

LUAN DE OLIVEIRA DAPUNT

Uma abordagem de segmentação de clientes utilizando o método *recency-frequency-monetary* adaptado e o método *Analytic Hierarchy Process*

Luan de Oliveira Dapunt

Uma abordagem de segmentação de clientes utilizando o método *recency-frequency-monetary* adaptado e o método *Analytic Hierarchy Process*

Trabalho de Graduação apresentado ao Conselho de Curso de Graduação em Engenharia de Produção Mecânica da Faculdade de Engenharia do Campus de Guaratinguetá, Universidade Estadual Paulista, como parte dos requisitos para obtenção do diploma de Graduação em Engenharia de Produção Mecânica.

Orientador: Prof. Dr. José Roberto Dale Luche

	Dapunt, Luan de Oliveira
D212a	Uma Abordagem de segmentação de clientes utilizando o método recency-frequency-monetary adaptado e método Analytic Hierarchy Process / Luan de Oliveira Dapunt – Guaratinguetá, 2018.
	52 f : il.
	Bibliografia: f. 38-40
	Trabalho de Graduação em Engenharia de Produção Mecânica – Universidade Estadual Paulista, Faculdade de Engenharia de Guaratinguetá, 2018.
	Orientador: Prof. Dr. José Roberto Dale Luche
	Co-orientador: Prof. Stefano Petrini de Oliveira
	1. Processo decisório. 2. Segmentação de mercado. 3. Planejamento estratégico. I. Título
	CDU 65.012.4

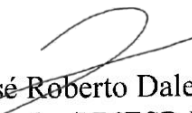
Luan de Oliveira Dapunt

ESTE TRABALHO DE GRADUAÇÃO FOI JULGADO ADEQUADO COMO
PARTE DO REQUISITO PARA A OBTENÇÃO DO DIPLOMA DE
"GRADUADO EM ENGENHARIA DE PRODUÇÃO MECÂNICA"

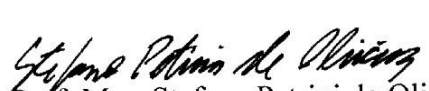
APROVADO EM SUA FORMA FINAL PELO CONSELHO DE CURSO DE
GRADUAÇÃO EM NOME DO CURSO

Prof. Dr. Andreia Maria Pedro Salgado
Coordenador

BANCA EXAMINADORA:


Prof. Dr. José Roberto Dale Luche
Orientador/UNESP-FEG


Prof. Dr. Aneirson Francisco da Silva
UNESP-FEG


Prof. Msc. Stefano Petrini de Oliveira
Membro Externo

DADOS CURRICULARES

LUAN DE OLIVEIRA DAPUNT

NASCIMENTO	19.05.1994 – São Paulo / SP
FILIAÇÃO	Luis Umberto Dapunt Heloisa de Oliveira
2012/2018	Engenharia de Produção Mecânica – Bacharelado Faculdade de Engenharia de Guaratinguetá Universidade Estadual Paulista “Júlio de Mesquita Filho”

dedico este trabalho
de modo especial, à minha família

AGRADECIMENTOS

Em primeiro lugar agradeço à minha mãe *Heloisa de Oliveira*, que apesar de todas as dificuldades enfrentadas durante minha trajetória acadêmica e desenvolvimento pessoal, sempre incentivou meus estudos e me apoiou em todas as minhas decisões e desafios. Também agradeço aos meus irmãos *Kauê de Oliveira Dapunt* e *Endy de Oliveira Dapunt* que sempre estiveram ao meu lado em todos os momentos. Todas as minhas conquistas são dedicadas à minha família e devo tudo a eles.

Aos meninos da *República Masmorra* que são minha segunda família e minha segunda casa. Eles fizeram parte de toda a minha graduação, sempre me apoiando todos os dias. Sem eles não seria possível ter concluindo minha graduação.

Aos meus amigos que sempre me apoiaram e que foram essenciais nesta trajetória.

Ao meu orientador, *Prof. Dr. José Roberto Dale Luche* que me auxiliou neste projeto referente ao trabalho de graduação.

E aos professores e colaboradores da Faculdade de Engenharia do Campos de Guaratinguetá, responsáveis pelo nosso aprendizado e nossa formação como engenheiros.

