

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ
ESCOLA DE NEGÓCIOS
PROGRAMA DE PÓS-GRADUAÇÃO GESTÃO DE COOPERATIVAS -
PPGCOOP

ELTON CESAR FREIRE

ANÁLISE DA SEGMENTAÇÃO DINÂMICA DE BASE DE ASSOCIADOS EM
COOPERATIVA DE CRÉDITO

Curitiba
2017

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ
ESCOLA DE NEGÓCIOS
PROGRAMA DE PÓS-GRADUAÇÃO GESTÃO DE COOPERATIVAS -
PPGCOOP

ANÁLISE DA SEGMENTAÇÃO DINÂMICA DE BASE DE ASSOCIADOS EM
COOPERATIVA DE CRÉDITO

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Gestão de Cooperativas da Pontifícia Universidade Católica do Paraná.

Orientador: Prof. Dr. Alex Antonio Ferraresi
Co-Orientador: Prof. Dr. Edson Emílio Scalabrin

Curitiba
2017

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central
Giovanna Carolina Massaneiro dos Santos – CRB 9/1911

F866a Freire, Elton Cesar
2017 Análise da segmentação dinâmica de base de associados em cooperativa de crédito / Elton Cesar Freire ; orientador: Alex Antonio Ferraresi ; coorientador, Edson Emílio Scalabrin . – 2017.
90 f. : il. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2017

Bibliografia: f. 78-82

1. Cooperativas de crédito. 2. Cooperativismo. 3. Segmentação de mercado. 4. Algoritmos. I. Ferraresi, Alex Antonio. I. Scalabrin, Edson Emílio. II. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Gestão de Cooperativas. III. Título.

CDD 20. ed. – 334.2

TERMO DE APROVAÇÃO

ANALISE DA SEGMENTAÇÃO DINÂMICA DE BASE DE ASSOCIADOS EM COOPERATIVA DE CRÉDITO

Por


ELTON CESAR FREIRE

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Gestão de Cooperativas, área de concentração em Gestão de Cooperativas, da Escola de Negócios da Pontifícia Universidade Católica do Paraná.

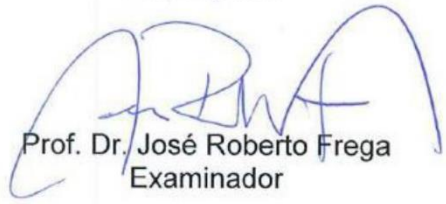


Prof. Dr. Alex Antonio Ferraresi
Coordenador do Programa de Pós-Graduação em Gestão de Cooperativas

Prof. Dr. Alex Antonio Ferraresi
Orientador



Prof. Dr. Vilmar Rodrigues Moreira
Examinador



Prof. Dr. José Roberto Frega
Examinador

AGRADECIMENTOS

Evidentemente não poderia deixar de agradecer a todos que de uma forma ou de outra participaram desta empreitada de minha vida, mas em especial:

A minha amada Esposa Fernanda, que sempre me apoiou em tudo. Sou eternamente grato a Deus por ela estar ao meu lado, assim como meus filhos Beatriz e Otávio que entenderam minha ausência em alguns momentos.

A minha mãe e minha avó Maria, que inundadas de sabedoria divina me instruíram pelos caminhos da vida.

Ao Sicoob Norte PR, a essa cooperativa sou eternamente grato por possibilitar as minhas experiências profissionais. Em especial ao Emerson que brilhantemente conduziu os rumos deste empreendimento.

Ao professor Dr. Alex Ferraresi, que esteve ao meu lado em todos os momentos da construção deste trabalho, mesmo nos momentos mais desesperadores.

Ao professor Dr. Scalabrin e o Sr. Luis Boiko, sem eles este trabalho seria apenas um sonho.

Ao Padre Bruno Athíla, amigo de todas as horas.

RESUMO

O sistema de crédito cooperativista possui muitos adeptos no Brasil. São mais de 6 milhões de pessoas utilizando os serviços financeiros prestados por este ramo do cooperativismo. Embora com uma demanda crescente, o cooperativismo de crédito possui apenas 2,44% de participação nos ativos do mercado financeiro como um todo, com base nos dados apresentados pelo Banco Central do Brasil. O mercado financeiro brasileiro possui uma concentração, em relação aos ativos, de 90% em apenas 8 bancos. A concentração bancária ocorreu principalmente após a implantação do plano real. Este período foi positivo ao cooperativismo de crédito já que as cooperativas cresceram em número saindo de 806 para 1.311 singulares, em apenas 10 anos no fim do século XX. Para este estudo foram consideradas as informações da cooperativa “Alfa”, que é uma cooperativa de crédito do norte do Paraná, fundada pelos comerciantes desta região e atualmente aderente ao Sistema de Cooperativas do Brasil, utilizando assim o Bancoob como acesso ao sistema de compensação bancária. Muito jovem, essa cooperativa com pouco mais de 10 anos de exercício já ultrapassaria a barreira de meio bilhão de ativos, e mais de 19 mil associados. São muitos associados para o gerenciamento em tão pouco tempo, e assim a cooperativa, tendo em vista a necessidade de adequação do atendimento, visa um novo posicionamento estratégico. A segmentação da base real de associados da cooperativa Alfa, evidenciada a partir do embasamento teórico das práticas de posicionamento estratégico de *marketing*, do gerenciamento de relação com o cliente, e ainda a segmentação de clientes, objetiva administrar o contexto em que esta instituição financeira está inserida. As técnicas de análise utilizadas foram: análise documental, análises estatísticas e matemáticas do banco de dados dos cooperados, bem como a técnica de árvore de decisão criada através do algoritmo C 4.5 implementado pelo software WEKA. A partir das identificações de padrões e aspectos preditivos, pretende-se associar essas informações aos possíveis pacotes de serviços personalizados para grupos distintos de cooperados.

Palavras Chave: Cooperativismo, Cooperativismo de Crédito, Segmentação, Algoritmo C 4.5.

ABSTRACT

The cooperative credit system has many adherents in Brazil, there are more than 6 million people, using the financial services provided by this branch of cooperativism. Although with a growing demand, credit cooperativism has only 2.44% participation in financial market assets as a whole, based on the data presented by the Central Bank of Brazil. The Brazilian financial market has a concentration, in relation to assets, of 90% in only 8 banks. The banking concentration occurred mainly after the implementation of the real plan. This period was positive for credit cooperatives as cooperatives grew in numbers from 806 to 1,311 in 10 years at the end of the 20th century. For this study we considered the information of the "Alfa" cooperative, which is a credit cooperative in the north of Paraná, founded by the merchants of this region and currently adherent to the Brazilian Cooperative System, thus using Bancoob as access to the bank clearing system. Very young, this cooperative with after 10 years of exercise would already surpass the barrier of half a billion assets, and more than 19 thousand associates. They are many associates for the management in such a short time, and thus the cooperative, in view of the need of adequacy of the service, aims at a new strategic positioning. The segmentation of the actual base of associates of the Alfa cooperative, evidenced from the theoretical basis of the practices of strategic marketing positioning, customer relationship management, and customer segmentation, aims to manage the context in which this financial institution is inserted. The analysis techniques used were: documentary analysis, statistical and mathematical analyzes of the cooperative database, as well as the decision tree technique created through the algorithm C 4.5 implemented by WEKA software. From the identification of patterns and predictive aspects, it is intended to associate this information with possible packages of personalized services for different groups of members.

Key Words: Cooperativism, Credit Cooperativism, Segmentation, Algorithm C 4.5.

LISTA DE GRÁFICOS

GRÁFICO 1 – Evolução Cooperativas Singulares no Brasil	15
GRÁFICO 2 – Número de Membros do Cooperativismo de Crédito no Brasil	15
GRÁFICO 3 – Distribuição Percentual de Associados Cooperativa Alfa	55
GRÁFICO 4 – Distribuição de crédito tomado por modalidade nos associados PF	57
GRÁFICO 5 – Distribuição de crédito tomado por modalidade nos associados PJ	58

LISTA DE FIGURAS

FIGURA 1 – Distribuição das Cooperativas de Crédito por Regiões no Brasil	21
FIGURA 2 – Modelo de Decisão e atividades da Segmentação	23
FIGURA 3 – Modelo de Segmentação do Banco Itaú	26
FIGURA 4 – Modelo de Segmentação do Banco Bradesco	26
FIGURA 5 – Modelo de Segmentação do Banco Santander	27
FIGURA 6 – Modelo de Árvore de Decisão	36
FIGURA 7 – Algoritmo de Criação de Árvore de Decisão	38

LISTA DE TABELA

TABELA 1 – Modalidades de Segmentação	24
TABELA 2 – Fórmulas para Classificar no Formato de Árvore de Decisão	39

LISTA DE QUADROS

QUADRO 1 – Funcionamento do Algoritmo k-Means	34
QUADRO 2 – Modelo de Árvore de Decisão em Regras	36
QUADRO 3 – Conjunto de Teste Previamente Ordenado	37
QUADRO 4 – Atributos selecionados na base de dados	50
QUADRO 5 – Objetivo de Aplicação da Técnica de Análise de Conglomerado	51
QUADRO 6 – Objetivo de aplicação da técnica de Análise de Correspondência	52
QUADRO 7 – Relatório de Desempenho Classificação do C 4.5 – Pessoa Física	52
QUADRO 8 – Relatório de Desempenho Classificação do C 4.5 – Pessoa Jurídica	53
QUADRO 9 – Exemplo Árvore de Decisão Gerada	54
QUADRO 10 – Percentual de Penetração do Convênio Débito Automático	59
QUADRO 11 – Resultados das Classes Pessoa Física	62
QUADRO 12 – Resultado das Classes Pessoa Jurídica	64
QUADRO 13 – Proposta de Segmentação para Cooperativa Alfa	68
QUADRO 14 – Proposta de segmentação para cooperativa Alfa com Convênio	68
QUADRO 15 – Exemplo Comportamento Observado I	71
QUADRO 16 – Exemplo Comportamento Observado II	71

SUMÁRIO

1 INTRODUÇÃO	11
1.1 PROBLEMA DE PESQUISA	12
1.2 OBJETIVO GERAL	13
1.3 OBJETIVOS ESPECÍFICOS	13
1.4 JUSTIFICATIVA	13
2 ABORDAGEM TEÓRICA	17
2.1 COOPERATIVISMO	17
2.1.1 Cooperativismo de Crédito	19
2.2 SEGMENTAÇÃO	21
2.2.1 Segmentação em Instituições Financeiras	25
2.3 POSICIONAMENTO ESTRATÉGICO	28
2.4 MINERAÇÃO DE DADOS	30
2.5 ANÁLISE DE AGRUPAMENTO	32
2.6 MÉTODO K-MEANS	33
2.7 ALGORITMO C 4.5	35
3 METODOLOGIA	48
3.1 ROTEIRO DA PESQUISA	48
4 DISCUSSÃO DOS RESULTADOS	55
4.1 ESTATÍSTICA DESCRITIVA	55
4.2 ANÁLISE DE CORRELAÇÃO	60
4.3 MINERAÇÃO DE DADOS	61
5 CONCLUSÃO, LIMITAÇÕES DA PESQUISA E IMPLICAÇÕES GERENCIAIS	73
5.1 CONCLUSÃO	73
5.2 LIMITAÇÕES DO ESTUDO	75
5.3 IMPLICAÇÕES GERENCIAIS	76
REFERÊNCIAS	78
APÊNDICE A	83
APÊNDICE B	86
APÊNDICE C	89
APÊNDICE D	90

1 INTRODUÇÃO

O Sistema financeiro nacional vem observando o crescimento de uma modalidade de instituição financeira que é o cooperativismo de crédito. Segundo a Organização das Cooperativas Brasileiras (2016), as cooperativas de crédito são destinadas a promover poupança e financiar necessidades ou empreendimentos dos seus cooperados, atuando no crédito rural e urbano.

Na Europa, segundo Meinen (2002), destaca-se como exemplos a Alemanha, onde cerca de 20% do mercado financeiro passa pelas cooperativas de crédito, e a Holanda, na qual 90% do crédito rural concedido é oriundo deste tipo de instituição.

Atualmente são mais de 6 milhões de pessoas no Brasil que fazem parte desse modelo, segundo a WOCCU (2016). O Banco Central do Brasil (2016) revela que o cooperativismo de crédito perfaz o percentual de 2,44% de representatividade no total de ativos do mercado financeiro, o qual concentra cerca de 90% do total desses ativos em 8 bancos. Comparativamente ao sistema bancário, todas as cooperativas de crédito juntas seriam o nono maior banco do país. Essa concentração bancária, ocorrida principalmente após a implantação do plano real, quando a estabilização da inflação trouxe grandes complicadores aos bancos que trabalhavam somente com o mercado “*overnight*”, ou seja, captavam dinheiro dos clientes e aplicavam apenas um dia em operações noturnas no mercado aberto e não faziam empréstimos. A queda das taxas de rentabilidade deste produto obrigou os bancos a criarem produtos com maiores rentabilidades e ainda ajustar de forma drástica os custos, trabalhando a eficiência. Todo esse cenário trouxe um impulso às cooperativas que saíram de 806 para 1.311 singulares, um salto de 63% entre os anos de 1990 a 2000, segundo dados do Banco Central do Brasil.

O cooperativismo de crédito no Brasil é representado por sistemas de cooperativas de crédito tais como: Sicoob, Sicredi, Unicred, Uniprime, Cresol, Ceced e ainda cooperativas independentes de sistema que são conhecidas como “solteiras”. Dentre os sistemas de cooperativas, o Sicoob é a maior instituição financeira cooperativista do Brasil, representando um terço do total dos ativos de todos os demais sistemas de cooperativas financeiras.

A cooperativa “Alfa” é uma cooperativa de crédito pertencente ao sistema Sicoob, fundada em maio de 2003, no norte do Paraná, e será utilizada como objeto para este

estudo. Após 11 anos, ao final do ano de 2014, essa cooperativa encerrou o exercício com mais de meio bilhão em ativos, e mais de 19 mil associados.

Para o gerenciamento deste novo cenário, a cooperativa Alfa percebeu a necessidade de um posicionamento estratégico diferenciado a fim de atender esta nova demanda de forma adequada.

Para Porter (1999), a segmentação é a origem do posicionamento estratégico de uma empresa, que emerge de três fontes distintas: a variedade de produtos e serviços; a necessidades dos clientes; e o acesso (forma de alcançar o cliente).

Ainda, segundo Porter (1999), o lema da estratégia competitiva é ser diferente. O que significa escolher, deliberadamente, um conjunto diferente de atividades para proporcionar um mix de valor único.

O estudo das práticas de posicionamento de *marketing*, bem como do gerenciamento de relação com o cliente e ainda a segmentação, pretende contribuir para evidenciar a oportunidade de segmentação da base real de associados da cooperativa Alfa.

1.1 PROBLEMA DE PESQUISA

Bueno e Ikeda (2013) afirmam que, quanto ao relacionamento entre um cliente e a instituição bancária de que ele utilizará os serviços, este processo não começa apenas quando a escolha por parte daquele ocorre. Além disso, segundo os mesmos autores, a crescente concorrência do setor, impulsionada pelos recentes movimentos de fusões, aquisições e reestruturações, que deram origem aos grandes conglomerados atuais, fez com que o principal caminho para as instituições gerassem vantagem competitiva e diferenciação fosse a inovação nos serviços prestados.

Neste mesmo contexto estão inseridas as cooperativas de crédito, já que prestam serviços idênticos aos bancos. A livre admissão de associados, para este ramo do cooperativismo, trouxe uma fragilidade no relacionamento destas cooperativas com seus associados, devido ao crescimento da quantidade numérica desses. Para Bialoskorski Neto (2007), as cooperativas, em seu processo de crescimento, tenderiam a passar de um estágio inicial para um estágio mais avançado, em uma nova posição estratégica que exigiria a agregação de valor.

Sendo assim, percebe-se a necessidade da criação de valor no atendimento a estes associados pelas cooperativas de crédito, e, neste sentido, apresenta-se o seguinte problema de pesquisa:

A base de associados da cooperativa “Alfa” evidencia oportunidades de oferta pautada na estratégia de segmentação de clientes?

1.2 OBJETIVO GERAL

Tratando dos aspectos de posicionamento de *marketing* e relacionamento com o cliente destaca-se o seguinte objetivo geral:

Identificar se a base de associados da cooperativa “Alfa” evidencia oportunidades de oferta pautada na estratégia de segmentação de clientes.

1.3 OBJETIVOS ESPECÍFICOS

Em desdobramento ao objetivo geral, ressaltam-se os objetivos específicos a seguir:

- a) Caracterizar a base de associados da cooperativa Alfa;
- b) Identificar padrões de comportamentais na base de associados;
- c) Verificar, entre esses padrões e comportamento, a possibilidade de oferta de produtos e serviços mais direcionados às necessidades dos cooperados.

1.4 JUSTIFICATIVA

A cooperativa Alfa foi fundada em maio de 2003 no norte do estado do Paraná com o intuito de ajudar os empresários desta região a fugir da dita exploração bancária com altas taxas de juros e tarifas. Até o ano de 2009, a cooperativa atendia somente um determinado número de associados do ramo de atividade, proposto pelo seu objeto social. Contudo, partir daquele ano, passou a ser do formato livre admissão de associados, que é uma forma estatutária das cooperativas de crédito que permite a adesão de qualquer

natureza de associados e não somente em nichos específicos, como ramos de saúde, confecções, etc.

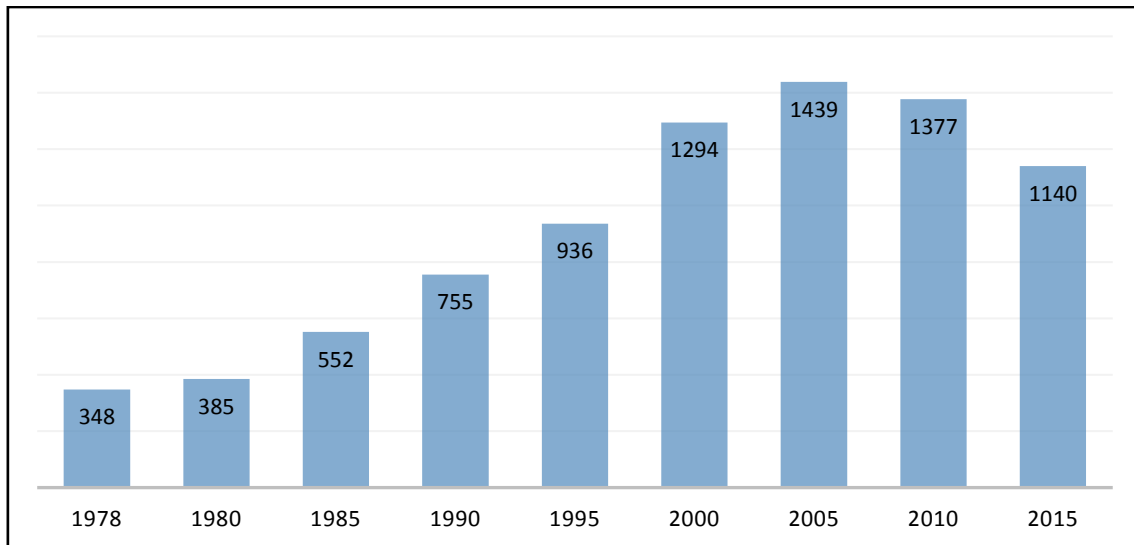
Com este novo cenário, o incremento de cooperados observou um crescimento acima do mercado, em apenas 4 anos, já somavam 19 mil associados que provinham dos mais variados ramos do mercado, e, além de empresas, um grande número de pessoas físicas. A força de vendas é formada principalmente por gerentes comerciais, que são liberados para abrir contas seguindo apenas diretrizes burocráticas, sem um posicionamento estratégico de *marketing*, ou seja, um mesmo gerente pode atender pessoas físicas e jurídicas dos mais variados ramos, ou, de qualquer nível de renda.

Observando esse fenômeno, algumas questões surgem, tais como: esta é a melhor forma de fazer negócios? Os associados são bem atendidos? E os gerentes, possuem competências específicas? É extraído o melhor da força de vendas com este modelo?

Tendo em vista que a questão estratégica que orienta o negócio não é mais “o que podemos produzir e vender?” Mas sim “o que podemos fazer para nosso cliente?” (DAWAR, 2013, p.2), a escolha do objeto de pesquisa deste estudo justifica-se devido ao crescimento do número de associado das cooperativas de crédito e a redução do número de singulares através das fusões e incorporações, como se pode observar a seguir nos gráficos 1 e 2.

Apresenta-se no gráfico 1 a evolução do número de cooperativas singulares no Brasil, evidenciando o crescimento em quantidade de singulares até o ano de 2005, já que após este ano, a tendência demonstrada, através do mesmo gráfico, é de redução do número de singulares, com o processo de fusões entre as cooperativas.

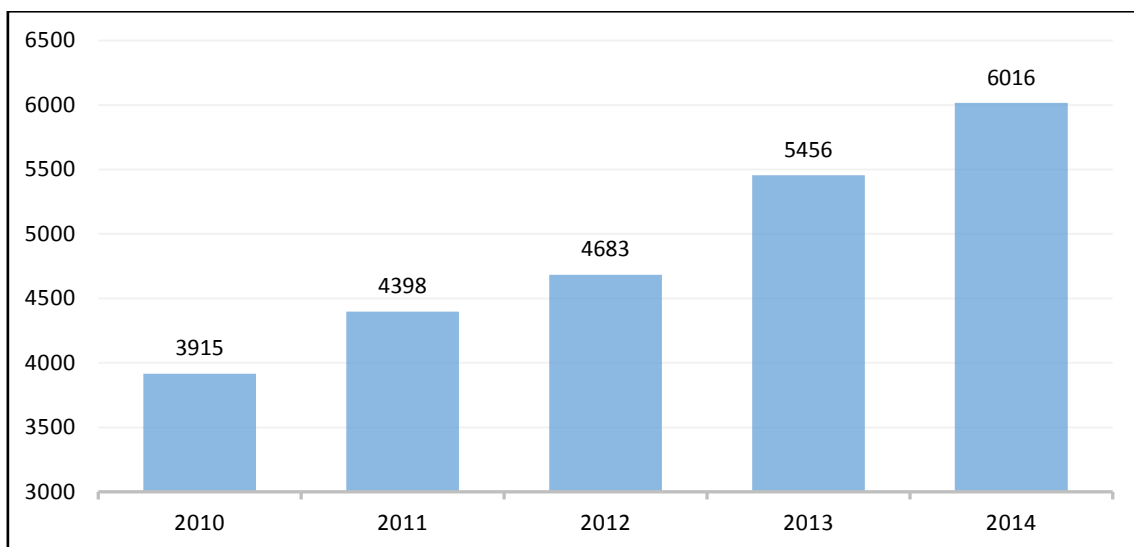
GRÁFICO 1 – Evolução Cooperativas Singulares no Brasil



Fonte: Banco Central do Brasil (2015)

O gráfico 2 apresenta a evolução do número de cooperados das cooperativas de crédito no Brasil e revela que em 4 anos este número praticamente dobou de tamanho, saindo de 3,9 milhões para 6,0 milhões de pessoas sendo associadas do cooperativismo financeiro.

