

3

Metodologia para Segmentação do Mercado Bancário

Este capítulo descreve a metodologia proposta nesta dissertação para a segmentação do mercado bancário a partir da abordagem *post-hoc*, servindo-se de ferramentas de análise de dados, como modelos de agrupamento, método automático para geração de regras *fuzzy* e sistema de inferência *fuzzy* para a segmentação, caracterização e classificação do mercado bancário.

A metodologia incorpora ao processo de segmentação informações relacionadas a aspectos psicológicos, como o estilo de vida, permitindo uma maior compreensão dos clientes e, conseqüentemente, uma maior adaptação e diferenciação do composto de marketing da instituição. A metodologia divide-se em quatro etapas: definição dos macros segmentos, coleta de dados, definição dos segmentos e classificação do mercado. A Figura 9 apresenta as quatro etapas.



Figura 9: Macro etapas para a segmentação do mercado bancário.

Nas seções seguintes são descritas todas as etapas que compõem a metodologia proposta.

3.1

Definição dos Macros Segmentos

Corresponde à primeira etapa da metodologia, sendo composta por apenas uma tarefa, conforme mostra a Figura 10.

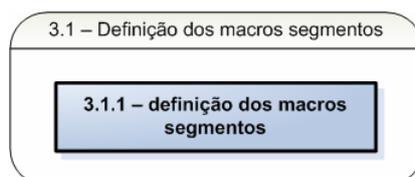


Figura 10: Tarefas que compõe a etapa de Definição dos macros segmentos.

3.1.1

Definição dos macros segmentos

Algumas instituições trabalham com diferentes estruturas de mercado, como varejo e atacado, o que, na maioria das vezes, exige uma abordagem diferenciada de segmentação para cada estrutura. A separação em estruturas de mercado corresponde à definição dos macros segmentos. Como exemplo pode-se citar o mercado de telefonia que divide o mercado em dois segmentos: corporativo (empresas) e varejo (consumidores finais).

3.2

Coleta de Dados

Corresponde à segunda etapa da metodologia, sendo dividida em duas tarefas: seleção das variáveis e levantamento dos dados. A Figura 11 descreve as tarefas que compõem esta etapa.

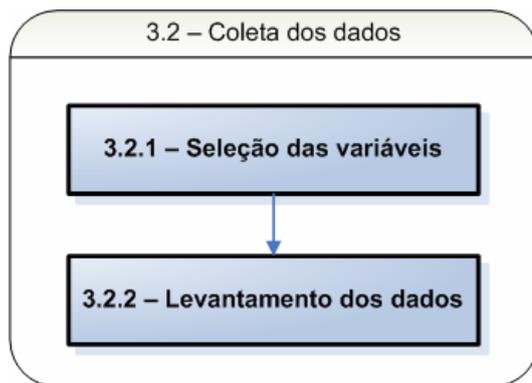


Figura 11: Tarefas que compõe a etapa de Coleta de dados.

3.2.1

Seleção das variáveis

Corresponde à seleção das variáveis aplicadas no modelo. Estas variáveis podem estar disponíveis nas bases, como as variáveis demográficas e socioeconômicas, ou demandar uma pesquisa, como as relacionadas a aspectos psicológicos, normalmente não disponíveis nas bases cadastrais.

Depois de selecionadas as variáveis que irão compor o estudo, estas são classificadas em duas classes: aquelas utilizadas em conjunto com a análise de agrupamento para formar os segmentos, e aquelas utilizadas para ajudar a descrever os segmentos depois de construídos.

3.2.2

Levantamento dos dados

A partir da seleção e classificação das variáveis inicia-se o levantamento dos dados necessários ao estudo, que envolve a coleta de informação das bases cadastrais e pesquisa com clientes.

A pesquisa com clientes pode ser dividida em duas fases: planejamento e execução. A fase de planejamento corresponde à elaboração do projeto de pesquisa, que abrange desde a definição do propósito, objetivos primários e secundários e técnica utilizada a composição da amostra, dentre outros. A fase de

execução trata da implementação do projeto de pesquisa definido na fase anterior. Normalmente, neste tipo de levantamento, relacionado a aspectos psicológicos, são realizados dois tipos de pesquisa de maneira seqüencial: qualitativa e quantitativa.

A primeira pesquisa, qualitativa ou exploratória, objetiva apurar todos os possíveis valores, atitudes, interesses e opiniões dos consumidores. Esta pesquisa subsidia a construção dos questionários da pesquisa quantitativa, permitindo maior aderência do questionário ao público alvo e, conseqüentemente, melhores resultados. As técnicas mais utilizadas para este tipo de pesquisa são: grupo focal, entrevistas em profundidade e observação de campo.

A segunda, quantitativa, fornece informações representativas da população, ou mercado a ser estudado. As técnicas mais utilizadas para este caso são: entrevistas ao telefone, internet, e-mail e envio de mala direta.