RESUMO

A constante mudança no contexto empresarial sugere necessidades estratégias para compreender melhor os clientes afim de se destacar perante à alta competitividade de mercado. A segmentação de clientes possibilita com que sejam identificados fatores comuns entre um conjunto de clientes e auxiliados em tomadas de decisões direcionadas aos grupos de clientes. O método *recency-frequency-monetary* (RFM) é utilizado para mapear as características dos clientes e a técnica de mineração de dados de agrupamento é utilizada para classificar os clientes. Além disso, um dos métodos mais aplicados para métodos de tomada de decisão multicritério, o *Analytic Hierarchy Process* (AHP) foi utilizado para definir os pesos dos critérios analisados. Por meio de uma pesquisa ação em uma empresa de multinegócios, o objetivo da pesquisa é de segmentar os clientes, que possuem uma relação *Bussiness to Bussiness* (B2B) com a companhia, para auxiliar em tomadas de decisões relacionadas à sua estratégia de marketing. Os resultados obtidos foram condizentes com a realidade em que a empresa em questão está inserida e auxiliará em estratégias de mercado.

PALAVRAS-CHAVE: Segmentação de cliente. MCDM. Decisão multicritério. RFM. AHP.

ABSTRACT

The business context is in a constant change, so strategies are needed to better understand clients in order to be able to stand out against the high market competitiveness. Customer segmentation helps to identify common factors between a set of customers and assist in decision-making directed to customers groups. The recency-frequency-monetary (RFM) method is used to map the factors and clustering, a technique from Data Mining, is used to classify the customers. In addition, one of the most applied methods for multicriteria decision-making methods, the Analytic Hierarchy Process (AHP) was used to define the weights of the criteria analyzed in this graduate thesis. The research was based on a case in a multi-business company, whose objective is to segment customers, who have a Business-to-Business (B2B) relationship with this company, to aid in decision-making related to their marketing strategy. The results obtained were consistent with the reality in which the company in question is inserted and will assist in market strategies.

KEYWORDS: Customer segmentation. MCDM. Multicriteria decision-making. RFM. AHP.

LISTA DE ILUSTRAÇÕES

Figura 1 – Pesquisa bibliométrica sobre métodos de decisão multicritério	14
Figura 2 – Pesquisa bibliométrica sobre o tópico <i>Customer Segmentation</i>	15
Figura 3 – Pesquisa bibliométrica sobre o tópico AHP	15
Figura 4 – Pesquisa bibliométrica sobre o tópico AHP e <i>Customer Segmentation</i>	16
Figura 5 – Dimensões do ciclo de gestão de CRM	18
Figura 6 – Aplicações CRM analítico	19
Figura 7 – Classificação de técnicas de mineração de dados no ciclo de gestão de CRM.....	20
Figura 8 – Ciclo de vida e etapas do CRISP-DM.....	22
Figura 9 – Técnicas de mineração de dados	24
Figura 10 – Exemplo de agrupamento.....	25
Figura 11 – Estrutura hierárquica do método AHP	29
Figura 12 – Método e classificação de pesquisa.....	31
Figura 13 – Estrutura hierárquica AHP	36
Figura 14 – Método <i>Elbow</i>	37
Figura 15 – Número de iterações do algoritmo <i>K-Means</i>	39
Figura 16 – Gráfico de dispersão da segmentação de clientes	40
Figura 17 – Média de características para cada <i>cluster</i>	41
Figura 18 – Silhueta de cada <i>cluster</i>	42

LISTA DE TABELAS

Tabela 1 – Interpretação de resultados da silhueta média	27
Tabela 2 – Escala fundamental AHP	29
Tabela 3 – Índice randômico por ordem de matriz	30
Tabela 4 – Matriz de comparação de critérios.....	36
Tabela 5 – Cálculo de consistência.....	36
Tabela 6 – Centroides <i>cluster</i> iniciais	38
Tabela 7 – Centroides <i>cluster</i> finais e <i>ranking</i>	38
Tabela 8 – Sumário de iterações do algoritmo <i>K-Means</i>	39
Tabela 9 – Resultados do algoritmo <i>K-Means</i> para os quatro <i>clusters</i>	39
Tabela 10 – Distância entre centroides dos <i>clusters</i>	40