GRÁFICO 2 – Número de Membros do Cooperativismo de Crédito no Brasil



Fonte: Word Council of Credit Unions (2015)

As cooperativas perceberam certa fragilidade nas relações com seus associados pelo grande número de pessoas e empresas que passaram a fazer parte de seus quadros, e exigindo assim, maiores cuidados dos gestores. Segundo Olson (apud BIALOSKORSKI NETO, 2012), poderá haver falta de comprometimento, de participação de associados em

grandes grupos, uma vez que eles não teriam uma percepção nítida da importância de sua participação para todo o grupo. E ainda Emmerich Bakonyi (1990) afirma que como um consumidor/cliente, o membro não se limita às transações bancárias ou de *merchandising* oferecidos por sua cooperativa de crédito.

Sendo assim, faz-se necessário evidenciar se a base de associados da cooperativa Alfa possui oportunidades de oferta baseada em segmentação contribuindo para uma assertividade no atendimento às necessidades do associado que, segundo Dibb (1998), está cada vez mais diversificada, e não podem ser satisfeitas com estratégias baseadas em *marketing* de massa. As empresas têm que lidar com essa diversidade de grupos de consumidores com similar comportamento de compras em um segmento.

Há contribuição também no sentido de otimizar a ação da força de vendas. Cherón, McTavish e Perrien (1989) afirmam que os bancários no mercado comercial não conseguem manter um longo relacionamento pessoal com todos os clientes, por conta da heterogeneidade, complexidade do mercado e a variedade de produtos oferecidos. Nesse contexto, a segmentação de mercado pode ajudar os gerentes de contas a entender melhor o mercado que enfrentam.

2 ABORDAGEM TEÓRICA

Neste capítulo serão apresentadas as principais teorias em relação ao desenvolvimento da pesquisa proposta. Sendo assim, esse momento será fundamental para orientar a análise e interpretação dos resultados deste trabalho. O primeiro conceito a ser abordado é cooperativismo, como pode se observar a seguir.

2.1 COOPERATIVISMO

Tanto como doutrina econômica, tanto como doutrina social, o cooperativismo tem princípios bem definidos, os quais, tanto uma como a outra denominação, se igualam, sendo citadas várias definições. Apresenta-se a seguir algumas consideradas relevantes para este estudo:

O Cooperativismo é a doutrina econômica que, consagrando os princípios fundamentais da liberdade humana, intervém na ordem econômica em defesa do interesse de agrupamentos, para proporcionar a cada um de seus participantes o melhor resultado de sua atividade econômica (ALESSIO; FORTES, 1977, p. 13).

Doutrina que tem por objetivo a solução de problemas sociais por meio da criação de comunidades de cooperação [...], formadas por indivíduos livres (SANDRONI, 2003, p. 132).

Não se pode afirmar que o cooperativismo é uma doutrina econômica tão somente tendo em vista suas vertentes sociais, tais como a justa divisão das sobras entre os cooperados e o controle democrático. Não é possível também declarar com firmeza que se trata de uma doutrina exclusivamente social, embora sua origem encontre-se nas propostas dos chamados socialistas utópicos. No cooperativismo existe uma forma de lucro que está subentendido nas sobras. Pode-se então definir o cooperativismo como uma doutrina sócio-econômica, que tenta balancear o capitalismo e o socialismo. Assim Pinho (1977) define:

O cooperativismo é tudo isso ao mesmo tempo: Movimento nascido nos meios populares da Europa Ocidental de meados do século XIX, para uma ação pacífica de defesa e de emancipação sócio – econômica [...]; Doutrina surgida originalmente em oposição ao capitalismo desbragado, mas que depois reagiu também contra os sistemas econômicos que se instalaram em vários países, embasados em crescente intervencionismo estatal ou no socialismo revolucionário (PINHO, 1977, p. 24).

Ou ainda, de forma mais evidente, Alessio e Fortes (1977, p.13) afirmam que “o cooperativismo trata-se de ‘uma sociedade que é formada com objetivo sócio-econômico’”.

Estudiosos do movimento cooperativista, como Hugon (1995), por exemplo, são unânimes em atribuir a Robert Owen o pioneirismo no movimento cooperativista, ao qual deve também a sua grande celebridade. É assim denominado a partir da afirmação de Beatriz Potter-Webb - historiadora do cooperativismo da Grã-Bretanha: Esta fé que inspira os primeiros socialistas de nosso país e que se tornou a fonte vital do movimento cooperativo – é de ‘origem puramente britânica’ (apud PINHO, 1977).

Nascido no País de Gales, provindo de uma família de modestos artesãos, Owen, após algum aprendizado, tornou-se, aos 30 anos, co-proprietário e diretor de indústrias escocesas de fiação em New Lanark. Começou a pôr em prática em suas próprias indústrias suas concepções sociais e econômicas, buscando soluções para problemas dessa mesma ordem. Em 1924 Owen cria nos Estados Unidos, em Indiana, a colônia comunista “*New Harmony*”, formada por 2.500 europeus. Em apenas dois anos essa tentativa fracassou. Outras tentativas existiram, mas sem sucesso.

Com fortes traços de suas ideias surge, em 1844, a sociedade “*Rochdale Society of Equitable Pioneers*” (Sociedade Pioneira Equitativos de Rochdale), a alma mater de todo o cooperativismo moderno, com um surto de rápida expansão. Segundo Hugon, Willian Thompson e Étienne Cabet (1995), Owen não chega a criar propriamente uma escola, e deixa poucos discípulos. Os autores afirmam ainda que o movimento cooperativista foi incentivado também por Charles Fourier, Benjamin Buchy, Saint-Simon, Louis Blanc e Guitton.

Diante deste cenário de expansão do modelo cooperativista, seria fundamental a criação de mecanismos que sustentassem este ideal ao longo do tempo e do espaço, já que outros países, como, por exemplo, o Brasil, adeririam à cooperação.

Foi assim que em 1895 foi criada a Aliança Internacional do Cooperativismo, uma entidade mundial, com o intuito de prover informações sobre o cooperativismo, realizar trocas entre as cooperativas pelo mundo e, principalmente, definir e manter os princípios do cooperativismo, que atualmente são:

- a) Adesão voluntária e livre;
- b) Controle democrático dos sócios;
- c) Participação econômica dos membros;
- d) Autonomia e independência;
- e) Educação, treinamento e informação;
- f) Intercooperação;

g) Preocupação com a comunidade.

No Brasil, a OCB (Organização das Cooperativas do Brasil) foi fundada em 1969 por iniciativa das cooperativas de todo o País. Segundo a OCB, quando o médico francês Jean Maurice Faivre em 1847, adepto dos ideais de Charles Fourier, fundou nos sertões do Paraná, juntamente com alguns europeus, a colônia Tereza Cristina. Lá não existia escravidão, e todos trabalhavam para o sustento de todos, dividindo os resultados das colheitas. Assim, atuando unidas, as pessoas teriam todas as necessidades supridas. Desse modo, deu-se início aos princípios cooperativistas no Brasil.

Segundo a OCB, em seu relatório de gestão do ano de 2015, neste mesmo ano, 50% da produção agropecuária do Brasil passou, de alguma maneira, por uma cooperativa, e, ainda, segundo a ANS, 32% de clientes de planos de saúde são atendidos por cooperativas médicas. Um estudo do FGCCoop (Fundo Garantidor das Cooperativas de Crédito) (2016) afirma que todas as cooperativas de crédito, juntas, possuem a maior rede de atendimento do sistema financeiro, com 5.432 pontos distribuídos por todo o País.

No Brasil são 13 ramos de cooperativas segundo a OCB: Agropecuário, Consumo, Crédito, Educacional, Especial, Habitacional, Infraestrutura, Mineral, Produção, Saúde, Trabalho, Transporte, Turismo e Lazer.

Dos ramos apresentados acima, e em consonância com os objetivos desta pesquisa, no texto que segue será tratado o do cooperativismo de crédito.

2.1.1 Cooperativismo de Crédito

O cooperativismo de crédito surgiu com o objetivo principal de proporcionar crédito e moeda aos seus cooperados, nos princípios básicos de Rochdale, com módicas taxas de juros, independentemente do ramo de atividade, seja agrícola, mercantil ou industrial.

Entendeu-se no envolver da atividade cooperativa, cujos princípios lapidares foram tão bem deduzidos pelos Pioneiros de Rochdale, que, como os consumidores podiam, associados cooperativamente, substituir os negociantes e intermediários, executando eles próprios essas atividades e excluindo assim, o lucro por aqueles auferidos – e da mesma forma os produtores, tanto agrícola como industriais – podiam também substituir os bancos e usuários, financiando-se a si próprios, quer na atividade isolada (associados), quer na atividade comum (associação), evitando os males da

usura e a percepção das taxas pelos órgãos creditícios – propiciando assim ao movimento sua total independência (BULGARELLI, 1962, p. 54).

Os primórdios do cooperativismo de crédito, segundo Bakonyi (1990), se desenvolveram com transações bancárias para clientes envolvidos na agricultura. As primeiras cooperativas de crédito surgiram na Alemanha, Itália e Canadá. Conforme Menezes (2004), para melhor entender o cooperativismo de crédito é preciso conhecer, ainda que em síntese, a trajetória formadora desse processo cooperativo a partir das primeiras experiências que são as da Alemanha, Itália e Canadá.

Foi na Alemanha, em meados do século XIX, que o crédito cooperativista teve seu aparecimento (BULGARELLI, 1962), com Raiffeisen e Schulze-Delitzch, o primeiro com o crédito rural, o segundo com o crédito urbano. Depois vieram Luzzatti na Itália, Desjardins na América do Norte, mais precisamente no Canadá. Em Nova Petrópolis, no Rio Grande do Sul, no ano de 1902, é que surgiu a primeira cooperativa de crédito no Brasil. Fundada pelo padre Teodor Amstadt, a Caixa Rural de Nova Petrópolis funcionava no modelo Raiffeisen.

Ainda no Brasil, Pinheiro (2008) descreve a cooperativa de crédito:

São instituições financeiras constituídas sob a forma de sociedade cooperativa, tendo por objeto a prestação de serviços financeiros aos associados, como concessão de crédito, captação de depósitos à vista e a prazo, cheques, prestação de serviços de cobrança, de custódia, de recebimentos e pagamentos por conta de terceiros sob convênio com instituições financeiras públicas e privadas e de correspondente no País, além de outras operações específicas e atribuições estabelecidas na legislação em vigor (PINHEIRO 2006, p. 7).

Pinho (1977) afirma que o legislador no Brasil definiu que as cooperativas de crédito têm o objetivo de proporcionar crédito mediante uma taxa módica de juros.

A distribuição das cooperativas no Brasil está demonstrada na figura 01, que evidencia a concentração destas instituições nas regiões Sul e Sudeste com 78%.

FIGURA 1 - Distribuição das Cooperativas de Crédito por Regiões no Brasil.



Fonte: Banco Central do Brasil (2016). Elaborado pelo Autor.

Apresentados os conceitos que serão utilizados em relação ao cooperativismo, bem como o cooperativismo de crédito, apresentam-se, na sequência, os conceitos relativos à segmentação.

2.2 SEGMENTAÇÃO

A segmentação, para Porter (1999), é a origem do posicionamento estratégico de uma empresa. Como já abordado à introdução deste estudo, o posicionamento manifesta-se de formas distintas sendo as seguintes: diversificação de produtos e serviços, baseado nas necessidades dos clientes, e no acesso ao cliente. Ainda, neste sentido Porter, trata do lema da estratégia competitiva que para ele é ser diferente, e ser diferente significa

deliberadamente decidir por um conjunto de atividades diferentes proporcionando assim um mix de valor único.

A segmentação pode ser definida como o processo de se colocar em subgrupos os compradores de um produto-mercado, de modo que os integrantes de cada segmento exibam uma responsividade semelhante em relação à determinada estratégia de posicionamento (CRAVENS E PIERCY, 2007).

O conceito de segmentação, segundo Brochado e Martins (2008), foi introduzido na literatura de *marketing* por Smith (1956) e tem inspiração na teoria econômica da concorrência imperfeita (ROBSON, 1938 apud BROCHADO e MARTINS), considerando o pressuposto de que os bens e serviços não podem ser vendidos sem a consideração das necessidades dos consumidores e o reconhecimento da sua heterogeneidade.

Para que uma segmentação seja eficiente, Gavett (2014) afirma que ela deve possuir seis características:

- a) Identificável: identificar clientes em cada segmento e medir as características como demografia ou comportamento de uso;
- b) Substancial: ser grande o suficiente para ser lucrativo;
- c) Acessível: alcançar os segmentos para comunicação e canais de distribuição;
- d) Estável: ser estável o suficiente por um período longo a fim de ser atingido pela estratégia;
- e) Diferenciável: clientes com necessidades similares e diferentes dos outros segmentos;
- f) Acionável: ser capaz de prover os produtos e serviços para o segmento.

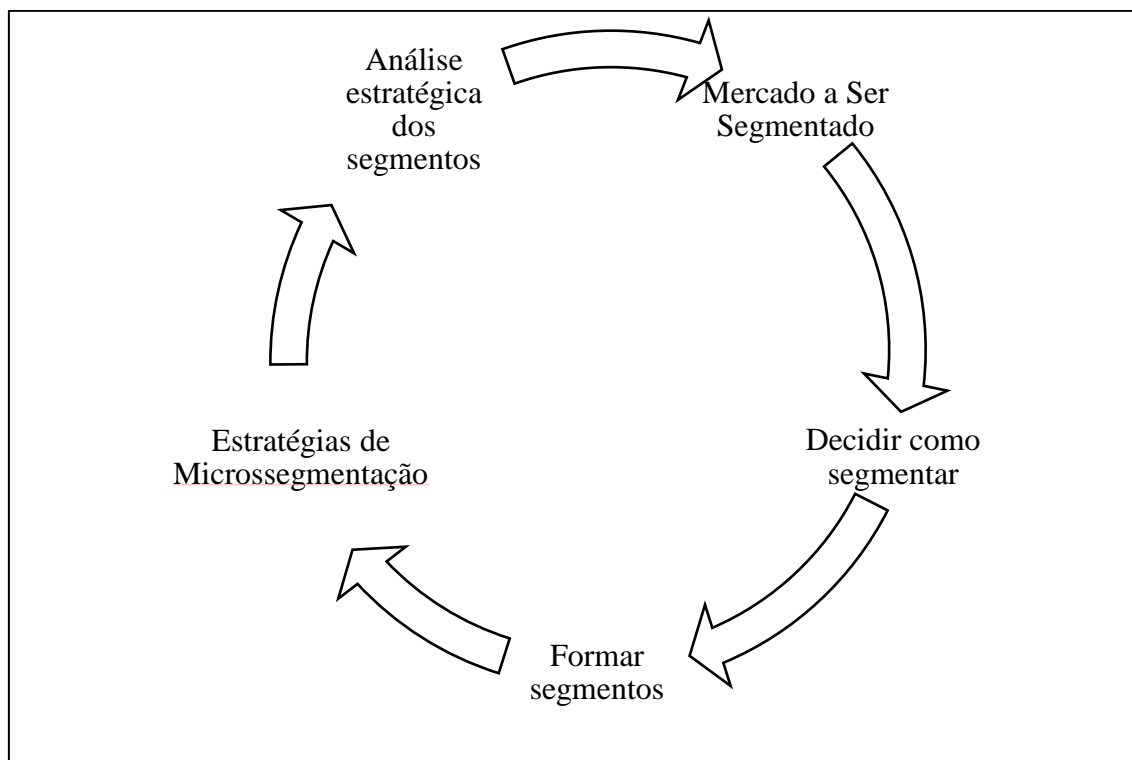
De acordo com Tavares, Afonso e Locatelli (2011), a segmentação de mercado é vista a partir de três níveis de abrangência. No primeiro, ela se restringe aos grupos que apresentam similaridades e diferenças entre si. O segundo nível de abrangência inclui o que decorre da identificação de grupos homogêneos com a empresa, selecionando segmentos que deseja tomar como prioritários. O terceiro compreende além do mapeamento de grupos homogêneos e, à escolha dos segmentos alvo, a elaboração do composto de *marketing*.

De uma forma direta, Berrigan e Finkbeiner (1994) afirmam que, para ser eficaz, a segmentação não deve ser complicada demais para ser compreendida e aplicada. Ao mesmo tempo, deve captar a diversidade da população de clientes que se almeja.

Para Tavares, Afonso e Locatelli (2011), segmentar é uma tarefa gerencial, deliberada a ser aplicada por qualquer organização frente a seu mercado. Ela tem muitas utilidades práticas: orienta a decisão sobre quais segmentos focar; estimula a identificação de possibilidades estratégicas de diferenciação de ofertas; e realça a importância de ação empresarial proativa na configuração do mercado de atuação.

Para segmentar um mercado, são necessárias decisões e atividades que devem ser planejadas de forma a gerar assertividade neste processo. Cravens e Piercy (2007) apresentam um modelo destas atividades e decisões, conforme pode ser observado na figura 2:

FIGURA 2 - Modelo de Decisão e atividades da Segmentação



Fonte: Cravens e Piercy (2007)

Observa-se no modelo que a primeira decisão a ser tomada para uma segmentação é a definição do mercado a ser segmentado. Em seguida como segmentar, e isso envolve a seleção de variáveis que serão utilizadas como base para identificar os segmentos. Na sequência, a forma dos segmentos, cujo processo é realizado através de estatísticas do

comportamento de compra dos consumidores. Depois, verificar se existirão segmentos mais sofisticados. E, finalmente, realizar análise estratégica de cada segmento para decidir qual o segmento terá foco.

Na visão de Dawar (2013, p. 2), a questão estratégica que orienta o negócio não é mais “o que podemos produzir e vender?”, mas sim, “o que podemos fazer para nosso cliente?”. Cravens e Piercy (2007) afirmam que os compradores variam de acordo com o modo como usam os produtos, com as suas necessidades e com as suas preferências, além de seus padrões de consumo. Essas diferenças evidenciam os segmentos de mercado. Reconhecer as diferenças entre os segmentos de mercado e como elas se modificam, antes dos concorrentes, é uma fonte cada vez mais importante de vantagem competitiva.

Ainda neste sentido, existem incontáveis maneiras para segmentar um determinado mercado, sobretudo se a empresa cogita várias de suas formas, conforme explicam Richers e Lima (1991).

A seguir apresenta-se um quadro comparativo entre autores para as bases de segmentação existente:

TABELA 1 - Modalidades de Segmentação

Modalidade	Significado	Autores
Geográfica	Divisão do mercado por localização: Nação, Estado, Cidade, etc;	Richers e Lima (1991); Tavares, Afonso e Locateli (2011); Brochado e Martins (2008).
Demográfica	Divisão do mercado por idade, tamanho da família, sexo, religião, etc;	Richers e Lima (1991); Tavares, Afonso e Locateli (2011); Brochado e Martins (2008).
Socioeconômica	Divisão do mercado por renda, instrução, ocupação, <i>status</i> , etc	Richers e Lima (1991); Brochado e Martins (2008)
Padrão de Consumo	Divisão do mercado por frequência de compra, local de compra, lealdade a marcas;	Richers e Lima (1991); Tavares, Afonso e Locateli (2011).
Benefícios Procurados	Divisão do mercado em função dos benefícios buscados;	Richers e Lima (1991); Tavares, Afonso e Locateli (2011).

Estilo de Vida	Divisão do mercado pela expectativa de vida, uso do tempo, interesses predominantes;	Richers e Lima (1991); Tavares, Afonso e Locateli (2011).
Personalidade	Divisão do mercado por bases culturais, atitudes e valores, liderança;	Richers e Lima (1991);
Caracterização Econômica	Divisão do mercado por setor de atividades, tamanho de empresas, atuação de concorrentes, acessibilidades.	Richers e Lima (1991);

Fonte: Elaborado pelo autor, 2016.

Após esta apresentação das bases de segmentação percebe-se que, no mercado financeiro, a prática de segmentar também é utilizada. Dessa forma, procura-se a seguir evidenciar este conceito em instituições financeiras.

2.2.1 Segmentação em Instituições Financeiras

Embora este estudo vise tratar da segmentação em cooperativas de crédito, o fato da literatura não identificar, ao menos preliminarmente, estudos relacionados a esse assunto aplicados em cooperativas, serão utilizadas as pesquisas realizadas em empresas com similaridade de mercado, que são as instituições financeiras bancárias convencionais.

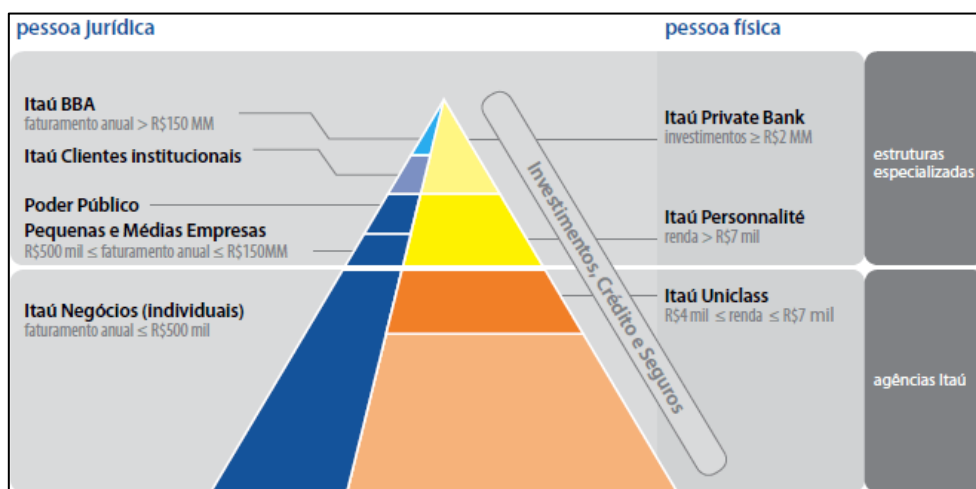
Para Bueno e Ikeda (2013), a evolução do conceito de *marketing* bancário deu-se a partir do momento em que as necessidades dos clientes bancários foram se tornando mais complexas, de modo que, tanto as soluções quanto as abordagens tradicionais, representadas por contas correntes, cartões, financiamentos e investimentos, não eram mais suficientes para atenderem a problemas específicos enfrentados por segmentos distintos.

A teoria de *marketing* avançou grandes passos para, possivelmente, se tornar a mais importante ferramenta atual no mercado competitivo de serviços financeiros (Wills, 1985). Morgan (1993) ressalta os desenvolvimentos do *marketing* no mercado bancário na França, com foco particular no crescimento da importância da segmentação de mercado.

Nesse mesmo aspecto, Morgan (1993) afirma que a cooperativa *Credit Agricole* é pioneira no mercado financeiro francês no uso do *marketing* bancário. Essa cooperativa francesa baseou a segmentação de seus associados por um perfil psicográfico, mais precisamente, um estudo publicado no pós-guerra sobre o estilo de vida dos europeus.

