscar as regiões de maior interesse quanto comportamental e psicográfica, identificando o quanto cada cidadão está familiarizado com os conceitos da Política Espacial Brasileira e qual o estilo de vida do cidadão, mais especificamente, quais são seus hábitos na internet e de onde o mesmo procura informações relacionadas às atividades do Programa Espacial Brasileiro. Com esse conhecimento em mãos, será possível estipular estratégias de marketing específicas para cada grupo encontrado.

## 6. Conclusão

Este trabalho apresenta uma proposta de classificação para técnicas de mineração de dados aplicáveis à segmentação de mercado, a partir de uma metodologia que leva em conta classificações existentes para a escolha de modelos. Ao apresentar a proposta também lança os fundamentos para a criação de uma ação permanente de monitoramento da opinião do cidadão a respeito das realizações da AEB no âmbito do Programa Espacial Brasileiro. Como o principal objetivo de tal ação é descobrir se as realizações na agência no âmbito do Programa Espacial estão sendo capazes de alcançar o cidadão, conhecer o perfil da população a ser atingido é etapa importante do processo.

Assim, imagina-se que o presente trabalho pode ser explorado como referência para a definição das técnicas a serem utilizadas em um projeto de segmentação de mercado que busca evidenciar as principais características demográficas, comportamentais e psicográficas do público interessado por temas do Programa Espacial Brasileiro. Com a utilização das técnicas encontradas para a identificação dos segmentos, as ações de marketing poderão ser direcionadas, uma vez que cada segmento possui características próprias que justificam o teor e os meios de se transmitir informação de maneira eficiente, assim como será possível obter acompanhamento se a mensagem do Programa Espacial está chegando ao cidadão de forma adequada.

O tema de mineração de dados é extenso, o que dificulta a tarefa de propor uma classificação capaz de agrupar todas as técnicas adequadamente. Nesse sentido, o presente trabalho avaliou revisões da literatura já existentes para propor uma classificação específica para o caso de segmentação de mercado. Do total de revisões selecionadas, poucas estavam de acordo com os critérios estabelecidos, deixando-se assim de avaliar possíveis trabalhos que poderiam trazer novas formas de classificação.

Como os trabalhos para cada técnica classificada foram escolhidos a partir de pesquisa manual de palavras chave com os nomes das técnicas, deixaram de ser avaliados trabalhos relevantes que poderiam acrescentar novas formas de classificação a partir de técnicas que não foram elicitadas. Outro limitador foi a escolha de trabalhos apenas na

língua inglesa.

Os próximos passos da pesquisa envolvem a seleção das fontes a serem utilizadas e uma realização de um pré-experimento para avaliar a viabilidade de se utilizar os dados adquiridos. Uma vez que os dados estejam disponíveis, será possível obter os segmentos a partir da clusterização *K-Means* e pela análise da topologia gerada com os dados de entrada em um *SOM*, seguindo a metodologia CRISP-DM. De posse do modelo teórico, será possível construir um painel que permita aos gestores da AEB monitorar, em tempo real, o alcance dos projetos relacionados ao Programa Espacial.

### Agradecimentos

Esta pesquisa foi financiada pela Agência Espacial Brasileira (AEB) em conjunto com o Centro de Pesquisa em Arquitetura da Informação da Universidade de Brasília (CPAI-UnB) através do Projeto de Arquitetura da Informação para o Programa Espacial Brasileiro 2014-2017.

### Referências

- Beane, T. P. and Ennis, D. M. (1994). Market Segmentation : A Review. *European Journal of Marketing*, 21(5):20–42.
- Capon, N. (1981). Marketing Strategy Differences Between State and Privately Owned Corporations: An Exploratory Analysis. *Journal of Marketing*, 45(2):11–18.
- Khadjeh Nassirtoussi, A., Aghabozorgi, S., Ying Wah, T., and Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications*, 41(16):7653–7670.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1):7–15.
- Liao, S.-h., Chu, P.-h., and Hsiao, P.-y. (2012). Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems With Applications*, 39(12):11303–11311.
- Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., and Coello, C. A. C. (2014). A survey of multiobjective evolutionary algorithms for data mining: Part i. *IEEE Transactions on Evolutionary Computation*, 18(1):4–19.
- Park, D. H., Kim, H. K., Choi, I. Y., and Kim, J. K. (2012). A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39(11):10059–10072.
- Rollemborg, R., Veloso, E. M., Queiroz Filho, A. P. d., Lins, B. F. E., Silva, E. F., Rocha, F. C. W., Faria, F. F., Juras, I. d. A. G., Menck, J. T. M., Camino, M. E. M. B., Vasconcelos, R. D. d. L., Martins, R. C. d. R., and Filho, R. d. M. G. (2010). A Política Espacial Brasileira.
- Smith, W. R. (1956). Product Differentiation and Market Segmentation As Alternative Marketing Strategies. *Journal of Marketing*, 21(1):3–8.
- Tsytarau, M. and Palpanas, T. (2012). Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3):478–514.