SUMÁRIO

1	INTRODUÇÃO	11
1.1	CONSIDERAÇÕES INICIAIS	11
1.2	OBJETIVO E QUESTÃO DE PESQUISA.....	12
1.3	DELIMITAÇÃO DA PESQUISA	12
1.4	JUSTIFICATIVA DA PESQUISA	13
1.5	ESTRUTURA DO TRABALHO	16
2	REFERENCIAL TEÓRICO	17
2.1	GESTÃO DE RELACIONAMENTO COM O CLIENTE	17
2.2	MINERAÇÃO DE DADOS	21
2.3	TOMADA DE DECISÃO MULTICRITÉRIO	27
3	MÉTODO E CLASSIFICAÇÃO DE PESQUISA.....	31
4	DESENVOLVIMENTO E RESULTADOS.....	32
4.1	DESCRIÇÃO DA EMPRESA E DA ÁREA	32
4.2	DETERMINAÇÃO DAS CARACTERÍSTICAS DE AVALIAÇÃO DOS CLIENTES	33
4.3	OBTENÇÃO E PREPARAÇÃO DE DADOS	34
4.4	MODELAGEM MÉTODO AHP	35
4.5	APLICAÇÃO MÉTODO K-MEANS E RESULTADOS	37
5	CONSIDERAÇÕES FINAIS.....	43
	REFERÊNCIAS	45
	APÊNDICE A – Dados primários	49
	APÊNDICE B – Dados normalizados	52
	APÊNDICE C – Pontuação geral e <i>ranking</i>	55
	APÊNDICE D – Dados clientes por <i>cluster</i>	58

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

O cenário empresarial tem passado por mudanças radicais nas últimas décadas, com mudanças nas condições de mercado, exigências dos clientes e crescente tecnologia (VINODH; JOY, 2012).

A necessidade das empresas em entender melhor os seus clientes tem sido cada vez mais importante para criar, sustentar e melhorar os seus relacionamentos levando a melhores resultados financeiros. As empresas devem alocar seus recursos nos clientes que serão mais lucrativos para a companhia, porém as tomadas de decisões nestes casos são complexos, uma vez que as empresas possuem diversos clientes e um plano de marketing em nível individual é uma prática inviável (MENDE; BOLTON; BITNER, 2013).

Com a finalidade de solucionar este problema, a segmentação de clientes tem como principal objetivo identificar elementos em comum entre um conjunto de clientes, como preferências, atitudes e valor percebido. Esta divisão de clientes tornou-se um conceito central em marketing e auxilia diversas empresas a entender melhor as necessidades de clientes para fazer tomadas de decisões direcionadas à suas estratégias de marketing (FLOH et al., 2014).

Com o intuito de realizar melhor esta segmentação é fundamental a compreensão de quais os critérios mais relevantes com o que as empresas entendem como prioritários (TAVARES; AFONSO; LOCATELLI, 2011).

A elevada relevância na tomada de decisão de quais os critérios mais importantes faz com que aumente a exigência e a complexidade do processo, por conta principalmente da natureza e quantidade destes incorporados na análise. Além disso, em diversas ocorrências os critérios qualitativos e quantitativos são conflitantes. A existência de métodos que combinam diferentes aspectos ao avaliar as alternativas vêm a suprir estas divergências (VIANA; ALENCAR, 2012).

Dentro deste contexto, este trabalho utiliza um modelo adaptado do método *Recency-Frequency-Monetary* (RFM) para estabelecer os critérios mais importantes a serem analisados para a segmentação de clientes de uma empresa multinacional. Após seleção de critérios é aplicado o método *Analytic Hierarchy Process* (AHP) para classificar a importância de cada um destes. Por fim, a segmentação de clientes é feita com uma técnica de mineração de dados, o *K-Means*, onde se tem por objetivo particionar a um número k definido de grupos mediante os diversos critérios a serem analisados.

1.2 OBJETIVO E QUESTÃO DE PESQUISA

Este trabalho tem como objetivo geral realizar a segmentação de clientes estabelecendo critérios importantes para a empresa em questão, determinando índices de relevância para cada um destes para classificar os grupos, possibilitando auxílio em tomadas de decisões de acordo com o direcionamento estratégico de marketing da companhia.

Os objetivos específicos foram:

- Identificação dos critérios para realizar a segmentação de clientes por meio do método RFM adaptado.
- Definição da relação de importância entre os critérios por meio de um método de decisão multicritério, o método AHP.
- Clusterização de clientes por meio do algoritmo de mineração de dados *K-Means*.