3.3

Definição dos Segmentos

Corresponde à terceira etapa da metodologia e descreve o processo interativo de geração de resultados a partir do modelo de agrupamento *fuzzy*, Fuzzy C-Means (FCM). Esta interação ocorre a partir da combinação dos diferentes parâmetros de entrada do modelo, gerando, para cada combinação, um resultado diferente. O objetivo é buscar a combinação de parâmetros que forneça os melhores resultados quanto ao arranjo dos agrupamentos.

A etapa de Definição dos segmentos pode ser dividida em: parametrização do modelo, tratamento dos dados, agrupamento dos dados, consolidação dos resultados, caracterização dos segmentos e análise dos clientes em transição entre segmentos. A Figura 12 descreve as tarefas que compõem esta etapa.

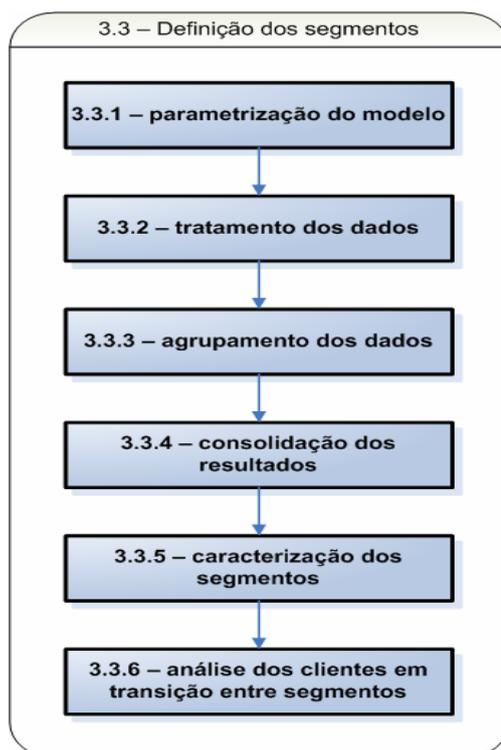


Figura 12: Tarefas que compõe a etapa de Definição dos segmentos.

3.3.1

Parametrização do modelo

O modelo requer que três parâmetros sejam definidos na sua inicialização, sendo eles: número de grupos, coeficiente *fuzzy* e critérios de parada. O número de grupos está associado ao número de segmentos que se deseja estabelecer para o estudo. O coeficiente *fuzzy* m corresponde à medida de nebulosidade que deve existir na definição dos grupos. Por fim, o critério de parada do algoritmo, que pode ser definido como: número máximo de interações, incremento na função objetivo, ou ambos.

3.3.2

Tratamento dos dados

Corresponde ao pré-processamento dos dados e envolve duas ações: tratamento de dados duplicados, *outliers*, valores nulos ou inválidos e normalização dos dados. Existe uma infinidade de métodos para a normalização

dos dados (Kaufman, Rousseeuw, 1990), mas para este estudo, será descrita, pela Equação 32 e 33, a normalização *z-score*, mais utilizada em problemas de segmentação (MatLab Reference 2005).

$$z_{if} = \frac{x_{if} - \bar{x}}{std_f} \quad (32)$$

$$std_f = \sqrt{\frac{1}{1-n} \{(x_{1f} - \bar{x}_f)^2 + (x_{2f} - \bar{x}_f)^2 + \dots + (x_{nf} - \bar{x}_f)^2\}} \quad (33)$$

3.3.3

Agrupamento dos dados

Equivale ao processamento dos dados, que envolve desde a inicialização do modelo, passando por sua execução e tratamento de suas respostas. Na inicialização são definidos os parâmetros e o modelo é executado, apresentando como resposta uma matriz com os graus de pertinência de cada elemento para com os grupos definidos pelo algoritmo. A partir desta matriz, os elementos são classificados nos grupos que possuem o maior valor de pertinência.

Para avaliar o grau de distância do elemento para com os demais se utiliza a matriz de dissimilaridade. Esta matriz apresenta a dimensão $n \times n$, onde n corresponde ao total de objetos do conjunto de dados, e cada elemento contém uma medida de dissimilaridade, ou seja, uma medida de diferença, de um objeto para outro (Vale 2005). O modelo utiliza a Equação 34 para o cálculo da dissimilaridade em variáveis escalares, e a Equação 35 para as nominais. Ao final, somam-se as dissimilaridades obtidas a partir de todas as variáveis, obtendo uma medida para dissimilaridade de dois objetos de um mesmo grupo (Kaufman, Rousseeuw 1990).

$$d_{ij}^{(f)} = \frac{|x_{if} - x_{nf}|}{\max_n x_{nf} - \min_n x_{nf}} \quad (34)$$

$$d_{ij}^{(f)} \begin{cases} 1, x_{if} = x_{nf} \\ 1, x_{if} \neq x_{nf} \end{cases} \quad (35)$$

Onde $d_{ij}(f)$ constitui a dissimilaridade do objeto x_{if} para os demais x_{nf} objetos do grupo f descoberto.