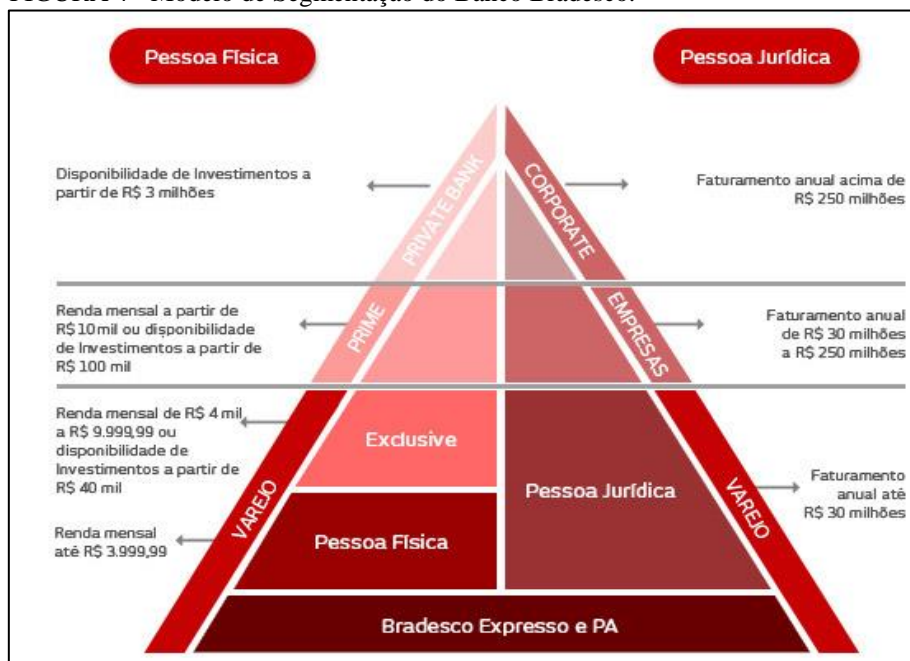
No Brasil o mercado financeiro atua com a segmentação de seus clientes, sendo que os três maiores bancos privados do país apresentam as seguintes formas de segmentação:

FIGURA 3 - Modelo de Segmentação do Banco Itaú



Fonte: Banco Itaú

FIGURA 4 - Modelo de Segmentação do Banco Bradesco.



Fonte: Banco Bradesco

FIGURA 5 - Modelo de Segmentação do Banco Santander.



Fonte: Banco Santander

As cooperativas de crédito estão inseridas no mercado financeiro. Porém até o momento não apresentam deliberadamente aos cooperados formatos de atendimento diferenciado. E neste sentido, a legislação brasileira, por meio da lei do cooperativismo a 5.764 de 1971, evidencia no artigo 4º item IX a indiscriminação religiosa, racial e social, e ainda, no artigo 37, que trata da igualdade de direitos dos associados, ressalta no item III que os associados devem ter livre exercício dos direitos sociais.

Tanto a legislação brasileira, como mencionado no parágrafo anterior, quanto a cultura do cooperativismo, que em seus valores essenciais trata da equidade entre os membros, podem limitar as ações de *marketing* no que diz respeito à diferenciação de atendimento aos associados. Para Bialoskorski Neto (2012), para se elaborar e implementar uma política de negócios consistente em cooperativas, existem especificidades doutrinárias e organizacionais que devem ser levadas em consideração.

Ainda para Bialoskorski Neto (2012), a falta de disponibilidade de informações inviabiliza a garantia de condições para elaboração de políticas de negócios em cooperativas. Nesse mesmo sentido, Dibb (1998) afirma que o *gap* entre a métrica do

impacto da segmentação nos negócios de uma empresa, torna falha a implementação dessa estratégia como solução de relacionamento.

O pouco entendimento dos princípios de segmentação, segundo Dibb (1998), é um exemplo clássico dos motivos que essa estratégia pode vir a falhar. Após a segmentação o posicionamento estratégico é o próximo conceito que se discute a seguir.

2.3 POSICIONAMENTO ESTRATÉGICO

Muitos pesquisadores de segmentação enfatizaram a importância crítica no relacionamento entre segmentação e a decisão de posicionamento (DOUGLAS; CRAIG, 1995; WIND, 1986 apud HASSAN; CRAFT, 2005). Por isso é imprescindível para a estratégia não parar apenas na seleção do segmento de mercado desejado, mas também incluir um posicionamento efetivo da marca em relação a este segmento (HASSAN; CRAFT, 2005).

A maioria dos gestores descreve o posicionamento estratégico em termos de clientes. Contudo, Porter (1999) define o posicionamento estratégico como a escolha entre executar as mesmas atividades de forma diferente ou de se concentrar em atividades diferentes em comparação com os rivais. Ou seja, a essência da estratégia está nas atividades.

De forma objetiva, Zineldin (1996) define o posicionamento estratégico como uma tentativa de ter uma posição clara ou única no mercado. Segundo Porter (1999), sua origem emerge de três fontes distintas que não são mutuamente excludentes e em geral se encontram imbricadas, a saber:

A) **Posicionamento baseado na variedade:** fundamenta-se na escolha de variedades de produtos ou serviços, não em segmentos de clientes. Este posicionamento faz sentido econômico quando a empresa é capaz de produzir determinados produtos ou serviços utilizando conjuntos de atividades diferenciadas.

B) **Posicionamento baseado nas necessidades:** está mais próximo do pensamento tradicional sobre orientação para um determinado segmento de clientes como alvo. Surge quando há um grupo de clientes com necessidades diferenciadas e quando um conjunto de atividades sob medida é capaz de satisfazer melhor a essas necessidades.

C) **Posicionamento baseado no acesso:** é a segmentação dos clientes em razão das diferenças nas modalidades de acesso. Embora as necessidades sejam

semelhantes as dos demais, a melhor configuração de atividades para o acesso a esses clientes é diferente. O acesso pode ser em função da geografia ou do porte do cliente – ou qualquer outra coisa que exija um diferente conjunto de atividades para melhor alcançá-lo.

Segundo Mintzberg (2010), Porter foi um divisor de águas, no que tange ao posicionamento estratégico, quando em 1980 publicou o Livro *Competitive Strategy*. Mesmo enaltecendo Porter, Henry Mintzberg trouxe sua contribuição para esta linha do estudo da estratégia.

Para Mintzberg (1988), as estratégias de diferenciação e, por conseguinte, o posicionamento estratégico, emanam de seis pontos básicos:

A) **Preço:** o ponto mais básico de se diferenciar é oferecer o menor preço do mercado em que se atua;

B) **Imagem:** em alguns momentos utilizar-se de marketing para criar diferenciação dos produtos, mesmo que sejam diferenças cosméticas. Cria-se uma imagem do produto;

C) **Suporte:** às vezes os efeitos do produto não estão no produto em si, mas sim no pós-venda, nas bases de suporte após a aquisição do produto;

D) **Qualidade:** a diferenciação por qualidade não trata de ser um pouco melhor, mas ser simplesmente “o melhor”;

E) **Design:** oferecer algo verdadeiramente diferente, que quebre paradigmas, entregando características únicas;

F) **Não-diferenciação:** não ter diferenciação também é um posicionamento estratégico, porém para ser definido como estratégia este deve de fato ter sido escolhido como posicionamento.

A chave do sucesso para obter um posicionamento favorável no mercado competitivo é oferecer valor nas necessidades e desejos dos atuais e possíveis clientes. (ZINELDIN, 1996).

Como este estudo trata de cooperativas de crédito que pertencem ao segmento de instituições financeiras, apresenta-se a seguir pelo menos quatro estratégias de posicionamento para o setor de serviços financeiros segundo Zineldin (1996):

A) **Posicionamento institucional:** Trata do escopo de operação. A instituição define se vai operar de forma nacional, internacional, regional ou numa comunidade, por exemplo. É possível selecionar mais de uma base.

B) **Posicionamento de linha de produtos e serviços:** Trata da natureza dos produtos. A instituição decide como vai oferecer os produtos e serviços, varejo, atacado, oferecer apenas para empresas (*corporate*), especialista em determinados produtos, por exemplo, investimentos.

C) **Posicionamento pelo sistema de distribuição/entrega e pessoal:** Trata de um *mix* entre recursos humanos, localização e equipamentos. O desenvolvimento de tecnologias e sistemas de informação, como os caixas eletrônicos, internet *banking*, tem contribuído para que as instituições financeiras ofereçam produtos a seus clientes 24 horas por dia. As habilidades humanas ganham importância para obter todas as vantagens competitivas possíveis da tecnologia e para prever as atuais tendências de tecnologias para que novas oportunidades sejam identificadas antecipadamente. A localização dos caixas eletrônicos e das agências também é fundamental para este posicionamento.

D) **Posicionamento por segmento:** Tentar prover de tudo a todos é difícil, se não impossível num mercado competitivo. Muitas instituições financeiras selecionam pequenas fatias-chave do mercado e concentram-se em servi-las muito melhor do que seus competidores.

A estratégia de posicionamento provê um quadro essencial de referências que guiam as decisões gerenciais, de acordo com Cravens e Piercy (2007). Para os autores, a rápida mudança no ambiente, as mudanças de preferências dos clientes, novos produtos e serviços e o crescimento da competição pela demanda exigem que as empresas monitorem continuamente a estratégia de posicionamento para capitalizar novas oportunidades e evitar armadilhas em potencial.

A seguir apresenta-se o conceito de mineração de dados que pode contribuir com os conceitos já apresentados até aqui.

2.4 MINERAÇÃO DE DADOS

Para Weiss e Indurkha, 1998 (apud REZENDE, 2005), a mineração de dados é definida como a pesquisa de informações valiosas em grandes escalas de dados. É ainda uma extração de informações não triviais que estão implícitas, não previamente conhecidas, e são potencialmente usuais (FRAWLEY, 1992 apud REZENDE, 2005).

A quantidade de informação gerada atualmente é enorme, criando assim, grandes bases de dados sejam eles originados da relação direta entre empresa e cliente ou,

comprados (FORCHT; COCHRAN, 1999). Diante desse cenário, a Mineração de Dados aparece como alternativa no sentido de utilizar essas informações de forma eficiente.

No rápido crescimento do número de informações geradas nas empresas, a MD (Mineração de Dados) vem sendo utilizada nos mais diferentes tipos de negócios. Ranjan e Bhatnagar (2009) afirmam que a aplicação da MD está crescendo cada vez mais no mundo empresarial. Isso pode ser confirmado por meio da análise da configuração de muitas organizações no ambiente de negócios atual.

Diante disso, cabe o seguinte questionamento: como a MD pode ajudar as empresas? Segundo Ranjan e Bhatnagar (2009), existem duas diferentes formas de informações extraídas que podem auxiliar o mundo dos negócios:

A) **Descritiva:** pode ser interpretada como a descoberta de informações desconhecidas de um grande volume de dados, que são extraídas dos dados do cliente e colaboram no *status* atual da organização.

B) **Preditiva:** utiliza a informação descritiva para prever tendências futuras. Isso subsidia os gerentes de negócios a tomar futuras decisões estratégicas para o crescimento e desenvolvimento na organização.

A aplicação das técnicas de MD em base de dados dos clientes tem impactado significativamente a área de *marketing*. Forcht e Cochran (1999) e Ranjan e Bhatnagar (2009) afirmam que a aplicação da MD no gerenciamento do relacionamento com o cliente vem produzindo imensos lucros para a indústria de serviços.

Para utilizar a mineração de dados, Lee e Siau (2001) propõem a utilização de três principais passos: preparação do banco de dados, redução do banco de dados e, finalmente, verificação das informações valiosas. Porém, estes passos podem variar de empresa para empresa e, assim, a IBM define quatro passos:

A) **Modelo Preditivo:** Utilizam-se técnicas de raciocínio indutivo, como redes de neurônios e algoritmos de raciocínio indutivo, para criar modelos de previsão;

B) **Segmentação do Banco de Dados:** Utilizam-se técnicas de *clusters* estatísticos para particionar a base de dados em *clusters*;

C) **Análise de Links:** identificam-se associações entre os dados;

D) **Deteção de Desvios:** Detecta-se e explica-se por que certos registros não podem estar em específicos segmentos.

A mineração de dados, ainda segundo Lee e Siau (2001), vem se desenvolvendo em várias técnicas e designs. Essas técnicas são classificadas em: embasamento no banco de dados, conhecimento a ser descoberto e técnicas a serem utilizadas.

A) **Baseado no Banco de Dados:** São vários os sistemas de banco de dados utilizados nas organizações como a base relacional (de transação, objeto orientado, especial, multimídia, de legado e *web*). O sistema de MD pode ser classificado pelo tipo de banco de dados que vai ser designado para análise.

B) **Baseado no Conhecimento:** A MD pode descobrir vários tipos de conhecimento, incluindo as regras de associação, de características, de classificação, os *clusters*, as evoluções e a análise de desvios. A MD ainda pode ser classificada de acordo com o nível de abstração do conhecimento gerado. O conhecimento pode ser classificado como geral, de nível primitivo e de múltiplo nível.

C) **Baseado nas Técnicas:** A MD pode ser classificada de acordo com as técnicas de mineração de dados utilizadas. Por exemplo: método dirigido, mineração autônoma de conhecimento, mineração orientada por banco de dados, mineração orientada por consulta.

Neste trabalho, a utilização da mineração de dados é confirmada pela afirmação de Lee e Siau (2001) de que a MD pode dar soluções para bancos ou indústria de crédito, tal como *score* de crédito, detecção de fraudes, e segmentação de clientes.

Para este estudo será utilizada a MD, que é baseada em banco de dados, após a análise de agrupamento, que se observa na sequência do estudo.

2.5 ANÁLISE DE AGRUPAMENTO

Uma metodologia estatística utilizada para análise de segmentação, sugerida pela literatura, seria a análise de agrupamento, que consiste em classificar atributos que permitam descrever as características fundamentais de similaridade entre indivíduos, (objetos), de um determinado grupo. Para Corrar, Paulo e Dias Filho (2017) a análise de agrupamento é definida assim:

A literatura nacional refere-se frequentemente à técnica em estudo como Análise de Conglomerado ou como análise de agrupamento. É uma análise que engloba uma variedade de técnicas e algoritmos, cujo objetivo é encontrar e separar objetos em grupos similares. O problema que se pretende resolver é: dada uma amostra de n objetos (ou indivíduos), cada um deles medido segundo p variáveis, procurar um esquema de classificação que agrupe objetos em g grupos. Devem ser determinados também o número e as características desses grupos (CORRAR; PAULO; DIAS FILHO, 2017, p. 324).

Ainda segundo os mesmos autores, as classificações dos objetos (ou indivíduos) pertencentes a um determinado grupo deve considerar um critério de seleção predeterminado.

Avançando pela história, Gouvêa, Prearo e Romeiro (2013) afirmam que a análise de agrupamentos foi utilizada pela primeira vez em 1939 por Tyron. Os mesmos autores chamam a análise de agrupamentos por análise de conglomerados, análise de *clusters*, análise de taxionomia e análise de segmentação.

A análise de agrupamento abriga uma variedade de algoritmos de classificação diferentes - todos voltados para a questão de como organizar dados observados em estruturas que façam sentido, ou como desenvolver taxinomias capazes de classificar dados observados em diferentes grupos (GOUVÊA; PREARO; ROMEIRO, 2013).

Ainda segundo os autores, dois dos principais algoritmos de classificação disponíveis nos *softwares* estatísticos são o método hierárquico e o método *k-means*. Método de agrupamento hierárquico é o mais apropriado para amostras inferiores a 250 casos (GARSON, 2007b apud GOUVÊA, PREARO E ROMEIRO 2013). Por outro lado, método *k-means* é mais apropriado para amostras grandes (a partir de 250 casos), ainda que também possa ser aplicado em amostras menores (GOUVÊA, PREARO e ROMEIRO, 2013). É este o método apresentado na próxima seção desta abordagem teórica.

2.6 MÉTODO *K-MEANS*

O algoritmo de classificação utilizado pela análise de agrupamentos chamado *k-means* é o mais popular, mais simples e também o mais utilizado (VASSILVITSKII, 2007; JIN, GOSWAMI & AGRAWAL, 2005; LIU & LU, 2015; WANG & BAI, 2016; SINGH & YADAV, 2013; GARCIA ESCUDERO & GORDALIZA, 1999; GOPINATHAN; PANDIYAN & THANGAVEL, 2017).

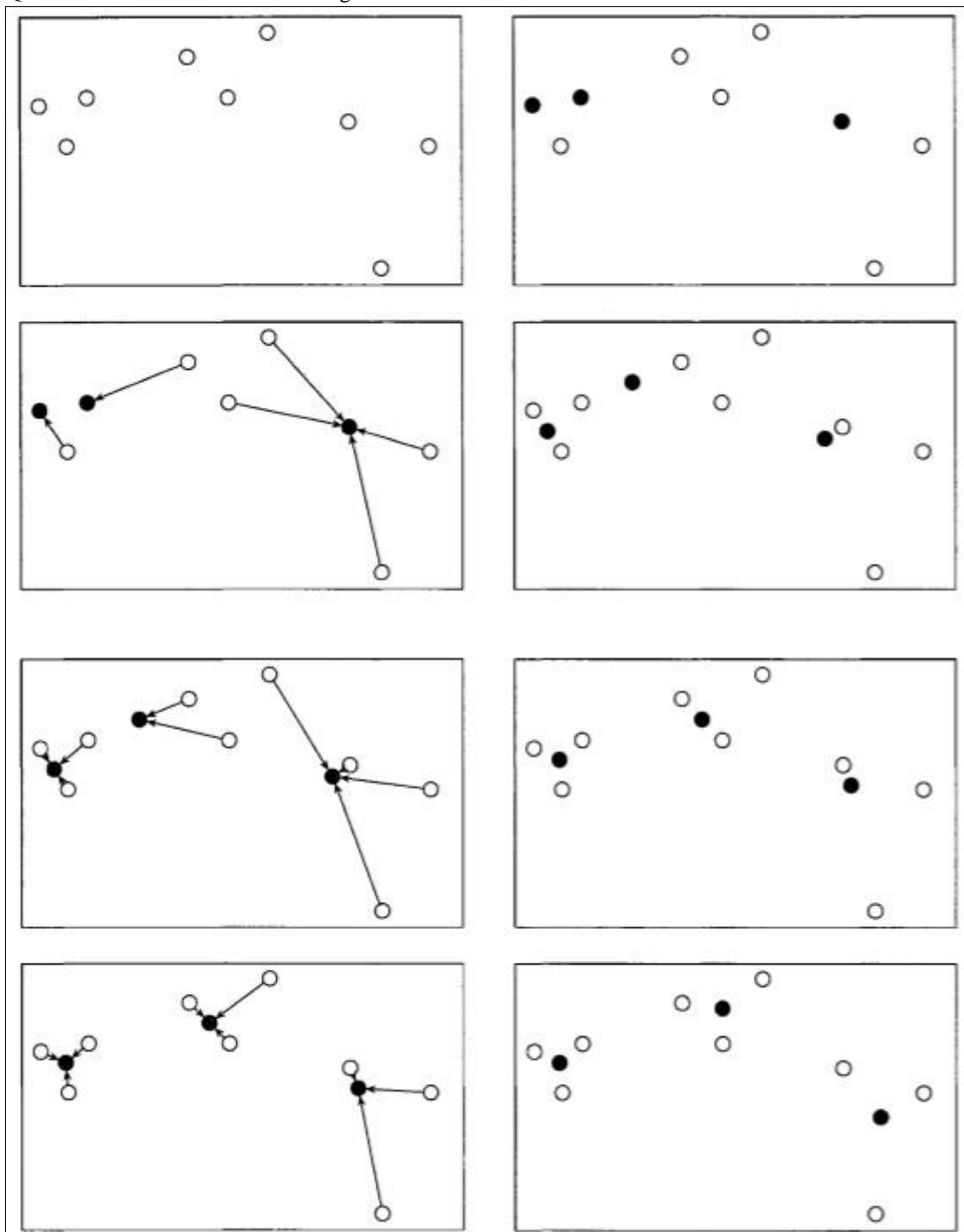
Este algoritmo possui uma história rica, e sua concepção foi atribuída a Stuart P. Lloyd em 1957, mas nunca apareceu em impressos até 1982 porque os cálculos nunca foram concluídos, até o avanço da computação. Embora seja um algoritmo velho, ele surpreende, pois é utilizado até hoje, muito por sua qualidade de precisão e tempo de execução (VASSILVITSKII, 2007).

O método do algoritmo *k-means* possui basicamente duas fases: na primeira ele calcula o *k* centroide e na segunda detecta cada ponto do *cluster* que está mais perto de

um determinado centroide (GOPINATHAN.S, PANDIYAN.M e THANGAVEL.P, 2017).

A seguir demonstra-se a execução do algoritmo k-means com 9 elementos n em 3 *clusters* k . Os círculos vazios são os dados distribuídos, enquanto os círculos pretos representam os centros dos *clusters*. Evidenciam-se os centros mais próximos em cada passo. O algoritmo converge em quatro interações.

QUADRO 1 – Funcionamento do algoritmo *k-means*



Fonte: Vassilvitskii, 2007

Depois de apresentar o método de clusterização k-means, que agrega a amostra em grupos homogêneos, é necessário apresentar os resultados desta agregação, ou seja, os grupos gerados, com suas respectivas características. Para isso evidencia-se a seguir a metodologia de criação de árvores de decisão e visualização de resultados utilizados nesta pesquisa que é o algoritmo C 4.5.

2.7 ALGORITMO C 4.5

O algoritmo C 4.5 foi criado por J. Ross Quinlan e concebido para criação de árvores de decisão e para representar regras de classificação, sendo um dos mais famosos indutores que usam árvore de indução (PRATI; BARANAUKAS; MONARD, 2001). As regras de decisão geradas pelo C 4.5 são de fácil entendimento e aceitação e possuem alta precisão (GAO; TANG; WANG, 2013).

A árvore de decisão é formada por um nó-raiz e vários nós-folha. O nó raiz é o atributo que melhor separa por si só os exemplos a serem classificados. Por sua vez, cada nó da árvore representa um atributo. Os ramos são formados pelos valores dos atributos e as folhas são as classificações dos exemplos de acordo com os nós e os ramos. O número de ramos dos nós-folha depende do tipo de árvore e do tipo de atributo e, também, do algoritmo utilizado para o processo de teste.

Por que a árvore de decisão é tão popular?