Com embasamento nos objetivos, o trabalho procurou responder as seguintes questões: Quais os benefícios de realizar a segmentação de clientes? Como selecionar critérios importantes, estabelecendo pesos entre estes, para realizar segmentação de clientes em grupos? De que forma essa segmentação pode auxiliar em estratégias de marketing e vendas?

1.3 DELIMITAÇÃO DE PESQUISA

O trabalho baseou-se em um projeto de uma empresa multinacional de grande porte do setor de linha branca localizada em São Paulo, SP. Este projeto teve como objetivo realizar a segmentação em grupos, dos clientes, que a empresa possui no Brasil, a fim de auxiliar nas tomadas de decisões.

Utilizando o projeto em questão como referência de objetivos e para posterior comparação com a pesquisa realizada neste trabalho, foi aplicado apenas um método para auxiliar na definição dos critérios para a segmentação, o método *Recency-Frequency-Monetary* (RFM), que é utilizado para analisar o valor de um cliente.

Além disso, foi utilizado um método de análise de decisão multicritério (MCDM), entre os diversos existentes utilizou-se o *Analytic Hierarchy Process* (AHP), para estabelecer relação entre os critérios analisados.

Os dados referentes aos critérios foram extraídos de bases de dados disponibilizados pela empresa em questão, a qual foi selecionada pela facilidade de localização, facilidade de acesso aos dados e exequibilidade do projeto, que facilitou o desenvolvimento deste e

possibilitou maior criticidade ao comparar os resultados deste trabalho com o projeto que a empresa realizou para segmentar os clientes.

Quanto a segmentação de clientes em *clusters* foi utilizado apenas um método de mineração de dados (*Data Mining*) o *K-Means*, realizado no XLSTAT, um *plugin* com aplicações estatísticas e análise de dados para o *Microsoft Office Excel*[®] desenvolvido em 1993.

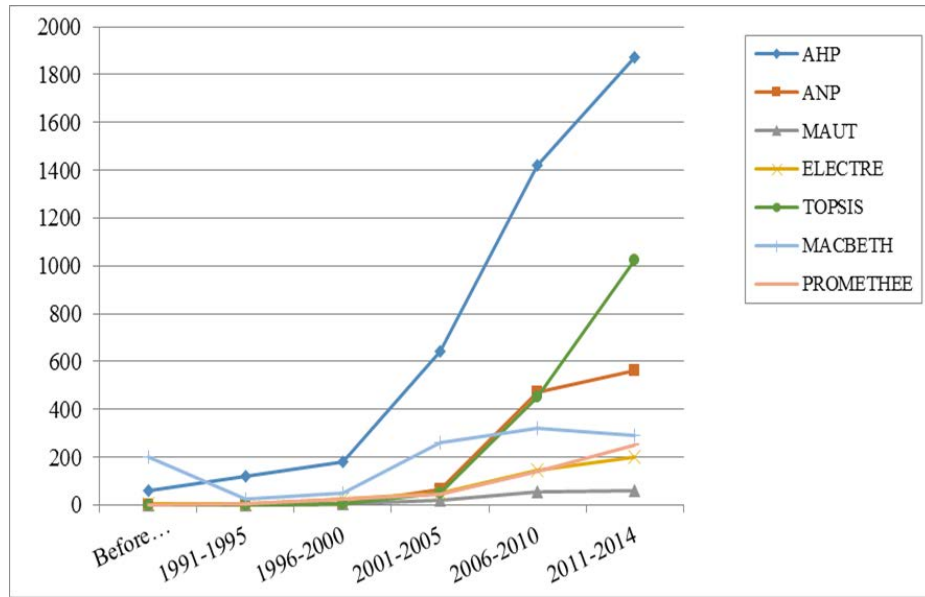
1.4 JUSTIFICATIVA DA PESQUISA

Saber de que maneira realizar tomadas de decisões baseadas em estratégias de marketing são de enorme complexidade para todas as empresas em qualquer contexto. Diante deste fato, a segmentação de clientes auxilia as companhias a ter um maior conhecimento sobre os seus clientes e facilita a visualização para tomadas de decisões gerenciais para alavancar os negócios de uma organização. Para auxiliar em quais critérios deverão ser utilizados, foi aplicado um dos métodos mais conhecido para análise de valor de clientes, o RFM.

Segundo Khajvand et al. (2011), o método RFM é um dos mais poderosos e simples para implementação de gestão de relacionamentos com clientes (*Customer Relationship Management*). E por este motivo, foi o método selecionado para o trabalho.