Wirth, R. (2000). CRISP-DM : Towards a Standard Process Model for Data Mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, (24959):29–39.

Ziafat, H. and Shakeri, M. (2014). Using Data Mining Techniques in Customer Segmentation. *International Journal of Engineering Research and Applications*, 4(9):70–79.

## Anexo I

A Tabela 1 a seguir apresenta todos trabalhos que foram identificados seguindo-se a metodologia proposta no presente artigo.

**Tabela 1. Classificação dos Trabalhos de Mineração de Dados para Segmentação de Mercado**

| Título   | Ano  | Técnicas                                    | Qualis | Market Segmentation                                 | Área             | Algoritmos e Técnicas   |
|--|------|---|--------|---|------------------|---|
| A case study of applying LRFM model in market segmentation of a children's dental clinic                                 | 2012 | Clustering, Redes Neurais                   | A1     | Demográfica, Comportamental (Lealdade)              | Saúde            | SOM   |
| A consumer-oriented segmentation study in the Swiss wine market  | 2011 | Clustering                                  | A2     | Comportamental, Imagem, Psicográfica                | Vinho            | Ward's Method   |
| A Decision Support System for market-driven product positioning and design   | 2015 | Clustering, Árvore de Decisão               | A1     | Comportamental (uso do produto)                     | Automotivo       | Principal Component Analysis (PCA), K-means, AdaBoost classification                            |
| A hybrid case-GA-based decision support model for warehouse operation in fulfilling cross-border orders                  | 2012 | Árvore de Decisão                           | A1     | P   | Armazenamento    | C-GADS  |
| A hybrid model of self-organizing maps (SOM) and least square support vector machine (LSSVM) for time-series forecasting | 2011 | Redes Neurais, Regressão                    | A1     | P   | Geral            | SOM, LSSVM  |
| A meta-learning approach for determining the number of clusters with consideration of nearest neighbors                  | 2013 | Clustering, KNN                             | A1     | P   | Geral            |   |
| A novel customer scoring model to encourage the use of mobile value added services                                       | 2011 | Redes Neurais, Regressão, Árvore de Decisão | A1     | Comportamental (Propensão ao uso)                   | Telecomunicações | Exploratory Data Analysis (EDA)   |
| A psycho-cognitive segmentation of organ donors in Egypt using Kohonen's self-organizing maps                            | 2011 | Redes Neurais                               | A1     | Psicográfica  | Saúde            | SOM   |
| A topic-based recommender system for electronic marketplace platforms  | 2012 | Análise de Links                            | A2     | P   | E-Marketing      | Latent Dirichlet Allocation   |
| A unified framework for market segmentation and its applications   | 2012 | Clustering, Redes Neurais, Regressão        | A1     | P   | Geral            |   |
| An application of particle swarm optimization algorithm to clustering analysis   | 2011 | Clustering, Redes Neurais                   | A2     | P   | Geral            | Particle Swarm Optimization (PSO), K-Means  |
| Application of particle swarm optimization and perceptual map to tourist market segmentation                             | 2012 | Clustering                                  | A1     | Demográfica, Comportamental (Propósito de compra)   | Turismo          | Particle Swarm Optimization (PSO), PSKO, K-Means, MDS   |
| Audience targeting by B-to-B advertisement classification: A neural network approach                                     | 2013 | Redes Neurais                               | A1     | P   | Geral            | FeedForward network, MLP (multilayer perceptron), Text-Mining, Sentiment Analysis, Topic Mining |
| Bagged Clustering and its application to tourism market segmentation   | 2013 | Clustering                                  | A1     | Demográfica, Comportamental (Motivação para compra) | Turismo          | Bagged Clustering (BC), K-means, fuzzy C-medoids Clustering (FCMdC)                             |