A) A construção de classificadores de árvore de decisão não requer nenhum domínio de conhecimento ou configuração de parâmetro e, portanto, é apropriada para explorar a descoberta de conhecimento;

B) As árvores de decisão podem lidar com dados multidimensionais;

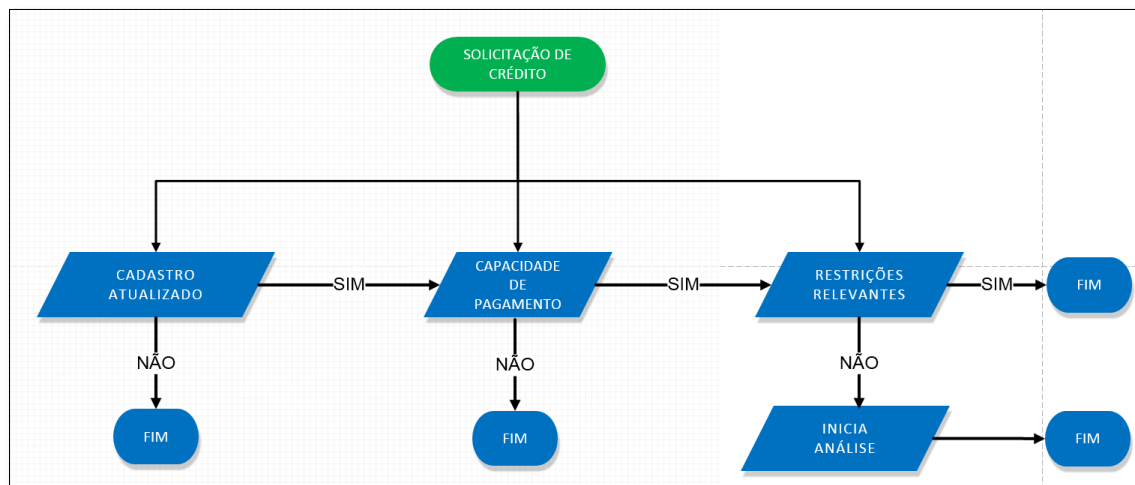
C) A representação do conhecimento adquirido em forma de árvore é intuitiva e geralmente é fácil de integrar pelos humanos;

D) Os passos de aprendizagem e classificação da indução das árvores de decisão são simples e rápidos. Em geral, os classificadores de árvores de decisão têm boa precisão;

E) O uso bem-sucedido dessa construção, no entanto, pode depender dos dados em mãos (AZAD; JAH, 2015).

A seguir um exemplo de árvore de decisão:

FIGURA 6: Modelo de Árvore de Decisão para Solicitação de Crédito



Fonte: elaborado pelo autor, 2016.

A árvore de decisão também pode ser escrita por meio de regras utilizando-se basicamente as expressões “SE”, “E” e “ENTÃO”. O quadro 2 a seguir apresenta um exemplo de como a árvore de decisão demonstrada na figura 6 poderia ser escrita:

QUADRO 2 - Modelo de Árvore de Decisão em Regras

Regra 1	SE existe a solicitação de uma proposta, E o cadastro está atualizado, ENTÃO segue a proposta.
Regra 2	SE existe a solicitação de uma proposta, E o cadastro não está atualizado, ENTÃO finaliza a proposta.
Regra 3	SE existe a solicitação de uma proposta, E existe capacidade de pagamento, ENTÃO segue a proposta.
Regra 4	SE existe a solicitação de uma proposta, E não existe capacidade de pagamento, ENTÃO finaliza a proposta.
Regra 5	SE existe a solicitação de uma proposta, E existem restrições relevantes, ENTÃO finaliza a proposta.

Regra 6	SE existe a solicitação de uma proposta, E não existem restrições relevantes, ENTÃO segue a proposta.
---------	--

Fonte: elaborado pelo autor, 2016

O Quadro 3 ilustra uma base de dados, criados hipoteticamente para ilustrar os dados de uma instituição financeira, onde se descrevem três atributos previsores e um atributo meta. Tal base de dados apresenta as seguintes características: 14 exemplos, 4 atributos e 2 classes, onde n é o número de exemplos, q é o número de atributos e m é o número de classes, lendo $n=14$, $q=4$ e $m=2$.

QUADRO 3: Conjunto de Teste Previamente Ordenado

	ATRIBUTOS			
Instâncias	PREVISORES			META
	Tipo de Operação de Crédito	Produto	Garantia Real	Inadimplência
1	Rotativo	Seguros	Não	Sim
2	Rotativo	Seguros	Não	Não
3	Rotativo	Seguros	Sim	Sim
4	Rotativo	Consórcio	Não	Não
5	Rotativo	Consórcio	Sim	Não
6	TD	Seguros	Não	Sim
7	TD	Seguros	Sim	Não
8	TD	Consórcio	Sim	Sim
9	TD	Consórcio	Não	Sim
10	Financiamento	Seguros	Não	Não
11	Financiamento	Seguros	Não	Sim
12	Financiamento	Seguros	Sim	Não
13	Financiamento	Consórcio	Sim	Não
14	Financiamento	Consórcio	Não	Não
LEGENDA	R=Rotativo	S=Seguros	S=Sim	S=Sim
	T=T	C=Consórcio	N=Não	N=Não
	F=Financiamento			

Fonte: elaborado pelo autor, 2016.

O processo de criação da árvore de decisão, a partir de um conjunto contendo n exemplos de teste $T = \{t_1, \dots, t_n\}$, onde cada t corresponde a um exemplo de teste; um conjunto contendo q atributos $A = \{a_1, \dots, a_q\}$, onde cada a é um atributo; um conjunto contendo m classes $C = \{c_1, \dots, c_m\}$, onde cada c é classe que rotula um exemplo t . Todo este processo ocorre de acordo com o algoritmo descrito a seguir na Figura 7:

FIGURA 7 - Algoritmo de Criação de Arvore de Decisão

```

Entradas:   uma base de dados  $T$ 
              um conjunto de atributos  $A$ 
              um conjunto de classes  $C$ 
Saída:     uma árvore de decisão  $D$ 

1 procedimento C4.5(  $T, A, D, C$  )
3 cria um nó  $\alpha$  de decisão e adiciona-o à árvore  $D$ 
4 se todas as instâncias de  $T$  pertencem a uma classe  $c \in C$ 
5 então
6   atribui ao nó  $\alpha$  uma única folha com classe  $c \in C$ 
7 senão
8   se  $A$  é um conjunto unitário
9   então
10    atribui ao nó  $\alpha$  uma única folha identificando o valor
        mais comum do atributo preditor.
11    senão
12     calcula o ganho de informação de cada um dos atributos de
         $A$ 
13     se um dos atributos de  $A$  possui ganho de informação médio
        maior que os demais
14     então
15       define  $at$  o atributo com maior ganho
16       para cada valor  $v$  do atributo  $at$  faça
17         adiciona uma sub-árvore  $d$  ao nó apontado por  $D$ 
18         define  $T_v$  como subconjunto de  $T$ , onde  $at = v$ 
19         C4.5(  $T_v, A - \{at\}, d, C$  )
20       senão
21         atribui ao nó apontado por  $D$  uma única folha
        ligando o valor mais comum do atributo preditor.
22 fim.

```

Na sequência serão apresentados, passo a passo, os diferentes cálculos necessários para obter um classificador na forma de uma árvore de decisão (C4.5). O processo começa pela busca da informação que melhor codifica uma classificação de um determinado caso pertencente a um conjunto de teste T . Em termos matemáticos, esta informação é determinada pelo cálculo da entropia.

A entropia é calculada por meio da frequência dos exemplos que resultam em casos positivos p_i , presentes no conjunto de teste, multiplicando tal frequência pelo logaritmo na base dois destes exemplos positivos e somando com os exemplos negativos. Estes últimos são obtidos da mesma forma. O resultado é a quantidade de informação

necessária para codificar a classificação de um determinado caso pertencente ao conjunto de teste T (MITCHELL, 1997 apud MANFRÉ, 2010).

Na Tabela 2 estão ordenadas as fórmulas utilizadas para classificação pelo método C4.5 na forma de árvore de decisão.

Tabela 2 - Fórmulas para Classificar no Formato de Árvore de Decisão

$Entropia(T) = -\sum_c^C p_c \log_2 p_c$	(1)
$Entropia(T_{a=v}) = -\sum_c^C p_{c,v} \log_2 p_{c,v}$	(2)
$info(T_a) = \sum_{v \in \text{dom}(a)} \frac{ T_{a=v} }{ T } \times Entropia(T_{a=v})$	(3)
$ganho(T_a) = entropia(T) - info(T_a)$	(4)
$infoDiv(T_{a_i}) = -\sum_{i=1}^n \frac{ T_{a_i} }{ T } \times \log_2 \left(\frac{ T_{a_i} }{ T } \right)$	(5)
$ganhoMedio(T_a) = \frac{ganho(T_a)}{infoDiv(T_a)}$	(6)

Para o conjunto de atributos T, o cálculo da entropia é dado pela Fórmula (1) da Tabela 2, onde p_c é a razão definida pelo número de instâncias que pertencem à classe c e número total de instâncias da base dados.

Tomando o Quadro 3 **Erro! Fonte de referência não encontrada.** como o conjunto de teste T, tem-se a seguir o cálculo do valor da entropia. As classes são definidas por $C = \{\text{sim}, \text{não}\}$. O valor 14 nos diferentes denominadores corresponde ao número de exemplos de T. Os valores 6 e 8 correspondem respectivamente aos números de exemplos que possuem como classe sim e não.

$$Entropia(T) = -\frac{6}{14} \times \log_2 \left(\frac{6}{14} \right) - \frac{8}{14} \times \log_2 \left(\frac{8}{14} \right) = 0.983$$

O cálculo da entropia para um dado atributo de predição a para o valor v é dado pela Fórmula (2) da Tabela 2, onde $p_{c,v}$ é a razão entre o número de instâncias que pertencem a cada classe $c \in C$ para um dado valor v do atributo de predição a . Caso uma dada razão seja nula, ou seja, quando nenhuma instância possuir o valor v , assume-se a entropia nula.

$$Entropia(T_{Operação=Rotativo}) = \left(-\frac{2}{5} \times \log_2 \left(\frac{2}{5} \right) - \frac{3}{5} \times \log_2 \left(\frac{3}{5} \right) \right) = 0,971$$

$$Entropia(T_{Operação=TD}) = \left(-\frac{3}{4} \times \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \times \log_2\left(\frac{1}{4}\right) \right) = 0,811$$

$$Entropia(T_{Operação=Financiamento}) = \left(-\frac{4}{5} \times \log_2\left(\frac{4}{5}\right) - \frac{1}{5} \times \log_2\left(\frac{1}{5}\right) \right) = 0,722$$

$$Entropia(T_{Produto=Consórcio}) = \left(-\frac{2}{6} \times \log_2\left(\frac{2}{6}\right) - \frac{4}{6} \times \log_2\left(\frac{4}{6}\right) \right) = 0,918$$

$$Entropia(T_{Produto=Seguro}) = \left(-\frac{4}{8} \times \log_2\left(\frac{4}{8}\right) - \frac{4}{8} \times \log_2\left(\frac{4}{8}\right) \right) = 1$$

$$Entropia(T_{GarantiaReal=sim}) = \left(-\frac{2}{6} \times \log_2\left(\frac{2}{6}\right) - \frac{4}{6} \times \log_2\left(\frac{4}{6}\right) \right) = 0,918$$

$$Entropia(T_{GarantiaReal=não}) = \left(-\frac{4}{8} \times \log_2\left(\frac{4}{8}\right) - \frac{4}{8} \times \log_2\left(\frac{4}{8}\right) \right) = 1$$

O critério de ganho de informação é a maneira mais eficaz de dividir o conjunto de teste T e selecionar o atributo que irá dividir a informação do conjunto de teste. Tal critério consiste em medir a frequência $\text{freq}(C_j, S)$, na qual um conjunto qualquer de exemplos S pertence à mesma classe C_j , sendo que $|S|$ é o número de exemplos em S. Tomando como exemplo a base de dados do Quadro2, a frequência de exemplos positivos (cujo valor da classe é “sim”) é 6/14 ou 0,428. Já a frequência de exemplos negativos (valor da classe “não”) é 8/14 ou 0,571.

O critério de ganho de informação, mostrado na Fórmula (4) da Tabela 2, utilizado na seleção de um teste que maximiza o ganho de informação, é medido pela entropia menos a informação de T do atributo a. A informação do atributo a é medida de acordo com a Fórmula (3) da Tabela 2, onde o conjunto de dados de teste T é particionado em subconjuntos, de acordo com o conjunto de valores $v \in \text{dom}(a)$ que é domínio do atributo a, sendo $T_a=v$ o subconjunto formado por exemplos que possuem valor v para o atributo a.

Pode-se, desta forma, calcular o ganho para cada atributo da base de dados, escolhendo assim o atributo que melhor separa por si só as informações. O exemplo a seguir mostra o ganho calculado em detalhes para o atributo Operação e também os valores do ganho de informação dos atributos Produto, e Garantia Real.

$$\begin{aligned} \text{info}(T_{\text{Operação}}) &= \frac{5}{14} \times \text{Entropia}(T_{\text{Operação=Rotativo}}) + \frac{4}{14} \times \text{Entropia}(T_{\text{Operação=TD}}) + \\ &\frac{5}{14} \times \text{Entropia}(T_{\text{Operação=Financiamento}}) = 0,836 \end{aligned}$$

$$\text{info}(T_{\text{Produto}}) = \frac{6}{14} \times \text{Entropia}(T_{\text{Produto=Consórcio}}) + \frac{8}{14} \times \text{Entropia}(T_{\text{Produto=Seguro}}) = 0,964$$

$$\text{info}(T_{\text{GarantiaReal}}) = \frac{6}{14} \times \text{Entropia}(T_{\text{GarantiaReal=sim}}) + \frac{8}{14} \times \text{Entropia}(T_{\text{GarantiaReal=não}}) = 0,964$$

Após calcular a informação para todos os atributos, o próximo passo é calcular o ganho de informação para os mesmos.

$$\text{ganho}(T, \text{Operação}) = \text{entropia}(T) - \text{info}(T, \text{Operação})$$

$$\text{ganho}(T, \text{Operação}) = 0,983 - 0,835 = 0,147$$

$$\text{ganho}(T_{\text{Produto}}) = \text{entropia}(T) - \text{info}(T_{\text{Produto}})$$

$$\text{ganho}(T_{\text{Produto}}) = 0,983 - 0,964 = 0,018$$

$$\text{ganho}(T_{\text{Garantia Real}}) = \text{entropia}(T) - \text{info}(T_{\text{Garantia Real}})$$

$$\text{ganho}(T_{\text{Garantia Real}}) = 0,983 - 0,964 = 0,018$$

O atributo X, que possuir o valor mais alto de ganho(X), é então escolhido como sendo a raiz da árvore, neste caso, o atributo operação. O critério de ganho possui um bias que favorece atributos com muitos valores possíveis. Quando há diferentes funções que possuem valores em comum para um determinado exemplo, deve haver a seleção de uma função que se encaixe para ambos os pontos a priori, tal seleção do valor a priori é chamada bias. Sem o valor do bias, um aprendizado útil se torna impossível (NILSSON, 1996). Tal fato ocorre quando há um atributo que possui valor único para cada instância do conjunto de dados, ocorrendo uma divisão que gera inúmeros subconjuntos com apenas um valor possível para cada um deles, onde o valor de info(S) é igual a zero, portanto, mínimo. A correção para este problema consiste em ajustar os resultados possíveis, sendo que a potencial divisão da informação utiliza em seu cálculo a subtração do conjunto de teste T e de cada subconjunto n, conforme Fórmula (5) da Tabela 2. A

divisão da informação resulta em informações úteis na classificação, eliminando assim atributos que podem por si só classificar amostras.

$$\text{infoDiv}(T_{\text{Operação}}) = \underbrace{-\frac{5}{14} \times \log_2\left(\frac{|5|}{|14|}\right)}_{\text{Rotativo}} + \underbrace{\frac{4}{14} \times \log_2\left(\frac{|4|}{|14|}\right)}_{\text{TD}} + \underbrace{\frac{5}{14} \times \log_2\left(\frac{|5|}{|14|}\right)}_{\text{Financiamento}} = 1,577$$

$$\text{infoDiv}(T_{\text{Garantia Real}}) = \underbrace{-\frac{6}{14} \times \log_2\left(\frac{|6|}{|14|}\right)}_{\text{sim}} + \underbrace{\frac{8}{14} \times \log_2\left(\frac{|8|}{|14|}\right)}_{\text{não}} = 0,985$$

$$\text{infoDiv}(T_{\text{Produto}}) = \underbrace{-\frac{6}{14} \times \log_2\left(\frac{|6|}{|14|}\right)}_{\text{Consórcio}} + \underbrace{\frac{8}{14} \times \log_2\left(\frac{|8|}{|14|}\right)}_{\text{Seguro}} = 0,985$$

A proporção de ganho de um atributo a_i é medida com base na informação relevante, Fórmula 6 da Tabela 2:

$$\text{ganhoMedio}(T_{\text{Operação}}) = \frac{\text{ganho}(T_{\text{Operação}})}{\text{infoDiv}(T_{\text{Operação}})} = \frac{0,147}{1,577} = 0,093$$

$$\text{ganhoMedio}(T_{\text{Produto}}) = \frac{\text{ganho}(T_{\text{Produto}})}{\text{infoDiv}(T_{\text{Produto}})} = \frac{0,018}{0,985} = 0,018$$

$$\text{ganhoMedio}(T_{\text{Garantia Real}}) = \frac{\text{ganho}(T_{\text{Garantia Real}})}{\text{infoDiv}(T_{\text{Garantia Real}})} = \frac{0,018}{0,985} = 0,018$$

Cada valor entre parênteses corresponde ao número de exemplos bem classificados/número de exemplos mal classificados. O critério de ganho de informação que apresentar o maior valor é o atributo do nó raiz da árvore, neste caso o atributo Operação. Após a identificação do nó raiz, este atributo recebe seus valores correspondentes, no caso, Rotativo, TD e Financiamento, e para cada valor serão novamente, calculados a entropia e o ganho de informação dos atributos Produto e

Garantia Real. Onde para Operação com o valor Rotativo, ou seja, para o subconjunto de instâncias 1, 2, 3, 4 e 5, obtêm-se os resultados:

$$Entropia(T_{Operação \Rightarrow Rotativo}) = -\frac{2}{5} \times \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \times \log_2\left(\frac{3}{5}\right) = 0,971$$

$$Entropia(T_{q \Rightarrow r, Produto=Seguro}) = \left(-\frac{2}{2} \times \log_2\left(\frac{2}{2}\right)\right) = 0$$

$$Entropia(T_{q \Rightarrow r, Produto=Consórcio}) = \left(-\frac{2}{3} \times \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) = 0,918$$

$$Entropia(T_{q \Rightarrow r, Garantia Real=sim}) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$Entropia(T_{q \Rightarrow r, Garantia Real=não}) = \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right) + \frac{2}{3} \times \log_2\left(\frac{2}{3}\right)\right) = 0,918$$

$$info(T_{q \Rightarrow r, Produto}) = \frac{2}{5} \times Entropia(T_{q \Rightarrow r, Produto=Consórcio}) +$$

$$\frac{3}{5} \times Entropia(T_{q \Rightarrow r, Produto=Seguro}) = 0,550$$

$$info(T_{q \Rightarrow r, Garantia Real}) = \frac{2}{5} \times Entropia(T_{q \Rightarrow r, Garantia Real=sim}) +$$

$$\frac{3}{5} \times Entropia(T_{q \Rightarrow r, Garantia Real=não}) = 0,950$$

$$ganho(T_{q \Rightarrow r, operação=produto}) = entropia(T_{operação \Rightarrow r}) - info(T_{q \Rightarrow r, operação=produto})$$

$$ganho(T_{q \Rightarrow r, operação=produto}) = 0,971 - 0,550 = 0,421$$

$$ganho(T_{q \Rightarrow r, operação=garantia real}) = entropia(T_{operação \Rightarrow r}) - info(T_{q \Rightarrow r, operação=garantia real})$$

$$ganho(T_{q \Rightarrow r, operação=garantia real}) = 0,971 - 0,950 = 0,021$$

$$infoDiv(T_{q \Rightarrow r, garantia real}) = \underbrace{-\frac{2}{5} \times \log_2\left(\frac{|2|}{|5|}\right)}_{sim} + \underbrace{\frac{3}{5} \times \log_2\left(\frac{|3|}{|5|}\right)}_{não} = 0,970$$

$$\text{infoDiv}(T_{q \Rightarrow r, \text{produto}}) = \underbrace{-\frac{2}{5} \times \log_2\left(\frac{|2|}{|5|}\right)}_{\text{consórcio}} + \underbrace{\frac{2}{5} \times \log_2\left(\frac{|2|}{|5|}\right)}_{\text{seguro}} = 0,970$$

$$\text{ganhoMedio}(T_{q \Rightarrow r, \text{produto}}) = \frac{\text{ganho}(T_{q \Rightarrow r, \text{operação}=\text{produto}})}{\text{infoDiv}(T_{q \Rightarrow r, \text{produto}})} = \frac{0,421}{0,970} = 0,434$$

$$\text{ganhoMedio}(T_{q \Rightarrow r, \text{garantia real}}) = \frac{\text{ganho}(T_{q \Rightarrow r, \text{operação}=\text{garantia real}})}{\text{infoDiv}(T_{q \Rightarrow r, \text{garantia real}})} = \frac{0,021}{0,970} = 0,022$$

De acordo com os resultados dos cálculos, o atributo Produto apresentou um maior ganho de informação. Portanto, este atributo será o nó de decisão para o valor de Operação Rotativo. Em seguida, será continuada a sequência do atributo Produto, para depois retornar aos outros valores do atributo Operação.

Para o atributo Produto com valor de Seguro, que é o subconjunto de instâncias 1, 2 e 3, será incluída uma folha com o valor mais comum do atributo preditor, que no caso será uma folha com valor Sim.

Enquanto para o atributo Produto com o valor de Consórcio, referentes às instâncias 4 e 5, será inserida uma folha com o valor comum a todas as instâncias deste subconjunto, o valor Não.