O método AHP foi escolhido, dentre os diversos disponíveis entre os métodos de decisão de multicritério, uma vez que, segundo Tramarico et al. (2015) este método foi o mais utilizado no período de 1991 até 2014 conforme estudo bibliométrico realizado com o objetivo para apontar quais métodos de tomada de decisão multicritério foram mais utilizados neste período. Na Figura 1 pode-se observar que a partir de 1995, o método *Analytic Hierarchy Process* foi o mais utilizado entre os métodos de tomada de decisão multicritério em publicações no periódico da *Web of Science*.

Figura 1 – Pesquisa bibliométrica sobre métodos de decisão multicritério



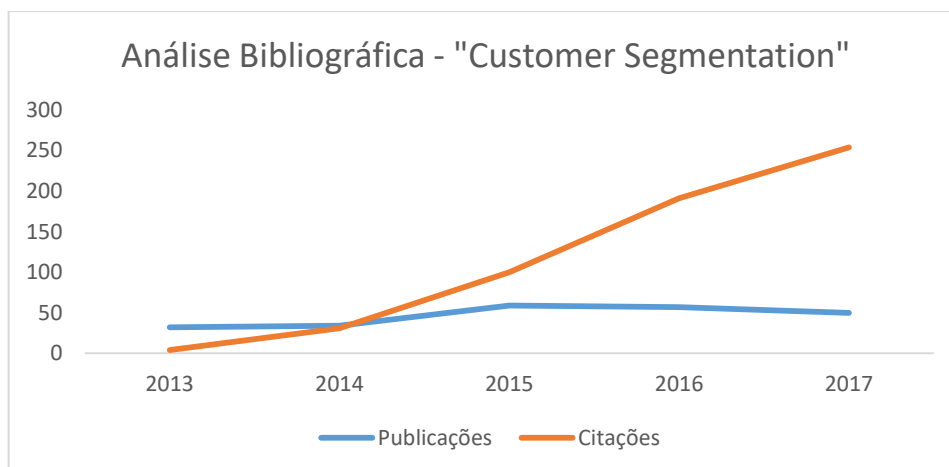
Fonte: Tramarico et al. (2015).

A segmentação de dados em grupos (*Clustering*) é muito importante para adquirir conhecimentos dentro destes dados, identificando recursos importantes, detectando anomalias, etc. Dentre os métodos de clusterização, o algoritmo de *K-Means* é sem dúvidas o mais utilizado devido à sua facilidade de implementação (CELEBI; KINGRAVI; VELA, 2013).

Além dos fatos apresentados para justificar os métodos utilizados para realizar este trabalho, como justificativa adicional, ao realizar pesquisa bibliométrica (observar Figuras 2, 3 e 4) em relação aos termos *Customer Segmentation* e *AHP* em periódicos como o *Web of Science*, foi encontrado um número elevado e constante de pesquisas nos últimos anos e o número de citações a cerca destes está crescendo. Todavia, existem poucas pesquisas quando estas palavras-chave são pesquisadas em conjunto.

Na Figura 2 é apresentado a relação dos últimos cinco anos de pesquisa, de 2012 até 2017, realizado com o tópico de *Customer Segmentation* na *Web of Science*. A pesquisa foi realizada no dia 17/09/2018.

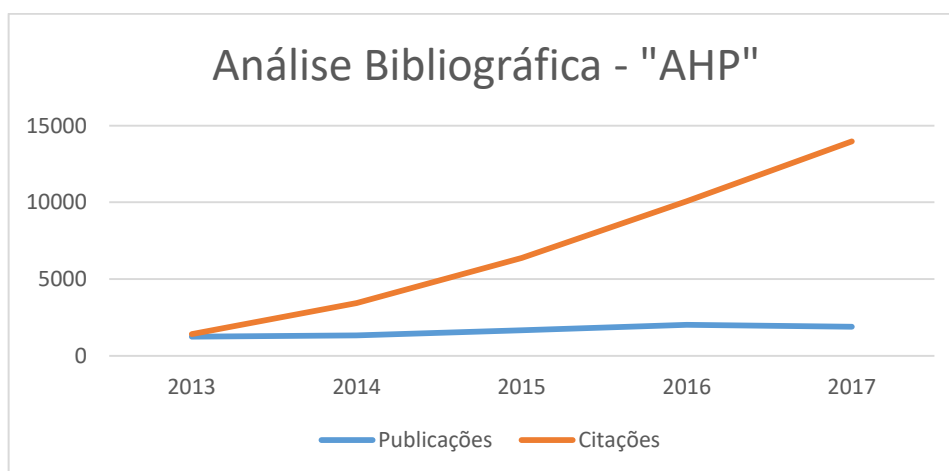
Figura 2 – Pesquisa bibliométrica sobre o tópico *Customer Segmentation*



Fonte: Adaptado de *Web of Science* (2018).