O valor de silhueta mede a qualidade de um agrupamento (Kaufman, Rousseeuw 1990). Sendo A o agrupamento ao qual o objeto i pertence, a dissimilaridade média do objeto i em relação a todos os outros objetos de A é dada pela Equação 36.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (36)$$

Onde $|A|$ representa o total de objetos presentes no agrupamento e $d(i, j)$ representa a dissimilaridade entre os objetos i e j . Para qualquer agrupamento C diferente de A , a dissimilaridade média do objeto i para todos os objetos de C será dado pela Equação 37.

$$d(i, C) = \frac{1}{|C| - 1} \sum_{j \in C, j \neq i} d(i, j) \quad (37)$$

Onde $|C|$ representa o total de objetos presentes no agrupamento e $d(i, j)$ representa a dissimilaridade entre os objetos i e j . A menor distância de dissimilaridade entre o objeto i a um dado agrupamento A será dada pela Equação 38.

$$b(i) = \min_{C \neq A} d(i, C) \quad (38)$$

Considere-se como B o agrupamento C que contém a menor distância acima. Esse agrupamento é chamado de vizinho do objeto i e é o segundo melhor agrupamento para este objeto. Com efeito, o valor de silhueta do objeto i é definido pela Equação 39.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (39)$$

Onde, para valores próximos de 1, têm-se objetos bem agrupados; para valores próximos de 0 o objeto está dividido entre dois agrupamentos, e para valores próximos de -1 o objeto está mal classificado.

3.3.4

Consolidação dos resultados

A consolidação dos resultados consiste no cálculo do valor de silhueta médio dos grupos de cada interação, dada pela combinação dos parâmetros de entrada do modelo. Os dados da combinação que apresenta os melhores resultados são armazenados para posterior análise de seus agrupamentos.

3.3.5

Caracterização dos segmentos

Nesta etapa os dados relativos à combinação que apresentou os melhores resultados são analisados, e os seus grupos caracterizados a partir das variáveis disponíveis no estudo. A partir desta caracterização os grupos são classificados quanto à importância financeira para a instituição.

3.3.6

Análise dos clientes em transição entre segmentos

Os clientes que apresentam valores de pertinência abaixo de 0,5 nos segmentos em que foram classificados são analisados, revelando que aspectos os distanciam do grupo atual. Esta análise permite o desenho de ações de marketing específicas para a correção destes desvios e conseqüente aumento de receita.

3.4

Classificação do Mercado

Após a segmentação e caracterização da amostra de clientes, inicia-se a classificação do mercado com base nos segmentos definidos pela amostra. Esta etapa é composta por duas tarefas relacionadas ao desenvolvimento do SIF, conforme apresenta a Figura 13.

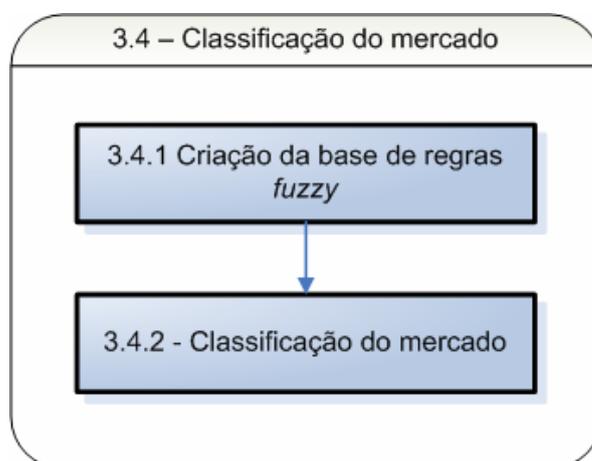


Figura 13: Tarefas que compõe a etapa de Caracterização dos segmentos.

3.4.1

Criação da base de regras

O desenvolvimento do sistema de inferência *fuzzy* inicia-se com a construção de sua base de regras. Para isto é necessário priorizar as variáveis utilizadas no estudo para que estas não gerem um número de regras que dificulte a

sua interpretação. A priorização das variáveis mais relevantes para este estudo utiliza o método ANFIS, descrito em seções anteriores. Após a priorização, são selecionadas as variáveis mais relevantes e seus domínios divididos em conjuntos *fuzzy*. Em seguida, são extraídas automaticamente regras do tipo SE-ENTÃO, por meio do software Fuzzy Rules 2001 (Vale, 2001), que implementa o método do Mendel de extração de regras *fuzzy*.

3.4.2

Classificação do mercado

As regras geradas na etapa anterior são aproveitadas na construção da base de regras de SIF o qual é aplicado em toda a base de clientes, classificando-os nos segmentos definidos. O próximo capítulo apresenta a aplicação da metodologia descrita neste capítulo ao Banco da Amazônia, de forma a avaliar o desempenho de todo o processo.