Concluída a sequência para o atributo Produto, retorna-se ao atributo Operação e calcula-se a entropia e o ganho de informação para os atributos Produto e Garantia Real. Onde para o atributo Operação com valor de TD, ou subconjunto 6, 7, 8 e 9, encontramos:

$$\text{Entropia}(T_{\text{operação}=\text{TD}}) = -\frac{3}{4} \times \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \times \log_2\left(\frac{1}{4}\right) = 0,811$$

$$\text{Entropia}(T_{q \Rightarrow t, \text{produto}=\text{consórcio}}) = \left(-\frac{2}{2} \times \log_2\left(\frac{2}{2}\right)\right) = 0$$

$$\text{Entropia}(T_{q \Rightarrow td, \text{produto}=\text{seguro}}) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropia}(T_{q \Rightarrow td, \text{garantia real}=\text{sim}}) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropia}(T_{q \Rightarrow td, \text{garantia real}=\text{não}}) = \left(-\frac{2}{2} \times \log_2\left(\frac{2}{2}\right)\right) = 0$$

$$\begin{aligned} \text{info}(T_{q \Rightarrow td, \text{produto}}) &= \frac{2}{4} \times \text{Entropia}(T_{q=td, \text{produto}=\text{consórcio}}) + \\ &\frac{2}{4} \times \text{Entropia}(T_{q \Rightarrow td, \text{produto}=\text{seguro}}) = 0,500 \end{aligned}$$

$$\begin{aligned} \text{info}(T_{q \Rightarrow td, \text{garantia real}}) &= \frac{2}{4} \times \text{Entropia}(T_{q \Rightarrow td, \text{garantia real}=\text{sim}}) + \\ &\frac{2}{4} \times \text{Entropia}(T_{q \Rightarrow td, \text{garantia real}=\text{não}}) = 0,500 \end{aligned}$$

$$\text{ganho}(T_{q \Rightarrow td, \text{operação}=\text{produto}}) = \text{entropia}(T_{\text{operação}=\text{TD}}) - \text{info}(T_{q \Rightarrow td, \text{produto}})$$

$$\text{ganho}(T_{q \Rightarrow td, \text{operação}=\text{produto}}) = 0,811 - 0,500 = 0,311$$

$$\text{ganho}(T_{q \Rightarrow td, \text{operação}=\text{garantia real}}) = \text{entropia}(T_{\text{operação}=\text{TD}}) - \text{info}(T_{q \Rightarrow td, \text{garantia real}})$$

$$\text{ganho}(T_{q \Rightarrow td, \text{operação}=\text{garantia real}}) = 0,811 - 0,500 = 0,311$$

$$\text{infoDiv}(T_{q \Rightarrow td, \text{garantia real}}) = \underbrace{-\frac{2}{4} \times \log_2\left(\frac{|2|}{|4|}\right)}_{\text{sim}} + \underbrace{\frac{2}{4} \times \log_2\left(\frac{|2|}{|4|}\right)}_{\text{não}} = 1$$

$$\text{infoDiv}(T_{q \Rightarrow td, \text{produto}}) = \underbrace{-\frac{2}{4} \times \log_2\left(\frac{|2|}{|4|}\right)}_{\text{consórcio}} + \underbrace{\frac{2}{4} \times \log_2\left(\frac{|2|}{|4|}\right)}_{\text{seguro}} = 1$$

$$\text{ganhoMedio}(T_{q \Rightarrow td, \text{produto}}) = \frac{\text{ganho}(T_{q \Rightarrow td, \text{operação}=\text{produto}})}{\text{infoDiv}(T_{q \Rightarrow td, \text{produto}})} = \frac{0,311}{1} = 0,311$$

$$\text{ganhoMedio}(T_{q \Rightarrow td, \text{garantia real}}) = \frac{\text{ganho}(T_{q \Rightarrow td, \text{operação}=\text{garantia real}})}{\text{infoDiv}(T_{q \Rightarrow td, \text{garantia real}})} = \frac{0,311}{1} = 0,311$$

Como podemos observar, o valor do ganho de informação para os atributos Produto e Garantia Real são os mesmos, impossibilitando a escolha de um atributo por maior ganho de informação. Será, portanto, realizada uma poda, a fim de simplificar a

árvore, incluindo uma folha com o valor mais frequente para o atributo Operação com valor TD, o valor Sim.

E para o atributo Operação com valor de Financiamento os valores de entropia e ganho de informação, são os seguintes:

$$Entropia(T_{\text{operação=financiamento}}) = -\frac{1}{5} \times \log_2\left(\frac{1}{5}\right) - \frac{4}{5} \times \log_2\left(\frac{4}{5}\right) = 0,721$$

$$Entropia(T_{q \Rightarrow f, \text{produto=consórcio}}) = \left(-\frac{2}{2} \times \log_2\left(\frac{2}{2}\right) \right) = 0$$

$$Entropia(T_{q \Rightarrow f, \text{produto=seguro}}) = \left(-\frac{2}{3} \times \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \times \log_2\left(\frac{1}{3}\right) \right) = 0,918$$

$$Entropia(T_{q \Rightarrow f, \text{garantia real=sim}}) = \left(-\frac{2}{2} \times \log_2\left(\frac{2}{2}\right) \right) = 0$$

$$Entropia(T_{q \Rightarrow f, \text{garantia real=não}}) = \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \times \log_2\left(\frac{2}{3}\right) \right) = 0,918$$

$$info(T_{q \Rightarrow f, \text{produto}}) = \frac{2}{5} \times Entropia(T_{q \Rightarrow f, \text{produto=consórcio}}) + \frac{3}{5} \times Entropia(T_{q \Rightarrow f, \text{produto=seguro}}) = 0,551$$

$$info(T_{q \Rightarrow f, \text{garantia real}}) = \frac{2}{5} \times Entropia(T_{q \Rightarrow f, \text{garantia real=sim}}) + \frac{3}{5} \times Entropia(T_{q \Rightarrow f, \text{garantia real=não}}) = 0,551$$

$$ganho(T_{q \Rightarrow f, \text{operação=produto}}) = entropia(T_{\text{operação=financiamento}}) - info(T_{q \Rightarrow f, \text{operação=produto}})$$

$$ganho(T_{q \Rightarrow f, \text{operação=produto}}) = 0,721 - 0,550 = 0,171$$

$$ganho(T_{q \Rightarrow f, \text{operação=garantia real}}) = entropia(T_{\text{operação=financiamento}}) - info(T_{q \Rightarrow f, \text{operação=garantia real}})$$

$$ganho(T_{q=f, \text{operação=garantia real}}) = 0,721 - 0,550 = 0,171$$

$$infoDiv(T_{q=f, \text{garantia real}}) = \underbrace{-\frac{2}{5} \times \log_2\left(\frac{|2|}{|5|}\right)}_{\text{sim}} + \underbrace{\frac{3}{5} \times \log_2\left(\frac{|3|}{|5|}\right)}_{\text{não}} = 0,971$$

$$infoDiv(T_{q \Rightarrow f, \text{produto}}) = \underbrace{-\frac{2}{5} \times \log_2\left(\frac{|2|}{|5|}\right)}_{\text{consórcio}} + \underbrace{\frac{3}{5} \times \log_2\left(\frac{|3|}{|5|}\right)}_{\text{seguro}} = 0,971$$

$$\text{ganhoMedio}(T_{q \Rightarrow f, \text{produto}}) = \frac{\text{ganho}(T_{q \Rightarrow f, \text{operação=produto}})}{\text{infoDiv}(T_{q \Rightarrow f, \text{produto}})} = \frac{0,171}{0,971} = 0,176$$

$$\text{ganhoMedio}(T_{q \Rightarrow f, \text{garantia real}}) = \frac{\text{ganho}(T_{q \Rightarrow f, \text{operação=garantia real}})}{\text{infoDiv}(T_{q \Rightarrow f, \text{garantia real}})} = \frac{0,171}{0,971} = 0,176$$

Como no exemplo anterior, os atributos Produto e Garantia real apresentaram o mesmo ganho de informação para o subconjunto de operação com valor financiamento, optando-se por incluir uma folha com o valor mais comum nestas instâncias: 10, 11, 12, 13 e 14, que é o valor Não.

Finalizando a apresentação dos conceitos teóricos envolvidos nesta pesquisa, sequencialmente será apresentado todo o procedimento a ser utilizado como metodologia na construção dos resultados relacionados ao problema de pesquisa.

3 METODOLOGIA

A pesquisa se caracteriza como um estudo de caráter descritivo e a estratégia utilizada foi o estudo de caso. Foram utilizados essencialmente dados secundários, oriundos de fontes internas à cooperativa Alfa. As fontes foram a base de dados dos cooperados da cooperativa Alfa, além de documentos internos e normas da legislação que rege o cooperativismo de crédito.

As técnicas de análise utilizadas foram: análise documental (MARCONI; LAKATOS, 2003 p. 177) e análises estatísticas e matemáticas do banco de dados dos cooperados. Especificamente, foi realizada a mineração dos dados da base que inclui todas as movimentações realizadas pelos cooperados da cooperativa Alfa no ano de 2015. Para realizar a descoberta de conhecimentos, a partir da base de dados de cooperados, utilizou-se um indutor de árvores de decisão. Tal indutor fornece uma saída simbólica compreensível não apenas pela máquina.

3.1 ROTEIRO DA PESQUISA

Iniciando a análise dos dados apresentados pela cooperativa Alfa utilizou-se a estatística descritiva no sentido de descrever de forma sumarizada as informações. A estatística descritiva sumariza vários aspectos sobre a base de dados, desvendando detalhes da amostra e fornece informações sobre a população de cada amostra desenhada (LARSON, 2006).

A base de dados dos associados foi dividida por personalidade civil em dois grupos denominados pessoa física e pessoa jurídica. É denominado pessoa física todo ser humano desde seu nascimento segundo o código civil brasileiro. A pessoa jurídica, segundo Sandroni (2003), é qualquer instituição (empresa, sociedade, corporação, etc) que se personaliza e individualiza, distinguindo-se das pessoas físicas que a formam.

O segundo passo, ainda no campo estatístico, foi a análise de correlação, que mede o grau de associação entre X e Y, variável dependente e independente, de toda a movimentação dos associados. Ressalta-se que esta medida não apresenta relação causa-efeito. Porém, em estudos que envolvem duas ou mais variáveis, é comum o interesse em conhecer o relacionamento entre elas, além das estatísticas descritivas normalmente calculadas (LIRA, 2004).

Após as duas análises estatísticas descritas acima foram aplicadas as etapas do processo de descoberta de conhecimento; as primeiras cinco etapas estão relacionadas à preparação dos dados e à descoberta de conhecimento propriamente dita, e as demais etapas são relativas à validação e interpretação dos achados. As etapas são as seguintes:

A) **Limpeza dos dados:** etapa em que são eliminados ruídos e dados inconsistentes. Nesta pesquisa, limitou-se basicamente em converter os valores dos atributos binários simbólicos em valores numéricos, sendo estes 0 e 1.

B) **Integração de dados:** esta é uma etapa na qual diferentes fontes de dados podem ser combinadas produzindo um único repositório. No caso desta pesquisa a realização desta etapa não se fez necessária pois a base de dados dos associados fornecida pela cooperativa Alfa era única, ou seja, sem necessidade de integração com outras fontes. Porém, em futuras pesquisas seria possível agregar bases de dados externas à cooperativa.

C) **Seleção:** esta etapa é quando o pesquisador seleciona os atributos que o interessam. Por exemplo, o pesquisador pode decidir que algumas informações são menos relevantes que outras na qualificação de um cliente - informações como o CEP (código de endereçamento postal) de um determinado endereço, grau de parentesco, entre outros.

Os itens selecionados na base de dados dos associados da cooperativa Alfa estão descritos no Quadro 4.

QUADRO 4 – Atributos selecionados na base de dados

NU1COOPERADO	Número do código interno do associado;
SEXO	Masculino e feminino no caso de pessoa física; Pessoa jurídica não se aplica;
IDADE	Reflete a idade da pessoa física no caso da pessoa jurídica tempo de abertura da empresa;
DIA	Dia do atributo idade;
MES	Mês do atributo idade;
ANO	Ano do atributo idade;
QCC	Quantidade de contas correntes que associado possui;
QCCU11ITE	Quantidade de contas correntes com limite de cheque especial para pessoa física ou conta garantida para pessoa jurídica;
CAPITAL	Saldo do capital integralizado de cada associado;
DAP	Valor de investimentos do associado;
DAV	Saldo do capital integralizado de cada associado;
U11ITEROTATIVO	Valor do limite de crédito rotativo;
U11ITEROTATIVOUSADO	Saldo em utilização do limite de crédito rotativo;
U11ITETD	Valor do limite de crédito para desconto de recebíveis;
U11ITECC	Valor do limite em conta-corrente, cheque especial para pessoa física e conta garantida para pessoa jurídica;
U11ITECCUSADO	Saldo em utilização do limite em conta-corrente, cheque especial para pessoa física e conta garantida para pessoa jurídica;
ADP	Saldo de crédito acima do limite em conta-corrente, cheque especial para pessoa física e conta garantida para pessoa jurídica;
INANCIA1ENTO	Saldo de financiamentos;
CONSIGNADO	Saldo de crédito consignado em folha de pagamento, privado ou público;
DE1AISCREDITO	Saldo dos demais crédito utilizados;
CREDITORURAL	Saldo de crédito tomado em operações de rural;
PREAPROVADO	Saldo em utilização de crédito pré-aprovado;
ICROREDITO	Saldo de microcrédito;
CONVENIOAGUA	Possui conta de água cadastrado como débito automático;
CONVENIOLUZ	Possui conta de energia elétrica cadastrado como débito automático;
CONVENIOOONE	Possui conta de telefone cadastrado como débito automático;
CONVENIO	Possui demais convênios cadastrados como débito automático;
COBRANCACREG	Possui boletos de cobrança com registro;
COBRANCASREG	Possui boletos de cobrança sem registro;
SEGUROVIDA	Possui seguro de vida;
SEGUROAUTO	Possui seguro de automóveis;
SEGUORRESIDENCIAL	Possui seguro residencial;
SEGURODIVERSO	Possui demais tipos de seguro;
CARTAOEBITO	Possui cartão de débito;
CARTAOREDITO	Possui cartão de crédito;
CONSAUTO	Possui consórcio de automóveis;
CONS1OTO	Possui consórcio de motocicleta;
CON1OVEL	Possui consórcio de imóvel;
RETABILIDADE	Valor da rentabilidade do associado;
CNAE	Código de Cadastro Nacional de Atividade Econômica;
INADIMPLENTE	Possui inadimplimento;

Fonte: elaborado pelo autor, 2016.

Ressalta-se que a cooperativa Alfa forneceu uma base de dados do ano de 2016 em 8 diferentes tempos, sendo que estas bases foram classificadas cronologicamente para efeito das análises como tempo1, tempo2, tempo3, tempo4, tempo5, tempo6, tempo7 e tempo8.

D) **Transformação dos dados:** etapa em que os dados são transformados em um formato apropriado para aplicação de algoritmos de mineração (por exemplo, através de operações de agregação).

Neste estudo optou-se em utilizar o algoritmo de classificação *k-means* por ser o mais adequado ao problema de pesquisa apresentado. Assim, Gouvêa, Prearo e Romeiro (2013) afirmam que a escolha dos métodos e dos tipos de análises empregadas nos trabalhos científicos deve ser determinada pelo problema de pesquisa.

Tendo em vista a avaliação realizada pelos autores acima citados, evidencia-se que *k-means* é utilizado para amostra grandes (acima de 250 casos), sendo compatível então com base de dados da cooperativa Alfa.

Percebe-se ainda que, comparando os quadros dos objetivos de aplicação de técnicas elaborados pelos autores, as técnicas de análise de conglomerado (*clusters*) são mais próximas aos objetivos deste estudo do que as técnicas de correspondência, já que a primeira tem por objetivo detectar grupos homogêneos, agrupar sujeitos, agrupar indivíduos semelhantes, em consonância com a teoria apresentada sobre segmentação.

A seguir apresentam-se os quadros de objetivos das análises de conglomerado e correspondência.

QUADRO 5 – Objetivo de aplicação da técnica de Análise de Conglomerado

Objetivo	Descrição	Fontes
Agrupamento	Detectar grupos homogêneos nos dados, podendo os grupos ser constituídos por variáveis ou casos.	Pestana e Gageiro (2000)
	Agrupar sujeitos ou variáveis em grupos homogêneos a partir de uma ou mais características comuns.	Maroco (2003)
	Agregar objetivos com base nas características que eles possuem	Hair <i>et al.</i> (2005)
	Agrupar indivíduos semelhantes de acordo com a similaridade das variáveis em estudo.	Mingoti (2005)
	Identificar um conjunto de grupos que minimize a variação dentro do grupo e maximize a variação entre os grupos.	Garson (2007b)

Fonte: Gouvêa, Prearo e Romeiro (2013)

QUADRO 6 – Objetivo de aplicação da técnica de Análise de Correspondência.

Categoria	Principal Objetivo Indicado no Estudo	Fontes
Agrupamento	Encontrar e representar as relações existentes entre as variáveis em um espaço multidimensional, reduzindo a dimensão de variáveis com dados qualitativos para obter um pequeno número de fatores.	Pestana e Gageiro (2000)
	Descobrir relações existentes entre variáveis nominais a partir de uma tabela de correspondências colocadas em um espaço de poucas dimensões.	SPSS (2001)
	Retratar a correspondência de categorias de variáveis, por meio de mapas percentuais.	Hair <i>et al.</i> (2005)
	Fatorar variáveis categóricas e as exibir em um espaço de propriedades que apresente a associação dessas variáveis em duas ou mais dimensões.	Garson (2007b)

Fonte: Gouvêa, Prearo e Romeiro (2013)

E) **Mineração:** Nesta etapa do processo, que é fundamental, aplicam-se as técnicas para a extração dos padrões de interesse, caso existam. Para este fim foi aplicado o algoritmo C 4.5 após a geração dos classificadores por meio do método *k-means*.

F) **Avaliação ou Pós-processamento:** etapa na qual são identificados os padrões interessantes de acordo com algum critério do usuário. Após o processamento do algoritmo C 4.5, apresenta-se o relatório de validação do processamento e pode-se observar que ocorreu uma correta classificação de 98,50% das instâncias, no caso da base de pessoa física, e de 97,07% na pessoa jurídica. Quanto à acurácia, esta foi identificada pelo indicador Kappa, que foi de 0.97 na pessoa física e de 0.95 na pessoa jurídica, ficando próximo de 1 evidenciando a precisão dos conhecimentos obtidos a partir da base de dados, pois quanto mais próximo de 1 melhor para este indicador. Na sequência apresenta-se os quadros 7 e 8 com os resultados citados:

QUADRO 7– Relatório de Desempenho Classificação do C 4.5 – Pessoa Física

```

=== Stratified cross-validation ===
Correctly Classified Instances      15751      98.5053 %
Incorrectly Classified Instances    239        1.4947 %
Kappa statistic                    0.9713
Mean absolute error                0.0129
Root mean squared error            0.0971
Relative absolute error            3.7013 %
Root relative squared error        23.2782 %

Total Number of Instances          15990

=== Detailed Accuracy By Class ===

MCC      ROC Area      TP Rate  FP Rate  Precision  Recall  F-Measure
0,976    0,990          0,990   0,015   0,987     0,990   0,989
0,972    0,988          0,985   0,013   0,983     0,985   0,984
0,919    0,954          0,882   0,001   0,963     0,882   0,921
Weighted Avg. 0,985   0,013   0,985     0,985   0,985
0,973    0,988          0,976

=== Confusion Matrix ===

  a    b    c  <-- classified as
8483  80    4  |   a = G1
  92 6880  11  |   b = G2
  16  36  388  |   c = G3

```

Fonte: Extraído sistema R.

QUADRO 8 – Relatório de Desempenho Classificação do C 4.5 – Pessoa Jurídica

```

=== Stratified cross-validation ===

Correctly Classified Instances      6933      97.0736 %
Incorrectly Classified Instances    209      2.9264 %
Kappa statistic                     0.9516
Mean absolute error                 0.023
Root mean squared error             0.135
Relative absolute error              5.6744 %
Root relative squared error         30.0148 %
Total Number of Instances          7142

=== Detailed Accuracy By Class ===

MCC      ROC Area  TP Rate  FP Rate  Precision  Recall  F-Measure
          PRC Area  Class
0,909    0,953    0,911    0,012    0,936      0,911    0,923
0,965    0,985    0,989    0,025    0,977      0,989    0,983
0,964    0,982    0,972    0,010    0,979      0,972    0,975
Weighted Avg. 0,971    0,018    0,971    0,971    0,971
0,955    0,978    0,951

=== Confusion Matrix ===

  a    b    c  <-- classified as
1079  58   47 |    a = G1
 38 3608   2 |    b = G3
 36  28 2246 |    c = G2

```

Fonte: Extraído sistema R.

A etapa de avaliação ou pós-processamento foi realizada consecutivamente com a etapa 7 - de visualização dos resultados-, e para exemplificar o critério de geração dos segmentos demonstra-se a seguir o formato de árvore de decisão gerada para um dos classificadores na base pessoa física.

QUADRO 9 – Exemplo árvore de decisão gerada.

```

IDADE <= 46
|  CONVENIOAGUA <= 0
|  |  CONVENIOLUZ <= 0
|  |  |  LI1ITECCUSADO <= 1800
|  |  |  |  IDADE <= 42
|  |  |  |  |  LI1ITECC <= 3700: G1 (6998.0/9.0)
|  |  |  |  |  LI1ITECC > 3700
|  |  |  |  |  |  IDADE <= 39
|  |  |  |  |  |  |  LI1ITECC <= 13000
|  |  |  |  |  |  |  |  CONSAUTO <= 0: G1 (189.0/3.0)
|  |  |  |  |  |  |  |  CONSAUTO > 0

```

Fonte: Extraído sistema R.

Observa-se pelo exemplo acima que os 6.998 associados pessoa física, que representam 82,49% e foram classificados como G1, tem a seguinte característica: 42 anos de idade ou menos, não possuem convênio de débito automático de água e luz, com limite de cheque especial em conta corrente de 3.700,00 ou menor.

Como no exemplo acima foi realizado este procedimento em cada classificador, identificando a maior representatividade de cada *cluster*, tanto para a base de pessoa física como de pessoa jurídica. Os resultados obtidos estarão descritos no capítulo de discussão dos resultados deste mesmo estudo.

G) Visualização dos Resultados: na etapa de visualização dos resultados são utilizadas técnicas para representar o conhecimento adquirido e apresentar de forma compreensível o conhecimento ao interessado.

A técnica aqui empregada foi a árvore de decisão gerada por meio do algoritmo C 4.5 recodificado para linguagem Java por meio do algoritmo j 48 implementado no software WEKA como J48 (WITTEN; FRANK, 2005, p. 373).

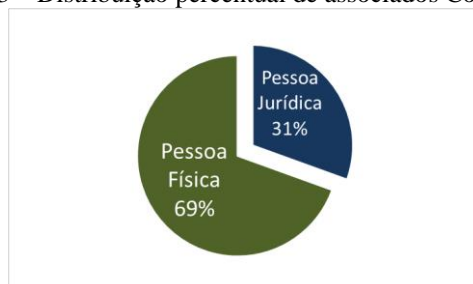
4 DISCUSÃO DOS RESULTADOS

Este capítulo traz os principais resultados encontrados após a execução da metodologia descrita anteriormente. Primeiramente apresenta-se os resultados da estatística descritiva, logo em seguida as análises de correlação e, por fim, os resultados oriundos da mineração de dados. Em todas as análises, a base de dados foi dividida em pessoa física e pessoa jurídica, que serão representados pelas siglas PF e PJ, respectivamente.

4.1 ESTATÍSTICA DESCRITIVA

A cooperativa Alfa, na base disponibilizada, possui 18.088 associados cadastrados como pessoa física, e 7.944 associados cadastrados como pessoa jurídica, totalizando 26.032 associados. Percentualmente a distribuição entre pessoa física e pessoa jurídica é demonstrada graficamente a seguir:

GRÁFICO 03 – Distribuição percentual de associados Cooperativa Alfa



Fonte: elaborado pelo autor, 2016

Em relação à média de renda para a PF, entre os associados desta classe, ficou em R\$ 5.279,25. Vale ressaltar que 85,8% dos associados estão abaixo da média apresentada. Já na PJ, o faturamento médio ficou em R\$ 462.628,66, porém 91,3% das empresas estão abaixo desta média.

As contas correntes dos associados PF totalizam 13.949, com 63,5% destas atribuído limite de cheque especial, ou seja, 5.090 associados em potencial para este produto. Na PJ a relação de associados com conta corrente e sem limite cai para 38,8%, sendo 6.967 contas para 2.706 limites.