A Figura 3 refere-se ao mesmos parâmetros da busca anterior, porém para a palavra-chave AHP.

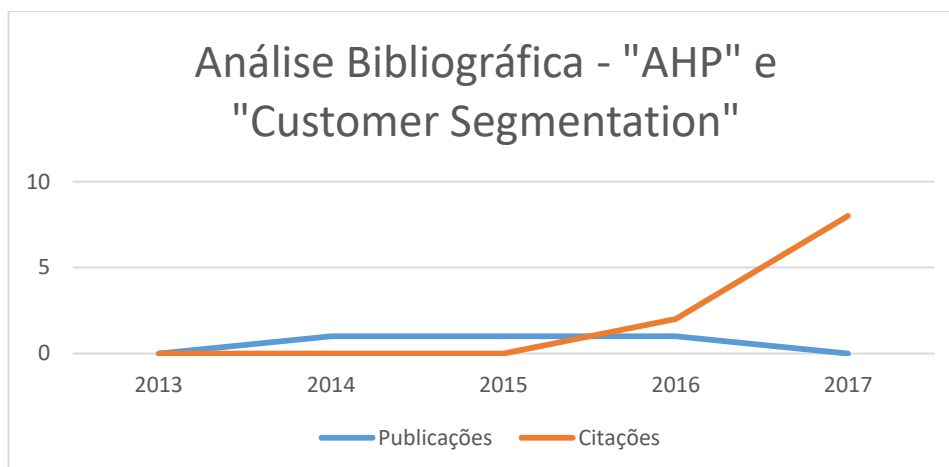
Figura 3 – Pesquisa bibliométrica sobre o tópico AHP



Fonte: Adaptado de *Web of Science* (2018).

Por sua vez, a Figura 4 representa o número de publicações e citações por ano ao utilizarmos os dois tópicos previamente analisados, evidenciando o baixo número de ambos os índices para os temas em conjunto no mesmo período dos últimos cinco anos.

Figura 4 – Pesquisa bibliométrica sobre o tópico *AHP* e *Customer Segmentation*



Fonte: Adaptado de *Web of Science* (2018).

Mediante estes contextos, este trabalho procura contribuir, abordando de forma simultânea os dois tópicos em referência, além de incluir outros métodos e assuntos relevantes para contribuição acadêmica. Em adição, foi demonstrado o desenvolvimento prático deste temas em um ambiente desafiador de uma das empresas líder do mercado mundial de linha branca.

1.5 ESTRUTURA DO TRABALHO

O trabalho está dividido em mais 4 capítulos. O Capítulo 2 contém a revisão sistemática da literatura, onde são apresentados o referencial teórico e os conceitos embasadores que sustentam o trabalho. São apresentados conceitos de *Customer Relationship Management*, método RFM, métodos de decisão multicritério, método AHP, mineração de dados e método *K-Means*.

No Capítulo 3 é descrito o método de pesquisa e como essa está classificada. Já no Capítulo 4 é apresentado o desenvolvimento do trabalho e a aplicação prática do conteúdo desenvolvido na empresa que contribuiu para o estudo deste trabalho e os resultados obtidos da segmentação de clientes. Por fim, o Capítulo 5 traz as considerações finais acerca do tema considerando os objetivos e questão de pesquisa.

2 REFERENCIAL TEÓRICO

2.1 GESTÃO DE RELACIONAMENTO COM O CLIENTE

As empresas reconhecem que os seus clientes são seu ativo mais importante e que diferem de um aos outros sobre questões de comportamento de compra, geografia, preferências, lucratividade, fidelidade e outros quesitos. Além disso, sabe-se que para gerar valor e obter um alto nível de satisfação aos clientes só é possível ao adquirir conhecimentos destes para iniciar e manter relacionamentos, além de melhorar a qualidade de produtos e serviços (TSENG; WU, 2014).