| Título  | Ano  | Técnicas                                    | Qualis | Market Segmentation                              | Área                    | Algoritmos e Técnicas   |
|---|------|---|--------|--|-------------------------|---|
| Bicycle commuting market segmentation analysis using attitudinal factors  | 2012 | Clustering                                  | A1     | Comportamental                                   | Transporte              | SEM, K-Means  |
| Building comprehensible customer churn prediction models with advanced rule induction techniques                                      | 2011 | Árvore de Decisão, Regressão                | A1     | P  | Geral                   | Ant Colony Optimization (ACO), ALBA, Ant Miner+, SVM, C4.5                                    |
| Category role aided market segmentation approach to convenience store chain category management                                       | 2014 | Clustering                                  | A1     | Demográfica, Comportamental (Padrões de compra)  | Varejo                  | weighted fuzzy K-means clustering algorithm (WFKM)  |
| Competitive strategies and market segmentation for suppliers with substitutable products  | 2011 | Clustering                                  | A1     | Comportamental (Lealdade)                        | Armazenamento           | Game Theory   |
| Consumer preferences for wine applying best-worst scaling: a Spanish case study   | 2012 | Clustering                                  | A2     | Demográfica, Comportamental (Razões para compra) | Vinho                   | BW  |
| Customer data mining for lifestyle segmentation   | 2012 | Clustering                                  | A1     | Psicográfica                                     | Varejo                  | VARCLUS   |
| Customer relationship management in the hairdressing industry: An application of data mining techniques                               | 2013 | Clustering                                  | A1     | Comportamental (Lealdade)                        | Cabelereiro             | SOM, K-Means  |
| Design of convergent product concepts based on functionality: An association rule mining and decision tree approach                   | 2012 | Regra de Associação, Árvore de Decisão      | A1     | P  | Produtos de Mídia       | Text Mining   |
| Discriminative clustering for market segmentation   | 2012 | Clustering, Regressão                       | A1     | Comportamental (Utilização do Produto)           | Site de Notícia         | SVM, Regularized Least Squares Regression (RLSR)  |
| Dynamics in the international market segmentation of new product growth   | 2012 | Clustering                                  | A1     | Demográfica                                      | Mercados Internacionais | Semi-Parametric HiddenMarkovmodel (HMM), p-splines, Stochastic Expectation-Maximization (SEM) |
| Event segmentation: A review and research agenda  | 2011 | Regressão, Clustering, Regra de Associação  | A1     | Geral  | Eventos                 |   |
| Experimental comparison of parametric, non-parametric, and hybrid multigroup classification   | 2012 | K-NN, Redes Neurais, Regressão              | A1     | Demográfica                                      | Investimentos           | K-NN, Linear Programming (LP), Multivariate Discriminant Analysis (MDA)                       |
| Exploring the use of travel information - identifying contextual market segmentation in Seoul, Korea                                  | 2011 | Árvore de Decisão                           | A1     | Demográfica, Comportamental (Tipo de Atividade)  | Turismo                 | Chi-square Automatic Interaction Detection (CHAID)  |
| External validity of market segmentation methods: A study of buyers of prestige cosmetic brands                                       | 2012 | Clustering, Regressão                       | A1     | Comportamental (Benefícios)                      | Cosméticos              | Mixture Regression, K-Means   |
| Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques | 2011 | Regressão, Árvore de Decisão, Redes Neurais | A1     | Demográfica, Comportamental (Uso do produto)     | Telecomunicações        | Genetic Algorithms (GA), SAVG, MVOTE, GAOW  |
| How to design and utilize online customer center to support new product concept generation  | 2011 | Árvore de Decisão, Regra de Associação      | A1     | Comportamental (Uso do Produto)                  | Telecomunicações        | Text Mining, K-Means  |
| Identity, demographics, and consumer behaviors: International market segmentation across product categories                           | 2011 | Clustering, Regressão                       | A1     | Demográfica, Psicográfica                        | Mercados Internacionais | SEM, Stepwise Multiple Linear Regressions (MLRs).   |
| Improved market segmentation by fuzzifying crisp clusters: A case study of the energy market in Spain                                 | 2014 | Clustering                                  | A1     | Psicográfica                                     | Energia                 | LAMDA   |
| International market selection and segmentation: perspectives and challenges  | 2011 | Clustering                                  | A1     | Demográfica, Psicográfica                        | Mercados Internacionais |   |
| Key drivers of airline loyalty  | 2011 | Árvore de Decisão                           | A2     | Comportamental (Lealdade) e Imagem               | Aviação                 |   |
| Market segmentation and ideal point identification for new product design using fuzzy data compression and fuzzy clustering methods   | 2012 | Clustering, Redes Neurais                   | A1     | Comportamental (Uso do Produto)                  | Produtos de Mídia       | Fuzzy Data Compression Method, Fuzzy Clustering Method, NPCA                                  |
| Market segmentation and willingness to pay for green electricity among urban residents in China: The case of Jiangsu Province         | 2012 | Regressão                                   | A2     | Demográfica                                      | Energia                 | Mlogit Model  |
| Midpoint radius self-organizing maps for interval-valued data with telecommunications application                                     | 2011 | Clustering, Redes Neurais                   | A1     | Psicográfica                                     | Telecomunicações        | MR-SOM  |
| Mining customer knowledge for direct selling and marketing  | 2011 | Regras de Associação, Clustering            | A1     | Psicográfica, Comportamental (Motivações)        | Cosméticos              | K-Means, Apriori  |
| Mining customer knowledge to implement online shopping and home delivery for hypermarkets   | 2011 | Regras de Associação                        | A1     | Demográfica, Comportamental (Produtos)           | E-commerce              | Apriori, K-Means  |