Dos associados PF que possuem limite de cheque especial, a utilização média ficou em 41,9%, enquanto na PJ 43,2% dos limites atribuídos são utilizados. Não é

objetivo deste trabalho a análise financeira, contudo a implantação de limites na metade do número de associados sem esse limite em conta corrente, considerando a utilização média e a taxa de juros padrão divulgada pela cooperativa Alfa nestas modalidades, existe uma possível receita desperdiçada na ordem de R\$ 8,8 milhões.

Percebe-se que o total de contas correntes não corresponde exatamente com o número de associados, isso porque não necessariamente um associado precisa ter conta corrente, basta integralizar o capital social para se tornar um cooperado. A medida estatística moda identifica que na PF o capital com maior frequência é de R\$ 50,00, que é o mínimo definido pelo estatuto social da cooperativa, já a média aponta R\$ 1.798,79 de capital por associado, o que indica um ponto de atenção, para investigações futuras, com relação à concentração de capital social em poucos cooperados.

Na PJ o conceito acaba se repetindo, porém com valores diferenciados. A moda é de R\$ 500,00, mesmo o estatuto social indicando R\$ 50,00. Isso por conta de cultura interna da cooperativa em relação ao capital social de cooperados PJ. A média de capital na PJ ficou em R\$ 3.921,38, indicando os mesmos cuidados da PF.

Os associados PJ possuem um saldo maior de investimentos do que a PF, sendo maior em R\$ 46,7 milhões. Com relação a produtos de investimento, a cooperativa Alfa possui poucas opções, restringindo-se ao RDC (Recibo de Depósito Cooperativo), e ainda num estágio embrionário a LCA (Letra de Crédito do Agronegócio).

Com relação ao saldo médio em conta corrente, a PJ também leva vantagem em relação a PF, já que este saldo médio por cooperado na PJ é 6,8 vezes maior do que na PF. Isto se dá por conta da própria atividade empresarial.

A proporção nas operações de crédito não é muito diferente, sendo 6,4 vezes maior a média de operações por associado PJ em relação a PF. Já em relação ao número de associados tomadores de crédito a PF fica com 11.330 com 62,6% do total de associados PF, enquanto que na PJ 54,4% dos associados são tomadores, sendo 4.323 associados. O total concedido em operações de crédito na base apresentada é de R\$ 406 milhões, que a cooperativa Alfa subdivide em algumas modalidades de crédito para seus associados às quais seguem:

A) **Limite Rotativo e Limite Rotativo Utilizado:** Trata-se de um limite concedido ao associado, que o utiliza conforme a sua necessidade, difere do limite em conta corrente pelo valor, garantias, e formato jurídico;

B) **Limite TD e Limite TD Utilizado:** Trata-se de limite concedido ao associado para realizar descontos de recebíveis como, duplicatas, cheques pré-datados, promissórias;

C) **Limite C/C e Limite C/C Utilizado:** Trata-se de limite atribuído na conta corrente do associado para cobrir eventuais cheques a descobertos, ou necessidade eventual de saldo em conta corrente;

D) **Financiamento:** Trata-se de financiamento de um bem móvel ou imóvel;

E) **Consignado:** Trata-se de empréstimo realizado com desconto de parcelas diretamente na folha de pagamento do trabalhador;

F) **Pré-Aprovado:** Trata-se de limite concedido diretamente na conta corrente para empréstimos parcelados. Difere do limite em conta corrente por ser parcelado.

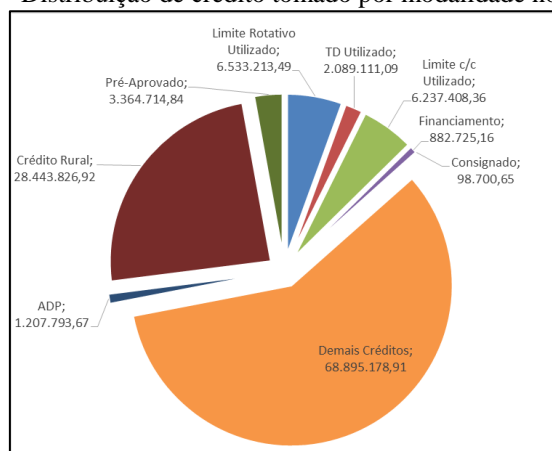
G) **Crédito Rural:** Trata-se de operação de crédito com fim de financiar a agricultura como um todo, implementos, compra de grãos, compra de bovinos, investimento na propriedade rural;

H) **ADP (Adiantamento a Depositante):** Trata-se de valor concedido a título de excepcionalidade, quando o associado ultrapassa o limite de conta corrente em suas operações, especialmente cobertura de cheques emitidos pelo associado;

I) **Demais Créditos:** Quando não enquadra em nenhuma modalidade acima citada.

A seguir apresenta-se a distribuição destas modalidades de crédito nas carteiras de associados PF e PJ.

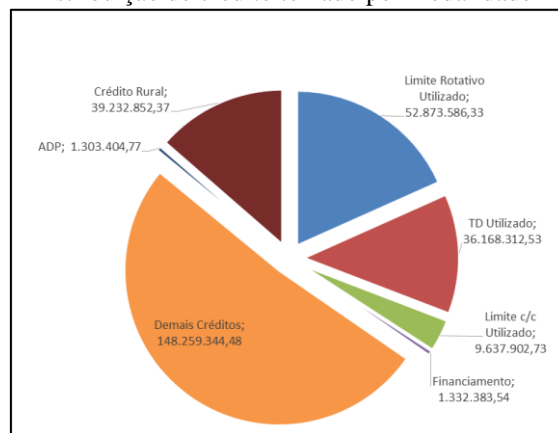
GRÁFICO 04 – Distribuição de crédito tomado por modalidade nos associados PF



Fonte: elaborado pelo autor, 2016

Percebe-se que a modalidade mais utilizada pelos associados PF é a de crédito rural, seguida pelo rotativo e utilização do cheque especial. Um problema identificado é que 58,5% da concessão de crédito para PF está cadastrado como demais créditos, o que inviabiliza uma análise mais apurada da carteira de crédito.

GRÁFICO 05 – Distribuição de crédito tomado por modalidade nos associados PJ



Fonte: elaborado pelo autor, 2016

Os associados PJ contratam mais operações de créditos na modalidade de rotativo, seguido pelo crédito rural e TD (Desconto de Recebíveis). O mesmo problema identificado na carteira PF está na PJ, com 51,3% da carteira classificada como demais créditos, inviabilizando maiores análises.

Até aqui a análise concentrou-se o que tange à intermediação financeira, ou seja, fontes de captação de recursos e sua aplicação em operações de crédito. Os próximos parágrafos tratam dos produtos e serviços que a cooperativa oferta para seus associados. O primeiro produto será o convênio de débito automático.

Os convênios de débitos automáticos de faturas concentram-se em quatro produtos: água, luz (energia elétrica), telefone e cartão de crédito. Na PJ o convênio com maior cadastro é o de luz com 1.496 registros, seguido pelo telefone com 817, água com 799 e cartão com 464. Com exceção do convênio de débito automático da fatura do cartão de crédito, para os demais pode-se considerar uma base de 7.944 associados PF para efeito de comparação.

No caso da PF, o convênio segue o mesmo critério, sendo água, luz (energia elétrica), telefone e cartão de crédito. Os números ficam abaixo da PJ, cartão em primeiro com 1.753, telefone com 667 é o segundo, luz com 657 seguido pela água com 367 cadastros. A base na PF é maior com 18.808 para comparação. A seguir apresenta-se um

quadro do percentual de penetração dos convênios, com exceção da fatura do cartão, que será tratada na análise do cartão de crédito:

QUADRO 10 – Percentual de penetração do convênio débito automático

	Nº Associados	Convênio Água	Convênio Luz	Convênio Telefone
PJ	7.944	10,06%	18,83%	10,28%
PF	18.808	1,95%	3,49%	3,55%
Total	26.752	4,36%	8,05%	5,55%

Fonte: elaborado pelo autor, 2016

O produto cartão de crédito em números absolutos possui 8.447 para PF e 1.341 na PJ. Na discussão sobre os convênios, conforme visto acima, este cartão possui fatura e pode ser cadastrado como débito automático. Em termos percentuais, o débito automático é cadastrado em 34,6% dos cartões PJ e 19,8% na PF, fechando em 21,7% do total de cartões com débito automático de fatura.

O cartão de crédito possui um limite para compras definido por uma análise interna da cooperativa Alfa. Embora tenha a conotação de operação de crédito é tratado como produto, e não linha de crédito na cooperativa.

Existe também o cartão de débito, que permite ao associado movimentar a conta corrente realizando compras sem a necessidade de dinheiro em espécie, bem como, caso esta necessidade exista, o cartão pode realizar saque nos terminais de autoatendimento. O número de cartões de débito é maior do que os de crédito, sendo respectivamente 16.038 e 9.788.

A seguridade também faz parte do *hall* de produtos da cooperativa Alfa, sendo seguro de vida, seguro auto (automóvel), seguro residencial e diversos. As empresas (associados PJ) que contrataram seguro de vida totalizam 555, enquanto que na PF este número sobe para 2.686. O seguro de auto também possui maiores números na PF com 548 apólices contra 83 na PJ. Este contexto se repete no seguro residencial com 665 na PF e 6 na PJ. Os seguros diversos possuem 50 associados na PJ e 8 na PF, contrariando os demais seguros.

Além dos seguros, a cooperativa Alfa oferece um meio de compra de bens de forma parcelada diferente das operações de financiamentos, esta opção é o consórcio. Existem três principais modalidades que são: consórcio de auto (automóveis), consórcio de moto (motocicletas), e consórcio de imóveis. Na carteira de associados PF o percentual de aquisição de consórcio de auto é de 2,78% com 503 associados. Este percentual sobe

para 4,56% na carteira PJ com 362 associados. No caso do consórcio de moto, o total de associados, que fizeram aquisição desta modalidade fica em 216, sendo 75% destes na PF. O consórcio de imóvel, embora pertencente à gama de produtos da cooperativa, não foi divulgado.

O último produto da base disponibilizada é a cobrança registrada, que é um produto mais direcionado a PJ, mesmo assim alguns casos de autônomos na base de dados acabam incluindo este produto na PF. Este item trata do convênio para emissão de boletos enviados para os clientes dos associados. Do total de associados com este produto 95,7% está na PJ, ou dos 1.127 associados 1.079 está na PJ, enquanto que na PF são 48 casos.

Após esta análise descritiva da base de dados dos associados da cooperativa Alfa na sequência serão apresentados os resultados da análise de correlação entre as variáveis desta base.

4.2 ANÁLISE DE CORRELAÇÃO

Em consonância com a metodologia apresentada nesta pesquisa, a discussão nesta seção é o resultado das correlações realizadas entre as variáveis que compõem a base de dados dos associados da cooperativa Alfa. Vale frisar que esta análise não é apresentada como fator de causa/efeito, ou seja, as mudanças em uma variável não necessariamente explicam mudanças em outra somente por esta medida estatística. Também nesta análise ocorreu a distinção entre pessoa física e pessoa jurídica. O quadro com as correlações está descrito no apêndice 3 e 4 deste trabalho. A seguir são apresentados os pontos a se destacar desta análise:

A) Em relação à renda, tanto na carteira PF quanto na PJ, existe correlação com o total de crédito concedido, porém esta correlação é maior na PJ;

B) No que se refere à captação de recursos em depósitos a prazo, depósitos à vista e capital social, a renda possui baixa correlação. Isso ocorre tanto na PF quanto na PJ. O teste (p) para estas correlações demonstrou que mesmo com baixa correlação esta é significativa;

C) A modalidade de crédito com maior correlação a renda é o crédito rural, no caso das carteiras PF e PJ. A utilização do rotativo também é correlata à renda, porém é menos do que o rural;

- D) O capital social integralizado está relacionado ao total de crédito utilizado, e de uma forma mais elevada nos cooperados PJ, que também relacionam capital com investimentos;
- E) Os associados que possuem saldo médio em conta, tem uma tendência a serem investidores;
- F) As modalidades de crédito rotativo e rural possuem correlação positiva ao investimento;
- G) Os investidores possuem uma tendência a colocarem a fatura do telefone em débito automático no caso da PF;
- H) As modalidades de crédito rotativo, TD (Títulos Descontados), rural e limite em conta corrente são as variáveis que possuem o maior número de correlações nas movimentações dos associados (PJ e PF);
- I) A utilização dos limites em conta corrente, do TD, e do Rotativo, possui relação com a utilização do ADP, no caso da PF. Esta relação acontece na PJ, contudo com uma menor correlação;
- J) Existe relação entre o ADP e o crédito rural na base PF;
- K) A relação positiva existente entre conta corrente e quantidade de contas com limite na PF é maior do que na PJ;
- L) A utilização do limite em conta corrente tem relação linear ao crédito pré-aprovado para cooperados PF.
- M) Normalmente os associados que possuem débito de água, possuem débito de luz e telefone, e este movimento ocorre na PF e PJ;
- N) Existe uma correlação entre a aquisição de seguro de vida e de seguro de automóvel com o seguro residencial;
- O) O cartão de débito possui correlação positiva em relação ao de crédito na PJ, mas na PF esta relação é maior;
- P) O débito automático da fatura do cartão possui correlação com o débito de luz na PF, enquanto na PJ esta correlação se dá com o débito automático de telefone.

4.3 MINERAÇÃO DE DADOS

Depois de passar pela estatística descritiva e da análise de correlação das variáveis da base de dados da cooperativa Alfa, nesta seção apresentam-se os resultados obtidos

com a mineração de dados seguindo os passos já descritos no capítulo da metodologia da pesquisa.

Especificamente será apresentado a seguir o resultado dos passos 4 (transformação) e 5 (mineração), de forma que se atende os passos 6 (avaliação) e 7 (resultados).

Na sequência, apresenta-se por meio dos quadros 11 e 12, as principais folhas de cada árvore de decisão, geradas por pessoa física, e pessoa jurídica, nos 8 diferentes tempos destacados no capítulo 3.

QUADRO 11 – Resultados das Classes Pessoa Física (PF)

Tempo	Classes		
	G1	G2	G3
1	Idade de 42 anos ou menos, sem convênio de débito de água e luz, com limite em conta corrente de 3.700,00 ou menor.	Idade maior que 48 anos, sem convênio de água, luz e telefone, utilizando 35.000,00 ou menos de rotativo.	Idade menor que 46 anos, com convênio de água e luz.
2	Idade menor que 45 anos, sem convênio de água e de luz com limite de 1.900,00 ou menor.	Idade menor que 47 anos, com convênio de água e de luz.	Idade de 48 anos ou mais, sem convênio de água, luz e telefone, utilizando 27.000,00 ou menos de rotativo.
3	Idade de 46 anos ou menos, com convênio de água e de luz.	Idade maior que 48 anos, sem convênio de água, luz e telefone, utilizando 32.000,00 ou menos de rotativo.	Idade menor que 43 anos, sem convênio de água, luz e telefone, limite em conta de 4.300,00 ou menos, limite pré-aprovado de 8.212,18 ou menos.
4	Idade maior que 48 anos, sem convênio de água, luz e telefone, utilizando	Idade menor que 43 anos, sem convênio de água e de luz, com	Idade de 46 anos ou menos, com convênio de água e de luz.

	10.000,00 ou menos de rotativo.	limite em conta de 4.300,00 ou menos e limite pré-aprovado de 7.984,68 ou menos.	
5	Idade de 43 anos ou menos, sem convênio de água, luz e telefone, utiliza 1.999,77 ou menos do limite em conta, e possui limite de 4.500,00 ou menos.	Idade de 47 anos ou mais, possui convênio de água e de luz.	Idade de 48 anos ou mais, sem convênio de água, luz e telefone, possui limite de rotativo de 40.000,00 ou menos e crédito rural de 151.192,91 ou menos.
6	Possui limite em conta corrente, convênio de água e de luz.	Possui limite em conta corrente, sem convênio de água e de luz, com possibilidade de possuir convênio de telefone, utiliza 30.000,00 ou menos de limite de rotativo.	Não possui limite em conta corrente, sem convênio de água, luz e telefone, com uma conta corrente, limite pré-aprovado de 4.017,90 ou menos, um ADP de 449,04 ou menos e uma rentabilidade de 1.015,99 ou menos.
7	Não possui convênio de água, luz e pode ter o de telefone, com limite em conta corrente de 25.000,00 ou menos, limite de rotativo de 35.000,00 ou menos, e capital social de 41.868,45 ou menos.	Não possui convênio de água, luz ou telefone.	Possui convênio de água e de luz.

8	Não possui convênio de água, luz e de telefone, não possui consórcio de auto e moto, possui uma conta corrente, ADP (Adiantamento a depositante) de 1.870,50 ou menos, crédito rural de 81.004,19 ou menos, capital social de 54.152,46 ou menos, investimento de 162.071,60 ou menos e pré-aprovado de 4.675,89 ou menos.	Não possui convênio de água, luz ou telefone e utiliza o limite em conta acima de 7.429,59.	Possui convênio de luz e utiliza o limite em conta corrente em 3.547,89.
---	--	---	--

Fonte: Algoritmo C 4.5

QUADRO 12 – Resultados das Classes Pessoa Jurídica (PJ)

Tempo	Classes		
	G1	G2	G3
1	Conta corrente sem limite, com idade de 25 anos ou mais.	Limite em conta corrente, sem convênio de água, pode ter de telefone, utiliza o rotativo em 189.997,08 ou menos, com idade de 20 anos ou menos.	Sem limite em conta corrente, menos de 22 anos de idade, sem convênio de água e telefone, com limite de rotativo de 201.260,00 ou menos, com 1 conta corrente, e saldo médio de 44.054,06 ou menos, e demais créditos em 65.813,31.

2	Idade de 21 anos ou menos, com convênio de água, luz e telefone, com limite em conta corrente de 25.000,00 ou menos.	Sem convênio de água e de telefone, com 23 anos de idade ou mais.	Idade menor que 16 anos, sem convênio de água, sem telefone, com limite em conta corrente utilizado de 5.063,13 ou menos, com limite de rotativo de 83.000,00 ou menos, com rentabilidade de 1.385,48 ou menos, com limite de TD de 130.000,00 ou menos, limite em conta corrente de 12.000,00 ou menos e investimentos de 583.657,43 ou menos.
3	Idade de 21 anos ou mais, com convênio de água e telefone.	Idade de 23 anos ou mais, sem convênio de água e telefone.	Idade menor que 14 anos, sem convênio de água, com limite em conta corrente utilizado de 5.089,83 ou menos, com limite de rotativo de 160.000,00 ou menos, com limite de TD de 220.000,00 ou menos, com uma conta corrente, com rentabilidade de 1.394,72 ou menos, sem consórcio de auto.

4	Idade menor que 16 anos, sem convênio de água, com limite rotativo utilizado de 94.656,23 ou menos, com limite em conta corrente utilizado de 9.681,36 ou menos, uma conta corrente, limite de TD de 240.000,00 ou menos e capital social de 28.877,87 ou menos.	Idade de 18 anos ou menos, sem convênio de água, e limite rotativo utilizado de 511.134,94 ou mais.	Idade de 22 anos ou mais, com rentabilidade de 7.976,23 ou menos.
5	Idade de 17 anos ou menos, com convênio, com limite em conta corrente.	Idade de 20 anos ou mais, convênio, com convênio de luz, com limite de rotativo de 330.000,00 ou menos, rentabilidade de 2.195,14 ou menos.	Idade menor que 17 anos, sem convênio, sem convênio de água, limite de TD de 150.000,00 ou menos, limite de conta corrente de 25.000,00 ou menos, uma conta corrente, rentabilidade de 3.626,97 ou menos, limite de rotativo de 83.000,00 ou menos, sem consorcio de auto.
6	Idade de 23 anos ou mais, com conta corrente e sem limite em conta corrente.	Idade de 21 anos ou mais, conta corrente sem limite, sem convênio, sem convênio de água, com limite de rotativo 160.000,00 ou menos, com uma conta corrente,	Idade menor que 23 anos, conta corrente com limite, sem convênio de água e luz, com capital de 35.069,37 ou menos.

		com rentabilidade de 1.717,01 ou menos, sem consórcio de auto.	
7	Idade maior que 2 anos, sem convênio, com limite de TD de 110.000,00 ou menos.	Conta corrente sem limite, saldo médio de 27.816,96 ou menos, sem limite de rotativo, sem convênio de telefone, com capital de 4.047,73 ou menos.	Conta corrente com limite.
8	Conta corrente com limite, rentabilidade de 10.433,56 ou menos, ADP de 10.032,73 ou menos.	Conta corrente sem limite, sem convênio, pode ter convênio de água, limite rotativo utilizado maior de 500,00, com limite de rotativo maior de 800.000,00.	Conta Corrente sem limite, sem convênio, pode ter convênio de água, luz e telefone, limite de rotativo usado de 500,00 ou menos, demais créditos de 350.662,85 ou menos, saldo médio de 121.340,40 ou menos, investimentos de 825.357,37 ou menos.

Fonte: Algoritmo C 4.5

Basicamente os quadros acima são o resumo das características de cada segmento identificado por meio da árvore de decisão gerada, ou seja, nesta etapa demonstra-se exclusivamente a sequência dos ramos, sem qualquer manipulação. Para facilitar o entendimento, demonstra-se a seguir a proposta de segmentação para a cooperativa Alfa obtida após a análise das características evidenciadas acima.

QUADRO 13 – Proposta de segmentação para cooperativa Alfa.

Segmentos	PF	PJ
	Critérios	
S₁	Mais de 48 anos ou mais de 190 mil em crédito.	Mais de 25 anos ou mais de 330 mil em crédito.
S₂	Entre 45 e 47 anos ou entre 35 mil e 189.999,99 em crédito.	Entre 20 e 24 anos ou entre 189 mil e 329.999,99 em crédito.
S₃	Entre 43 e 44 anos ou entre 1.900 e 34.999,99 em crédito.	Entre 16 e 19 anos ou entre 83 mil e 188.999,99 em crédito.
S₄	Menos de 43 anos ou menos de 1.899,99 em crédito.	Menos de 16 anos ou menos de 82.999,99 em crédito.

Fonte: elaborado pelo autor, 2016

Percebe-se que as características mais evidentes para a segmentação foram a faixa etária, no caso da pessoa física a idade, e no caso da pessoa jurídica o tempo de abertura da empresa e o valor total tomado em operações de crédito.

Outra característica que ficou evidente após análise dos quadros 11 e 12 foi o convênio de débito automático, que no primeiro quadro não foi mencionado e para isso propõe-se a seguir uma forma de incluir esta característica na metodologia proposta de segmentação.

QUADRO 14 – Proposta de segmentação para cooperativa Alfa com convênio.

Segmentos	PF	PJ
	Critérios	
S₁₊	Mais de 48 anos ou mais de 190mil em crédito + débito automático (água, luz ou fone).	Mais de 25 anos ou mais de 330mil em crédito + débito automático (água, luz ou fone).
S₂₊	Entre 45 e 47 anos ou entre 35 e 189.999,99 em crédito + débito automático (água, luz ou fone).	Entre 20 e 24 anos ou entre 189 mil e 329.999,99 em crédito + débito automático (água, luz ou fone).
S₃₊	Entre 43 e 44 anos ou entre 1.900 e 34.999,99 em crédito + débito automático (água, luz ou fone).	Entre 16 e 19 anos ou entre 83 mil e 188.999,99 em crédito + débito automático (água, luz ou fone).
S₄₊	Menos de 43 anos ou menos de 1.899,99 em crédito + débito automático (água, luz ou fone).	Menos de 16 anos ou menos de 82.999,99 em crédito + débito automático (água, luz ou fone).

Fonte: elaborado pelo autor, 2016

Objetivamente se propõe que a cooperativa obtenha segmentos de associados baseados nas características identificadas e demonstradas no quadro 13 por meio de S₁, S₂, S₃ e S₄. Em seguida uma segmentação *Plus*, apresentada no quadro 14, que seria basicamente a adição do produto convênio de débito automático obtendo-se assim os segmentos S₁₊, S₂₊, S₃₊ e S₄₊. O fato de o associado possuir um débito automático o colocaria num segmento diferenciado.

Além das características utilizadas para a formação de segmentos, a árvore de decisão pode apresentar possíveis padrões de comportamentos de um número restrito de cooperados contribuindo assim para a assertividade de oferta e posicionamento estratégico da cooperativa nos segmentos identificados.

A seguir apresenta-se exemplos de comportamento padrão observado e extraídos da árvore de decisão gerada:

Quadro 15 – Exemplo Comportamento Observado I

									LI1ITECCUSADO <= 3200
									DAP <= 107023.5: G1 (144.0/1.0)
									DAP > 107023.5
									IDADE <= 39: G1 (3.0)
									IDADE > 39: G3 (2.0)
									LI1ITECCUSADO > 3200
									PREAPROVADO <= 3316.69
									ADP <= 1462.87: G1 (8.0)
									ADP > 1462.87: G3 (2.0)
									PREAPROVADO > 3316.69: G3 (4.0)

Fonte: Extraída árvore de decisão algoritmo C 4.5. Elaborado pelo autor 2016.

Como pode-se observar no quadro 15 acima que foi extraído da árvore de decisão gerada para a pessoa física nesta pesquisa, 144 associados que utilizam o limite em conta corrente no valor de R\$ 3.200,00 ou menos possuem investimentos menores ou iguais a R\$ 107.023,50, já os que possuem investimos acima deste valor 3 têm idade igual ou inferior a 39 anos e 2, superior a 39 anos.

Para associados com utilização acima de R\$ 3.200,00 do limite em conta corrente a característica identificada é a utilização do empréstimo pré-aprovado no valor inferior ou igual a R\$ 3.316,69 que são 10 associados. Estes ainda se dividem em utilização de adiantamento a depositante onde 8 utilizam valores iguais ou menores que R\$ 1.462,87 e 2 valores maiores de R\$ 1.462,87. Na utilização de empréstimo pré-aprovado com valores acima de 3.316,69 são 4 associados.

Um outro exemplo de comportamento observado é demonstrado a seguir:

Quadro 16 - Exemplo Comportamento Observado II

									LI1ITEROTATIVO <= 70000
									CONSAUTO <= 0
									LI1ITECC <= 10000: G2 (37.0/1.0)
									LI1ITECC > 10000: G3 (4.0/1.0)
									CONSAUTO > 0: G3 (5.0/1.0)
									LI1ITEROTATIVO > 70000: G3 (4.0)

Fonte: Extraída árvore de decisão algoritmo C 4.5. Elaborado pelo autor 2016.

Neste outro trecho a característica que se destaca é a de que os associados que possuem limite de empréstimo rotativo menores ou iguais a R\$ 70.000,00, e não utilizam limite em conta corrente, possuem consorcio de auto, sendo 5 associados demonstrando este comportamento.

Conhecendo os detalhes de comportamento apresentados no quadro 15 a cooperativa pode por exemplo inibir a utilização de adiantamento a depositante na carteira

de associados que utilizam o empréstimo pré-aprovado nos valores apresentados. Pode também oferecer a novos associados que utilizem limite em conta corrente abaixo de R\$ 3.200,00 as modalidades de investimento da cooperativa.

Para o segundo exemplo de comportamento apresentado no quadro 16 a oferta do produto consórcio pode ser assertiva para associados que utilizam ou passem a utilizar o limite de rotativo abaixo de R\$ 70.000,00 e não utilizem o limite de conta corrente.

5 CONCLUSÃO, LIMITAÇÕES DA PESQUISA E SUGESTÕES DE PESQUISAS FUTURAS

No último capítulo da pesquisa será apresentada a conclusão, bem como algumas implicações gerenciais. Apresenta-se também limitações identificadas no estudo e ainda futuras pesquisas serão recomendadas. Toda esta seção foi embasada a partir dos resultados da aplicação da metodologia apresentada e dos respectivos resultados discutidos no capítulo anterior.

5.1 CONCLUSÃO

O estudo da estratégia de segmentação, para o cooperativismo de crédito, ganha relevância com aumento de pessoas e empresas que se associam a esta modalidade de instituição financeira, oferecendo os mesmos produtos e serviços dos bancos convencionais, que já possuem uma estrutura de segmentação desenvolvida e implantada.

Com a igualdade de oferta dos produtos e serviços financeiros, uma forma de buscar um diferencial competitivo junto aos associados do cooperativismo de crédito é a excelência no atendimento. Assim, conhecer os cooperados e alocá-los em grupos homogêneos em relação às suas características de compra e utilização da cooperativa, configura-se numa tarefa essencial.

Inserida neste contexto, a cooperativa Alfa, que foi utilizada como base deste estudo, é uma cooperativa de crédito do sistema Sicoob, com mais de 19 mil associados. A cooperativa possui milhares de registros das movimentações dos seus associados, o que dificultou a tarefa de segmentar a base de cooperados, sendo possível somente com o auxílio da tecnologia.

Antes de efetivamente concluir as observações realizadas neste estudo vale ressaltar alguns pontos importantes que respondem aos objetivos propostos. O primeiro deles é a característica da base de associados da cooperativa Alfa. Pela pesquisa realizada verifica-se que embora a cooperativa possua 69% da base de associados como pessoa física, em relação aos negócios, estes de modo geral são realizados em maior volume financeiro com os cooperados pessoa jurídica.

Verificou-se também que os negócios da cooperativa se concentram na operação base de toda instituição financeira, que é captar recursos no mercado através dos

investimentos de seus clientes, no caso da cooperativa associados, e em seguida emprestá-los por meio das operações de crédito concedidas, ou seja, a intermediação financeira.

A venda de produtos e serviços como seguros, consórcios e convênios ainda é incipiente registrando uma grande lacuna em relação ao número de associados com estes produtos adquiridos.

Com a mineração de dados, pode-se também verificar alguns padrões comportamentais na base de associados. Nesse momento da pesquisa, foi possível captar a diversidade da população de associados que compõe a base de dados e assim identificar as características descritivas apresentadas nos quadros do capítulo de análise dos resultados, permitindo à cooperativa Alfa identificar grupos de associados distintos e, assim, obter assertividade na oferta de produtos.

Utilizando estes mesmos comportamentos discutidos pode-se tratar do terceiro ponto importante que é a verificação, entre esses padrões e comportamento, da possibilidade de oferta de produtos e serviços mais direcionados às necessidades dos cooperados.

Todas as características identificadas atendem o objetivo proposto e podem conduzir a cooperativa criar alertas nos sistemas de tecnologia e assim contribuir para assertividade na ação da equipe comercial e ainda permitir a gestão estratégica de forma eficiente.

Diante das considerações em relação aos objetivos específicos pode-se resumir estes com o objetivo geral do estudo que é: identificar se a base de associados da cooperativa “Alfa” evidencia oportunidades de oferta pautada na estratégia de segmentação de clientes.

De acordo com o referencial teórico deste estudo, a segmentação é o processo de colocar em subgrupos os compradores de um produto-mercado de modo que os integrantes de cada segmento exibam uma responsividade semelhante em relação à determinada estratégia de posicionamento. Assim, os segmentos S_1 , S_2 , S_3 e S_4 , tanto na pessoa física, quanto na pessoa jurídica, extraídos da árvore de decisão criada por meio do algoritmo C 4.5, implementado no WEKA, configuram-se uma proposta de estrutura de segmentação para a cooperativa Alfa.

Esta proposta de segmentação permitirá à cooperativa identificar segmentos alvo bem como definir seu posicionamento estratégico em cada segmento, na tentativa de ter uma posição clara ou única no mercado.

Um exemplo claro desta conclusão seria o segmento S₁ Pessoa Física que possui 2.146 associados pelo critério de idade e 8.936 associados pelo critério de concessão de crédito. Após esta identificação do número de associados e as características percebe-se que são pessoas mais jovens com pequenas operações de crédito.

Surgem então as questões que a cooperativa deverá responder: Como vou atender estes associados? Será por meio eletrônico, ou vou designar uma equipe para este segmento? Quantos colaboradores são necessários para atender este segmento? Quais produtos devo ofertar a este segmento? Será meu segmento alvo?

Estas mesmas perguntas deverão ser respondidas para cada um dos segmentos identificados.

A metodologia de criação destes segmentos não é complicada de ser utilizada e possui flexibilidade de uso para atualização de futuros dados complementares. Ou seja, é dinâmica na geração dos segmentos, tanto com relação aos critérios, quanto a questão temporal.

Por fim, os resultados alcançados por esta pesquisa permitem concluir que a cooperativa Alfa possui oportunidades de oferta baseadas na estratégia de segmentação, atendendo positivamente o problema deste estudo, que, apesar de respondido, possui algumas limitações tratadas na seção seguinte.

5.2 LIMITAÇÕES DO ESTUDO

A abrangência do estudo possui limitações, como as elencadas a seguir:

A) A base de dados dos associados e suas respectivas movimentações possui deficiência em relação à totalidade dos produtos ofertados pela cooperativa Alfa. Ela não indica 100% dos produtos adquiridos, limitando os achados do estudo;

B) Os itens selecionados sobre produtos e serviços apresentam somente a informação sobre a aquisição ou não do produto, ou seja, somente se cooperado possuía ou não o produto e serviço, e não apresentam valores, por exemplo, das cartas de consórcio ou dos seguros contratados;

C) Por ser uma singular e utilizar um sistema de *input* de dados fornecidos por uma confederação, os dados gerados, em relação aos associados, tem uma defasagem de atualização, o que tornou a pesquisa ainda mais difícil;

D) O tamanho da base de dados exige tratamento para que a qualidade da mineração em relação à capacidade de processamento em determinados computadores seja atingida;

E) A utilização de base de dados é interna da cooperativa, não avaliando assim o comportamento do associado com o mercado financeiro como um todo;

F) A metodologia restrita ao uso interno à cooperativa, pois a exposição de segmentação aos associados poderia incorrer em possíveis equívocos de interpretações, tendo em vista, por exemplo, um dos valores do cooperativismo, que é o tratamento igualitário.

5.3 SUGESTÕES PARA FUTURAS PESQUISAS

A necessidade de ganhos de desempenho no atendimento aos associados, que além de associados, também são usuários de outras instituições financeiras participantes de um mercado extremamente competitivo, pode trazer a utilização da metodologia apresentada neste estudo, em consonância com a prática diária da cooperativa Alfa.

Neste sentido, este trabalho pode ser um passo inicial para que as cooperativas consigam prever futuras necessidades de determinados segmentos de associados, ou mesmo de cada associado. Uma pesquisa focada na criação de modelos de previsão de demanda pode ser desenvolvida em adição a este estudo. A inovação da utilização das técnicas tratadas ao longo desta pesquisa pode contribuir para além dos ganhos em atendimento. Pode ser expansiva à estrutura de custos e processos da cooperativa, já que, com modelos de predição de demanda, seria maior a assertividade na alocação de recursos. Aqui uma pesquisa votada para a gestão de custo em cada segmento ou ainda com foco em processos poderia ser desenvolvida.

Toda a estratégia de vendas poderá ser mais focada e dirigida aos associados de forma mais personalizada, pois irá permitir entender melhor cada empresa, cada indivíduo. A movimentação dos cooperados será mais bem entendida por parte dos gerentes de negócios, permitindo-os oferecer condições diferenciadas proporcionando melhores negócios aos correntistas, seja pessoa jurídica ou pessoa física. A implantação de metodologia de CRM (*Customer Relationship Manager*) com vistas aos segmentos identificados neste estudo pode configurar-se como uma pesquisa futura.

Entender melhor a movimentação dos associados pode conduzir a uma cobrança mais justa de tarifas de serviços, adequando cada pacote tarifário ao perfil de utilização

dos serviços oferecidos, das transações efetuadas em conta corrente. Neste sentido diminuir a dependência da cooperativa Alfa em relação a intermediação financeira e ajustar a oferta de produtos e serviços também justifica um estudo posterior.

É preciso ir além da movimentação dos associados com a cooperativa, incluindo no futuro base de dados externas, com as movimentações destes associados com o mercado financeiro como um todo, utilizando parceiros como o SERASA, o próprio Banco Central do Brasil. Ultrapassar esta barreira contribuiria para afinar a previsão de demandas e ainda para prever possíveis inadimplências, mitigando assim riscos financeiros para a cooperativa.

REFERÊNCIAS

- ALESSIO, Odir Mario Rubin; FORTES, Darcy Alvares de Siqueira. **Manual de Cooperativismo**. Porto Alegre-RS: Feplam, 1977.
- AZAD, Chandrashekhar. JHA, Vijay Kumar. **Genetic Algorithm to Solve the Problem of Small Disjunct In the Decision Tree Based Intrusion Detection System**. Computer Network and Information Security, Índia, 2015.
- BAKONYI, Emmerich. **Multi-Use Credit Co-operatives**. Universität Hohenheim, vol 7, Stuttgart, 1990.
- BANCO Central do Brasil. Disponível em: <<http://www.bcb.gov.br/?INDECO>>. Acesso em: 03 de agosto de 2016.
- BERRIGAN, John. FINKBEINER, Carl. **Marketing de Segmentação**. São Paulo: Makron Books, 1994.
- BIALOSKORSKI, Neto. **Economia e Gestão de Organizações Cooperativas**. 2. ed. São Paulo: Atlas, 2012.
- BRADESCO, Disponível em <<http://www.bradesco.com.br/abertura>> acesso em 03 de maio de 2016.
- BROCHADO, Ana Oliveira. MARTINS, Francisco Vitorino. Aspectos Metodológicos da Segmentação de Mercado: Base de Segmentação e Métodos de Classificação. **Revista Brasileira de Gestão de Negócios**, São Paulo, v. 10, n. 27, p. 132-149, abr/jun 2008.
- BUENO, Regiane. IKEDA, Ana Akemi. Segmentação de Consumidores de Produtos e Serviços Bancários: Um Estudo Exploratório. **Revista Brasileira de Marketing**, São Paulo, v. 12, n. 2, p. 133-157, abr./jun. 2013
- BULGARELLI, Waldirio. **Tratado Geral de Crédito Cooperativo**. Rio de Janeiro: Instituto Superior de Pesquisas e Estudos Sociais de cooperativismo, 1962.
- BRASIL. **Código Civil Brasileiro**. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/2002/L10406.htm> acesso em 15 de março de 2017.
- CHÉRON, Emmanuel J. MCTAVIS, Ronald. PERRIEN, Jean. *Segmentation of Bank Commercial Markets*. **International Journal of Bank Marketing**, Vol. 7 Iss 6 pp. 25 - 30
- CORRAR, Luiz J. PAULO, Edilson DIAS FILHO, José Maria. **Análise Multivariada**. São Paulo: Atlas, 2017.
- CRAVENS, David W. PIERCY, Nigel F. **Marketing Estratégico**. 8. ed. São Paulo: McGraw-Hill, 2007

- DAWAR, Niraj. *When Marketing is a Strategy*. In. **Harvard Business Review**. From the December, 2013 Issue.
- DIBB, Sally. *Market segmentation: strategies for success*. **Marketing Intelligence & Planning**, Vol. 16 Iss 7 pp. 394 406.
- DOUGLAS, Susan P. CRAIG, C. Samuel. **Global Marketing Strategy**. McGraw-Hill, 1995.
- FORCHT, Karen A. COCHRAN, Kevin. **Using Data Mining and Datawarehousing Techniques**. Industrial Management & Data Systems, Virginia, 1999.
- FUNDO GARANTIDOR DO COOPERATIVISMO. Disponível em: <<http://www.fgcoop.coop.br/home>> acesso em 06 maio de 2016.
- GARCIA ESCUDERO, Luis Ángel. GORDALIZA, Alfonso. *Robustness Properties of k Means and Trimmed k Means*. **Journal of the American Statistical Association**, p. 959, 1999.
- GAO, Wushi. TANG, Weifang. WANG, Xingang. . **Application of a improved C4.5 algorithm in Performance Analysis**. Trans Tech Publications, Switzerland, 2013.
- GAVETT, Gretchen. *What You Need to Know About Segmentation*. **Harvard Business Review**. From the July, 2014.
- GOPINATHAN.S, PANDIYAN.M e THANGAVEL.P. *Performance Evaluation of Extreme Learning Machine with k-Means Techniques for Clustering*. **International Journal of Advanced Research in Computer Science**, v. 8, nº5, 2017
- GOVÊA, Maria Aparecida. PREARO, Leandro Campi. ROMEIRO, Maria do Carmo. Avaliação do Emprego das Técnicas de Análise de Correspondência e Análise de Conglomerados em Teses e em Dissertações de Algumas Instituições de Ensino Superior. **Revista de Ciências da Administração**, Universidade Federal de Santa Catarina, 2013.
- HASSAN, Salah S. CRAFT, Stephen H. *Linking Global Market Segmentation Decisions with Strategic Positioning Options*. **The Journal of Consumer Marketing**, p. 81, 2005
- HUGON, Paul. **Histórias das Doutrinas Econômicas**. 11. ed. São Paulo: Atlas, 1995.
- ITAÚ. Relatório de sustentabilidade. Disponível em: <<http://www.itauunibanco.com.br/relatoriodesustentabilidade/2010/06>> acesso em 03 de maio de 2016.
- JIN, Ruoming. GOSWAMI, Anjan. AGRAWAL, Gagan. **Fast and exact out-of-core and distributed k-means clustering**. *Knowledge and Information Systems*, Springer-Verlag, London, 2005.
- LARSON, Martin G. *Descriptive Statistics and Graphical Displays*. In: **American Heart Association**, 114:76-81, Dallas, 2006.

- LEE, Jun Sang. SIAU, Keng. *A Review of Data Mining Techniques. Industrial Management & Data Systems*, Nebraska, 2001.
- LIRA, Sachiko Araki. **Análise de Correlação: Abordagem Teórica e de Construção dos Coeficientes com Aplicações.** Universidade Federal do Paraná, Curitiba, 2004.
- LIU, Huiming. LU, Jingrong. *Brief Survey of K-means Clustering Algorithms. Mechanics and Materials*, v. 740 (2015) pp 624-628.
- MANFRÉ, Patrícia Nascimento. **Concepção de um Sistema de Informação Integrado Inteligente para apoio ao Profissional Fisioterapeuta: Construído sobre uma Base de Dados Simulados.** Curitiba, Pontifícia Universidade Católica do Paraná, 2010.
- MARCONI, Marina de Andrade. LAKATOS, Eva Maria. **Fundamentos de Metodologia Científica.** 5. ed. São Paulo: Atlas, 2003.
- MEINEN, Ênio. Cooperativismo de Crédito: Raízes Evolução e Particularidades. In: MEINEN, E.; DOMINGUES, J. N.; DOMINGUES, J. A. S. **Cooperativas de crédito no direito brasileiro.** Porto Alegre: Ed. Sagra Luzzatto, 2002.
- MENEZES, Antônio. **Cooperativa de Crédito: O que é e Quais Seus Benefícios.** Brasília: Confebrás, 2004.
- MINTZBERG, Henry. *Generic Strategies. Advances in Strategic Management*, v. 5, 1988.
- MINTZBERG, Henry. AHLSTRAND, Bruce. LAMPEL, Joseph. **Safári de Estratégia.** 2. Ed. Porto Alegre: Bookman, 2010.
- MORGAN, Andrew Sturdy Glenn. "Segmenting the Market: A Review of Marketing Trends in French Retail Banking". *International Journal of Bank Marketing*, v. 11, Iss 7, pp. 11 – 19, 1993.
- NILSSON, N.J. 1996. **Introduction to machine learning.** Stanford : Stanford university, 1996.
- Organização das Cooperativas do Brasil. Disponível em: <http://www.portaldocooperativismo.org.br/sescoop/cooperativismo/estrutura_cooperativismo.asp> . Acesso em 20 agosto de 2016.
- PINHEIRO, Marcos Antônio Henriques. **Cooperativas de Crédito: História da evolução normativa no Brasil.** 6. ed. Brasília: Banco Centra do Brasil, 2008.
- PINHO, Diva Benevides. **Cooperativas e Desenvolvimento Econômico.** São Paulo: Boletim 289, 1963.
- PINHO, Diva Benevides. **Economia e Cooperativismo.** São Paulo: Saraiva, 1977.
- PORTER, Michael E. **Competição On Competition Estratégias Competitivas Essenciais.** Rio de Janeiro: Campus, 1999.

- PRATI, Ronaldo Cristiano. BARANAUKAS, José Augusto. MONARD, Maria Carolina. **Uma Proposta de Unificação da Linguagem de Representação de Conceitos de Algoritmos de Aprendizado de Máquina Simbólicos**. Instituto de Ciências Matemáticas e de Computação. Universidade de São Paulo, ISSN – 01030-2569 São Carlos, 2001.
- QUINLAN, J. R. *Learning Decision Tree Classifiers*. Sydney, 1996.
- RANJAN, Jayanthi. BHATNAGAR, Vichal. *A holistic framework for mCRM – data mining perspective*. *Information Management & Computer Security*, Índia, 2009.
- REZENDE, Solange Oliveira. Mineração de Dados. **Anais do XXV Congresso da Sociedade Brasileira de Computação**, UNISINOS 22 a 29 de Julho de 2005 pag. 397 a 433, São Leopoldo, RS.
- RICHERS, Raimar. LIMA, Cecília Pimenta. **Segmentação: Opções Estratégicas para o Mercado Brasileiro**. São Paulo: Nobel, 1991.
- SANDRONI, Paulo. **Novíssimo Dicionário de Economia**. São Paulo: Best Seller, 2003.
- SANTANDER. Disponível em:
<<https://www.santander.com.br/portal/wps/script/templates/GCMRequest.do?page=6779&entryID=7829>> acesso em 03 de maio de 2016.
- SICOOB Norte do Paraná – Cooperativa de Crédito de Livre Admissão do Norte do Paraná. Acervo pessoal.
- SINGH, S. P. YADAV, Asmita. *Study of K-Means and Enhanced K-Means Clustering Algorithm*. *International Journal of Advanced Research in Computer Science*, v. 4, nº10, 2013.
- TAVARES, Mauro Calixta. AFONSO, Tarcisio. LOCATELLI, Ronaldo Lamounier. Segmentação de Mercado, Diferenciação de Produtos e a Perspectiva da Antropologia do Consumo. **Revista Gestão & Tecnologia, Pedro Leopoldo, v. 11, n. 1, p. 106-122, jan./jun. 2011**
- VASSILVITSKII, Sergei. *K-means: Algorithms, Analyses, Experiments*. Stanford University, Stanford, 2007.
- ZINELDIN, Mosad. *Bank strategic positioning and some determinants of bank selection*. *International Journal of Bank Marketing*, v. 14, Iss 6, pp. 12-22, 1996.
- WANG, Xiaoyan. BAI, Yanping. *The Global Minmax k- means Algorithm*. SpringerPlus, School of Science, North University of China, Republic of China, 2016.
- WILLS, Gordon. *Dividing and Conquering: Strategies for Segmentation*. *International Journal of Bank Marketing*, v. 3, Iss 4, pp. 36 – 46, 1985.
- WITTEN, Ian H. FRANK, Eibe. *Data mining : practical machine learning tools and techniques*. 2. ed. Elsevier San Francisco, 2005.