| <b>Título</b>  | <b>Ano</b> | <b>Técnicas</b>                      | <b>Qualis</b> | <b>Market Segmentation</b>                     | <b>Área</b>            | <b>Algoritmos e Técnicas</b>  |
|--|------------|--------------------------------------|---------------|--|------------------------|---|
| Mobile customer segmentation based on smartphone measurement   | 2014       | Clustering                           | A2            | Demográfica, Psicográfica                      | Telecomunicações       | Latent Class Analysis (LCA)   |
| Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium | 2014       | Regras de Associação                 | A2            | Comportamental (Uso do Produto)                | Turismo                | Apriori   |
| Preference-based clustering reviews for augmenting e-commerce recommendation   | 2013       | Regressão, Clustering, K-NN          | A2            | Comportamental (Review de Produto)             | E-commerce             | Latent Class Regression Model (LCRM), Probabilistic Regression Model (PRM), K-Means, K-NN |
| Recommending inexperienced products via learning from consumer reviews   | 2012       | Clustering, K-NN                     | A2            | Comportamental (Review de Produto)             | E-commerce             | Latent Class Regression Model (LCRM), K-Means, K-NN                                       |
| Segmentation of telecom customers based on customer value by decision tree model   | 2012       | Árvore de Decisão                    | A1            | Comportamental (Lealdade)                      | Telecomunicações       | Classification and Regression tree (CART), C5.0 Decision Tree, AHP                        |
| Segmenting customers in online stores based on factors that affect the customer's intention to purchase                                  | 2012       | Clustering, Redes Neurais            | A1            | Psicográfica                                   | E-commerce             | SOM, K-Means  |
| Segmenting visitors of cultural events by motivation: A sequential non-linear clustering analysis of Italian Christmas market visitors   | 2012       | Clustering, Redes Neurais            | A1            | Demográfica, Comportamental (Motivação)        | Turismo                | SOM   |
| Stability of market segmentation with cluster analysis—A methodological approach   | 2014       | Clustering                           | A2            | Psicográfica                                   | Alimentos              | K-Means, Hierarchical Ward Method   |
| To mine association rules of customer values via a data mining procedure with improved model: An empirical case study                    | 2011       | Regra de Associação                  | A1            | Comportamental (Benefício)                     | E-commerce             | Ward's Method, RFMDR, Apriori   |
| Tourist attitudes towards an accommodation tax earmarked for environmental protection: A survey in the Algarve                           | 2012       | Árvore de Decisão                    | A2            | Comportamental (Disposicao a pagar)            | Turismo                | Chi-square Automatic Interaction Detection (CHAID)  |
| Understanding China's long-haul outbound travel market: An overlapped segmentation approach  | 2013       | Clustering                           | A1            | Psicográfica                                   | Turismo                | Canonical Correlation Analysis (CCA)  |
| Understanding Japanese tourists' shopping preferences using the Decision Tree Analysis method  | 2011       | Árvore de Decisão                    | A2            | Demográfica, Comportamental (Motivação)        | Turismo                | Chi-square Automatic Interaction Detection (CHAID)  |
| Using a back propagation network combined with grey clustering to forecast policyholder decision to purchase investment-linked insurance | 2011       | Redes Neurais, Clustering, Regressão | A1            | Demográfica, Psicográfica (Propensao a risco)  | Investimentos          | Grey Clustering Statistic, Logistic Regression, Back Propagation NN                       |
| Using the Taguchi method for effective market segmentation   | 2012       | Clustering, Redes Neurais            | A1            | Demográfica, Comportamental (Benefícios)       | Alimentos              | Taguchi Method, K-Means, SOM, Genetic Algorithms (GA)                                     |
| Variety-seeking: Using the CHAID segmentation approach in analyzing the international traveler market                                    | 2015       | Árvore de Decisão                    | A2            | Comportamental                                 | Turismo                | Chi-square Automatic Interaction Detection (CHAID)  |
| Visualizing market segmentation using self-organizing maps and Fuzzy Delphi method - ADSL market of a telecommunication company          | 2011       | Clustering, Redes Neurais            | A1            | Demográfica, Comportamental (Propósito de Uso) | Telecomunicações       | Fuzzy Delphi Method, SOM  |
| Water conservation behavior in Australia   | 2012       | Árvore de Decisão, Regressão         | A1            | Comportamental (Uso do Produto)                | Proteção Meio Ambiente | Multivariate Linear Regression Model, Unbiased Recursive Partitioning                     |