WOCCU. **Word Council of Credit Unions**. Disponível em:
<<http://www.woccu.org/publications/statreport>> Acesso em: 03 de Agosto de 2016.

APÊNDICE A – ESTATÍSTICA DESCRITIVA PESSOA FÍSICA

<i>Renda / Faturamento</i>		<i>Qtde CC</i>		<i>Qtde CC Limite</i>		<i>Capital</i>	
Média	5729,25453	Média	1,0153416	Média	1,006434135	Média	1798,785663
Erro padrão	288,130142	Erro padrão	0,00111692	Erro padrão	0,000849524	Erro padrão	130,9024841
Mediana	1800	Mediana	1	Mediana	1	Mediana	68,925
Modo	0	Modo	1	Modo	1	Modo	50
Desvio padrão	38751,094	Desvio padrão	0,13191496	Desvio padrão	0,079959104	Desvio padrão	17605,28916
Variância da amostra	1501647283	Variância da amostra	0,01740156	Variância da amostra	0,006393458	Variância da amostra	309946206,5
Curtose	2695,24406	Curtose	135,765718	Curtose	150,5131415	Curtose	2625,79772
Assimetria	42,8601551	Assimetria	10,0826108	Assimetria	12,34824529	Assimetria	44,70638606
Intervalo	3006978,4	Intervalo	4	Intervalo	1	Intervalo	1246331,23
Mínimo	0	Mínimo	1	Mínimo	1	Mínimo	25
Máximo	3006978,4	Máximo	5	Máximo	2	Máximo	1246356,23
Soma	103630756	Soma	14163	Soma	8916	Soma	32536435,08
Contagem	18088	Contagem	13949	Contagem	8859	Contagem	18088
<i>DAP</i>		<i>DAV</i>		<i>Limite Rotativo</i>		<i>Limite Rotativo Utilizado</i>	
Média	10082,1208	Média	1200,24255	Média	1362,081621	Média	1204,723122
Erro padrão	677,542293	Erro padrão	111,20157	Erro padrão	350,7236596	Erro padrão	407,6962232
Mediana	0	Mediana	0	Mediana	0	Mediana	0
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	91123,7711	Desvio padrão	14955,681	Desvio padrão	28946,92153	Desvio padrão	30023,16621
Variância da amostra	8303541655	Variância da amostra	223672394	Variância da amostra	837924266	Variância da amostra	901390509,1
Curtose	3351,86108	Curtose	4030,33837	Curtose	2014,395425	Curtose	2155,096884
Assimetria	45,9695702	Assimetria	57,0178771	Assimetria	40,03235973	Assimetria	42,86339122
Intervalo	7832665,77	Intervalo	1242304,68	Intervalo	1700000	Intervalo	1700104,22
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	-104,22
Máximo	7832665,77	Máximo	1242304,68	Máximo	1700000	Máximo	1700000
Soma	182365401	Soma	21709987,2	Soma	9278500	Soma	6533213,49
Contagem	18088	Contagem	18088	Contagem	6812	Contagem	5423
<i>Limite TD</i>		<i>TD Utilizado</i>		<i>Limite c/c</i>		<i>Limite c/c Utilizado</i>	
Média	645,818165	Média	385,515979	Média	1742,119451	Média	731,5749894
Erro padrão	222,952544	Erro padrão	237,041832	Erro padrão	31,89124762	Erro padrão	23,33565091
Mediana	0	Mediana	0	Mediana	600	Mediana	0
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	16426,0202	Desvio padrão	17449,5636	Desvio padrão	3526,975904	Desvio padrão	2154,728636
Variância da amostra	269814139	Variância da amostra	304487272	Variância da amostra	12439559,03	Variância da amostra	4642855,496
Curtose	3727,74925	Curtose	5010,39708	Curtose	110,94539	Curtose	219,3680753
Assimetria	57,2759903	Assimetria	69,6547797	Assimetria	7,49212868	Assimetria	10,57832613
Intervalo	1100000	Intervalo	1259636,26	Intervalo	100000	Intervalo	70000
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	0
Máximo	1100000	Máximo	1259636,26	Máximo	100000	Máximo	70000
Soma	3505501	Soma	2089111,09	Soma	21307863	Soma	6237408,36
Contagem	5428	Contagem	5419	Contagem	12231	Contagem	8526

<i>ADP</i>		<i>Financiamento</i>		<i>Consignado</i>		<i>Demais Crédito</i>	
Média	169,872527	Média	162,65435	Média	18,32540847	Média	8466,901673
Erro padrão	10,6124827	Erro padrão	30,305537	Erro padrão	8,318907851	Erro padrão	663,0712778
Mediana	0	Mediana	0	Mediana	0	Mediana	0
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	894,853204	Desvio padrão	2232,5536	Desvio padrão	610,5194273	Desvio padrão	59812,55769
Variância da amostra	800762,257	Variância da amostra	4984295,59	Variância da amostra	372733,9711	Variância da amostra	3577542057
Curtose	441,246422	Curtose	605,24069	Curtose	2064,068412	Curtose	2710,014889
Assimetria	16,1288809	Assimetria	21,6291564	Assimetria	43,20042677	Assimetria	44,19560185
Intervalo	35327,32	Intervalo	81000	Intervalo	32646,67	Intervalo	4069642,74
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	0
Máximo	35327,32	Máximo	81000	Máximo	32646,67	Máximo	4069642,74
Soma	1207793,67	Soma	882725,16	Soma	98700,65	Soma	68895178,91
Contagem	7110	Contagem	5427	Contagem	5386	Contagem	8137
<i>Crédito Rural</i>		<i>Pré-Aprovado</i>		<i>Micro Crédito</i>		<i>Crédito Total Utilizado</i>	
Média	5197,11802	Média	545,157946	Média	0	Média	10392,99851
Erro padrão	798,838699	Erro padrão	22,8940563	Erro padrão	0	Erro padrão	709,8828657
Mediana	0	Mediana	0	Mediana	0	Mediana	402,46
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	59097,8688	Desvio padrão	1798,60462	Desvio padrão	0	Desvio padrão	75561,68758
Variância da amostra	3492558096	Variância da amostra	3234978,6	Variância da amostra	0	Variância da amostra	5709568629
Curtose	302,315816	Curtose	48,9812975	Curtose	0	Curtose	1080,756742
Assimetria	16,2244032	Assimetria	5,89578112	Assimetria	0	Assimetria	27,46960754
Intervalo	1420503,91	Intervalo	24815,32	Intervalo	0	Intervalo	4069746,96
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	-104,22
Máximo	1420503,91	Máximo	24815,32	Máximo	0	Máximo	4069642,74
Soma	28443826,9	Soma	3364714,84	Soma	0	Soma	117752673,1
Contagem	5473	Contagem	6172	Contagem	5384	Contagem	11330
<i>Convênio água</i>		<i>Convênio Luz</i>		<i>Convênio Telefone</i>		<i>Convênio Cartão</i>	
Média	0,06564121	Média	0,11514196	Média	0,117782094	Média	0,268165825
Erro padrão	0,00378991	Erro padrão	0,00568844	Erro padrão	0,006067554	Erro padrão	0,006121214
Mediana	0	Mediana	0	Mediana	0	Mediana	0
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	0,28338286	Desvio padrão	0,4296936	Desvio padrão	0,456601049	Desvio padrão	0,494910651
Variância da amostra	0,08030585	Variância da amostra	0,18463659	Variância da amostra	0,208484518	Variância da amostra	0,244936553
Curtose	29,5327455	Curtose	64,5220507	Curtose	62,30257345	Curtose	2,977380706
Assimetria	4,96208751	Assimetria	6,0921091	Assimetria	5,9922906	Assimetria	1,723364097
Intervalo	4	Intervalo	9	Intervalo	10	Intervalo	5
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	0
Máximo	4	Máximo	9	Máximo	10	Máximo	5
Soma	367	Soma	657	Soma	667	Soma	1753
Contagem	5591	Contagem	5706	Contagem	5663	Contagem	6537

<i>Cobrança C. Registro</i>		<i>Cobrança S. Registro</i>		<i>Seguro Vida</i>		<i>Seguro Auto</i>	
Média	0,00265369	Média	0	Média	0,148496241	Média	0,030296329
Erro padrão	0,00038253	Erro padrão	0	Erro padrão	0,002644038	Erro padrão	0,001274475
Mediana	0	Mediana	0	Mediana	0	Mediana	0
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	0,05144703	Desvio padrão	0	Desvio padrão	0,355601038	Desvio padrão	0,171406201
Variância da amostra	0,0026468	Variância da amostra	0	Variância da amostra	0,126452098	Variância da amostra	0,029380086
Curtose	371,939132	Curtose	0	Curtose	1,90942962	Curtose	28,04662606
Assimetria	19,3364424	Assimetria	0	Assimetria	1,977174374	Assimetria	5,481197401
Intervalo	1	Intervalo	0	Intervalo	1	Intervalo	1
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	0
Máximo	1	Máximo	0	Máximo	1	Máximo	1
Soma	48	Soma	0	Soma	2686	Soma	548
Contagem	18088	Contagem	18088	Contagem	18088	Contagem	18088
<i>Seguro Residencial</i>		<i>Seguro Diversos</i>		<i>Cartão Débito</i>		<i>Cartão Crédito</i>	
Média	0,03676471	Média	0,00044228	Média	0,658226448	Média	0,466994693
Erro padrão	0,00139926	Erro padrão	0,00015634	Erro padrão	0,003526741	Erro padrão	0,003709697
Mediana	0	Mediana	0	Mediana	1	Mediana	0
Modo	0	Modo	0	Modo	1	Modo	0
Desvio padrão	0,18818879	Desvio padrão	0,02102644	Desvio padrão	0,474317224	Desvio padrão	0,498923252
Variância da amostra	0,03541502	Variância da amostra	0,00044211	Variância da amostra	0,224976829	Variância da amostra	0,248924412
Curtose	22,2446482	Curtose	2256,62452	Curtose	-1,55495016	Curtose	-1,98271045
Assimetria	4,92363571	Assimetria	47,5223632	Assimetria	-0,66724942	Assimetria	0,132320781
Intervalo	1	Intervalo	1	Intervalo	1	Intervalo	1
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	0
Máximo	1	Máximo	1	Máximo	1	Máximo	1
Soma	665	Soma	8	Soma	11906	Soma	8447
Contagem	18088	Contagem	18088	Contagem	18088	Contagem	18088
<i>Cons. Auto</i>		<i>Cons. Moto</i>		<i>Cons. Imóveis</i>			
Média	0,02780849	Média	0,00895621	Média	0		
Erro padrão	0,0018903	Erro padrão	0,00082106	Erro padrão	0		
Mediana	0	Mediana	0	Mediana	0		
Modo	0	Modo	0	Modo	0		
Desvio padrão	0,25422969	Desvio padrão	0,1104255	Desvio padrão	0		
Variância da amostra	0,06463274	Variância da amostra	0,01219379	Variância da amostra	0		
Curtose	1116,38792	Curtose	519,814485	Curtose			
Assimetria	24,8248983	Assimetria	18,728218	Assimetria			
Intervalo	16	Intervalo	4	Intervalo	0		
Mínimo	0	Mínimo	0	Mínimo	0		
Máximo	16	Máximo	4	Máximo	0		
Soma	503	Soma	162	Soma	0		
Contagem	18088	Contagem	18088	Contagem	18088		

APÊNDICE B – ESTATÍSTICA DESCRITIVA PESSOA JURÍDICA

<i>Renda / Faturamento</i>		<i>Qtde CC</i>		<i>Qtde CC Limite</i>		<i>Capital</i>	
Média	462628,659	Média	1,09602411	Média	1,003695492	Média	3921,375741
Erro padrão	61942,377	Erro padrão	0,00548248	Erro padrão	0,001381155	Erro padrão	401,2124877
Mediana	37199,875	Mediana	1	Mediana	1	Mediana	607,32
Modo	0	Modo	1	Modo	1	Modo	500
Desvio padrão	5520869,54	Desvio padrão	0,45761491	Desvio padrão	0,071846636	Desvio padrão	35759,71589
Variância da amostra	3,048E+13	Variância da amostra	0,20941141	Variância da amostra	0,005161939	Variância da amostra	1278757281
Curtose	1671,25715	Curtose	68,9575559	Curtose	520,8798174	Curtose	1497,055962
Assimetria	36,5924622	Assimetria	7,0166216	Assimetria	21,79153488	Assimetria	33,22548173
Intervalo	311800380	Intervalo	9	Intervalo	2	Intervalo	1999104,74
Mínimo	0	Mínimo	1	Mínimo	1	Mínimo	50
Máximo	311800380	Máximo	10	Máximo	3	Máximo	1999154,74
Soma	3675122071	Soma	7636	Soma	2716	Soma	31151408,89
Contagem	7944	Contagem	6967	Contagem	2706	Contagem	7944
<i>DAP</i>		<i>DAV</i>		<i>Limite Rotativo</i>		<i>Limite Rotativo Utilizado</i>	
Média	28843,0987	Média	8238,12782	Média	26968,36972	Média	24154,21943
Erro padrão	3227,60784	Erro padrão	691,315519	Erro padrão	4261,777978	Erro padrão	4382,633615
Mediana	0	Mediana	41,305	Mediana	0	Mediana	0
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	287673,846	Desvio padrão	61616,3437	Desvio padrão	223408,2389	Desvio padrão	205049,1844
Variância da amostra	8,2756E+10	Variância da amostra	3796573810	Variância da amostra	49911241219	Variância da amostra	42045168035
Curtose	1365,04104	Curtose	1440,56258	Curtose	211,4667543	Curtose	252,7996582
Assimetria	31,5254362	Assimetria	33,9957604	Assimetria	13,3494344	Assimetria	14,28785973
Intervalo	15392383,2	Intervalo	3070512,64	Intervalo	5000000	Intervalo	5002620,88
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	-2620,88
Máximo	15392383,2	Máximo	3070512,64	Máximo	5000000	Máximo	5000000
Soma	229129576	Soma	65443687,4	Soma	74109080	Soma	52873586,33
Contagem	7944	Contagem	7944	Contagem	2748	Contagem	2189
<i>Limite TD</i>		<i>TD Utilizado</i>		<i>Limite c/c</i>		<i>Limite c/c Utilizado</i>	
Média	28604,4308	Média	14548,7983	Média	7089,709473	Média	3065,49069
Erro padrão	2863,52554	Erro padrão	2001,4459	Erro padrão	336,8978122	Erro padrão	241,3202841
Mediana	0	Mediana	0	Mediana	3000	Mediana	0
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	143975,832	Desvio padrão	99791,6992	Desvio padrão	21561,45998	Desvio padrão	13531,16209
Variância da amostra	2,0729E+10	Variância da amostra	9958383238	Variância da amostra	464896556,5	Variância da amostra	183092347,6
Curtose	278,548601	Curtose	472,440134	Curtose	1122,698561	Curtose	1265,547786
Assimetria	14,278774	Assimetria	19,0441662	Assimetria	26,68416458	Assimetria	30,40446066
Intervalo	3800000	Intervalo	3074041,84	Intervalo	1000000	Intervalo	603581,36
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	0
Máximo	3800000	Máximo	3074041,84	Máximo	1000000	Máximo	603581,36
Soma	72312001	Soma	36168312,5	Soma	29039450	Soma	9637902,73
Contagem	2528	Contagem	2486	Contagem	4096	Contagem	3144

<i>ADP</i>		<i>Financiamento</i>		<i>Consignado</i>		<i>Demais Crédito</i>	
Média	433,456857	Média	624,945375	Média	0	Média	51640,31504
Erro padrão	38,3709359	Erro padrão	164,771329	Erro padrão	0	Erro padrão	5355,656014
Mediana	0	Mediana	0	Mediana	0	Mediana	0
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	2104,11322	Desvio padrão	7608,08305	Desvio padrão	0	Desvio padrão	286965,2251
Variância da amostra	4427292,46	Variância da amostra	57882927,7	Variância da amostra	0	Variância da amostra	82349040410
Curtose	101,306389	Curtose	425,245151	Curtose		Curtose	177,3484042
Assimetria	9,18303757	Assimetria	19,3228561	Assimetria		Assimetria	11,72137446
Intervalo	32844,52	Intervalo	192258,66	Intervalo	0	Intervalo	6326960,52
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	0
Máximo	32844,52	Máximo	192258,66	Máximo	0	Máximo	6326960,52
Soma	1303404,77	Soma	1332383,54	Soma	0	Soma	148259344,5
Contagem	3007	Contagem	2132	Contagem	2092	Contagem	2871
<i>Crédito Rural</i>		<i>Pré-Aprovado</i>		<i>Micro Crédito</i>		<i>Crédito Total Utilizado</i>	
Média	18629,0847	Média	0	Média	0	Média	66807,26041
Erro padrão	5848,19472	Erro padrão	0	Erro padrão	0	Erro padrão	5559,056791
Mediana	0	Mediana	0	Mediana	0	Mediana	328,6
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	268380,531	Desvio padrão	0	Desvio padrão	0	Desvio padrão	365505,3419
Variância da amostra	7,2028E+10	Variância da amostra	0	Variância da amostra	0	Variância da amostra	1,33594E+11
Curtose	415,535616	Curtose		Curtose		Curtose	156,5827676
Assimetria	18,9170785	Assimetria		Assimetria		Assimetria	11,13595889
Intervalo	7506455,45	Intervalo	0	Intervalo	0	Intervalo	7509076,33
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	-2620,88
Máximo	7506455,45	Máximo	0	Máximo	0	Máximo	7506455,45
Soma	39232852,4	Soma	0	Soma	0	Soma	288807786,8
Contagem	2106	Contagem	2092	Contagem	2092	Contagem	4323
<i>Convênio água</i>		<i>Convênio Luz</i>		<i>Convênio Telefone</i>		<i>Convênio Cartão</i>	
Média	0,32732487	Média	0,59412232	Média	0,34084272	Média	0,192931393
Erro padrão	0,02019653	Erro padrão	0,04255849	Erro padrão	0,024001601	Erro padrão	0,009229315
Mediana	0	Mediana	0	Mediana	0	Mediana	0
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	0,99783935	Desvio padrão	2,13557128	Desvio padrão	1,175098386	Desvio padrão	0,452612969
Variância da amostra	0,99568337	Variância da amostra	4,56066468	Variância da amostra	1,380856218	Variância da amostra	0,2048585
Curtose	136,554	Curtose	190,923106	Curtose	729,0018211	Curtose	5,045788396
Assimetria	9,40113043	Assimetria	11,4273092	Assimetria	21,1102535	Assimetria	2,342359885
Intervalo	20	Intervalo	51	Intervalo	43	Intervalo	3
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	0
Máximo	20	Máximo	51	Máximo	43	Máximo	3
Soma	799	Soma	1496	Soma	817	Soma	464
Contagem	2441	Contagem	2518	Contagem	2397	Contagem	2405

<i>Cobrança C. Registro</i>		<i>Cobrança S. Registro</i>		<i>Seguro Vida</i>		<i>Seguro Auto</i>	
Média	0,13582578	Média	0	Média	0,069864048	Média	0,010448137
Erro padrão	0,00384414	Erro padrão	0	Erro padrão	0,002860275	Erro padrão	0,001140897
Mediana	0	Mediana	0	Mediana	0	Mediana	0
Modo	0	Modo	0	Modo	0	Modo	0
Desvio padrão	0,34262504	Desvio padrão	0	Desvio padrão	0,254933804	Desvio padrão	0,101687143
Variância da amostra	0,11739192	Variância da amostra	0	Variância da amostra	0,064991244	Variância da amostra	0,010340275
Curtose	2,52188884	Curtose		Curtose	9,395292768	Curtose	90,77928396
Assimetria	2,12632406	Assimetria		Assimetria	3,375341087	Assimetria	9,631013925
Intervalo	1	Intervalo	0	Intervalo	1	Intervalo	1
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	0
Máximo	1	Máximo	0	Máximo	1	Máximo	1
Soma	1079	Soma	0	Soma	555	Soma	83
Contagem	7944	Contagem	7944	Contagem	7944	Contagem	7944
<i>Seguro Residencial</i>		<i>Seguro Diversos</i>		<i>Cartão Débito</i>		<i>Cartão Crédito</i>	
Média	0,00075529	Média	0,00629406	Média	0,520140987	Média	0,168806647
Erro padrão	0,00030825	Erro padrão	0,00088736	Erro padrão	0,005605638	Erro padrão	0,004202943
Mediana	0	Mediana	0	Mediana	1	Mediana	0
Modo	0	Modo	0	Modo	1	Modo	0
Desvio padrão	0,02747383	Desvio padrão	0,07909002	Desvio padrão	0,499625624	Desvio padrão	0,374604628
Variância da amostra	0,00075481	Variância da amostra	0,00625523	Variância da amostra	0,249625764	Variância da amostra	0,140328627
Curtose	1319,83207	Curtose	153,98399	Curtose	-1,99399852	Curtose	1,128492211
Assimetria	36,3524386	Assimetria	12,487803	Assimetria	-0,08064462	Assimetria	1,768674126
Intervalo	1	Intervalo	1	Intervalo	1	Intervalo	1
Mínimo	0	Mínimo	0	Mínimo	0	Mínimo	0
Máximo	1	Máximo	1	Máximo	1	Máximo	1
Soma	6	Soma	50	Soma	4132	Soma	1341
Contagem	7944	Contagem	7944	Contagem	7944	Contagem	7944
<i>Cons. Auto</i>		<i>Cons. Moto</i>		<i>Cons. Imóveis</i>			
Média	0,04556898	Média	0,00679758	Média	0		
Erro padrão	0,00478628	Erro padrão	0,00101988	Erro padrão	0		
Mediana	0	Mediana	0	Mediana	0		
Modo	0	Modo	0	Modo	0		
Desvio padrão	0,42659658	Desvio padrão	0,09090099	Desvio padrão	0		
Variância da amostra	0,18198464	Variância da amostra	0,00826299	Variância da amostra	0		
Curtose	495,975702	Curtose	555,9771	Curtose			
Assimetria	19,2437876	Assimetria	18,8879619	Assimetria			
Intervalo	16	Intervalo	4	Intervalo	0		
Mínimo	0	Mínimo	0	Mínimo	0		
Máximo	16	Máximo	4	Máximo	0		
Soma	362	Soma	54	Soma	0		
Contagem	7944	Contagem	7944	Contagem	7944		