As organizações estão constantemente buscando maneiras de obter informações de qualidade para auxílio em tomadas de decisões de *marketing* importantes, além de garantir um melhor relacionamento com os clientes, ampliando as chances de sucesso nos negócios. Como resposta, a gestão de relacionamento com o cliente, ou *Customer Relationship Management* (CRM), no decorrer das últimas décadas tem sido uma ferramenta importante para auxiliar a aumentar a lucratividade de uma empresa ao possibilitar a identificação dos melhores clientes e satisfazer os desejos destes, tornando-os mais leais às atividades da empresa (NGUYEN; MUTUM, 2012; ŠEBJAN; BOBEK; TOMINC, 2016).

O CRM não é considerado um produto ou um serviço, e sim uma estratégia geral de negócios que possibilita as empresas a gerenciar suas relações com os clientes e prover uma visão integrada dos clientes para toda a companhia (SHAON; RAHMAN, 2015).

Além disso, de acordo com Nguyen e Mutum (2012), a história do *marketing* nos sugere que ocorreu mudanças de paradigmas nas últimas décadas. Antigamente, as buscas das companhias pela liderança de mercado eram determinadas pela eficiência da produção, onde a meta era reduzir os custos operacionais por produção, acarretando em produtos e serviços a um preço menor. Contudo, os concorrentes foram imitando as mesmas técnicas em um curto período de tempo, fazendo com que esta estratégia não fosse sustentável. Atualmente, as organizações possuem a noção de que uma estratégia mais sustentável obtida ao criar relações com os clientes baseado em qualidade, diálogo, inovação e aprendizado, o que gera uma vantagem competitiva de maior tempo.

Segundo (MOHAMMADHOSSEIN; ZAKARIA, 2012; SHAON; RAHMAN, 2015) os benefícios da gestão de relacionamento com os clientes são:

- Aumentar retenção e lealdade de clientes;
- Atingir expectativas dos clientes;

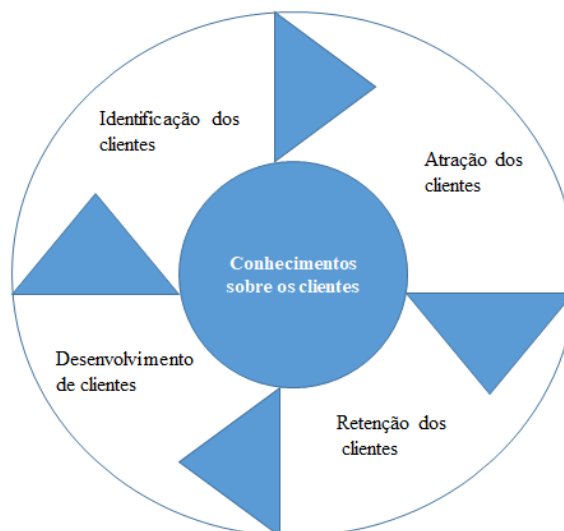
- Capacidade aprimorada de segmentar clientes lucrativos;
- Assistência integrada através de canais;
- Maior eficiência e eficácia da força de vendas;
- Melhor preço;
- Produtos e serviços personalizados;
- Melhoria da eficiência e eficácia do atendimento ao cliente;
- Mensagens de *marketing* individualizadas.

Conforme Kracklauer, Mills e Seifert (2004) e Ngai, Xiu e Chau (2009), o CRM consiste em quatro dimensões: identificação de clientes, atração de clientes, retenção de clientes e desenvolvimento de clientes. Estas dimensões podem ser caracterizadas como um ciclo fechado de um sistema de gestão de clientes e são essenciais para obter maior compreensão dos clientes.

Ainda de acordo com os mesmo autores, o CRM começa com a identificação do cliente, fase em que envolve a segmentação de uma dada população que tem potencial para se tornar clientes ou mais lucrativos para a companhia, além de clientes que estão sendo perdidos por competição e de que forma estes podem ser recuperados. Depois de reconhecido os clientes, as companhias podem direcionar seus esforços para atrair uma segmentação de clientes alvo. Após atrair o cliente, é necessário entender como satisfazê-lo para que este se torne leal. Por último, o desenvolvimento de clientes envolve uma expansão consistente de intensidade de transações e valores de vendas, garantindo uma maior lucratividade individual de cada cliente.

A Figura 5 ilustra o ciclo fechado conforme as quatro dimensões do CRM.

Figura 5 – Dimensões do ciclo de gestão de CRM



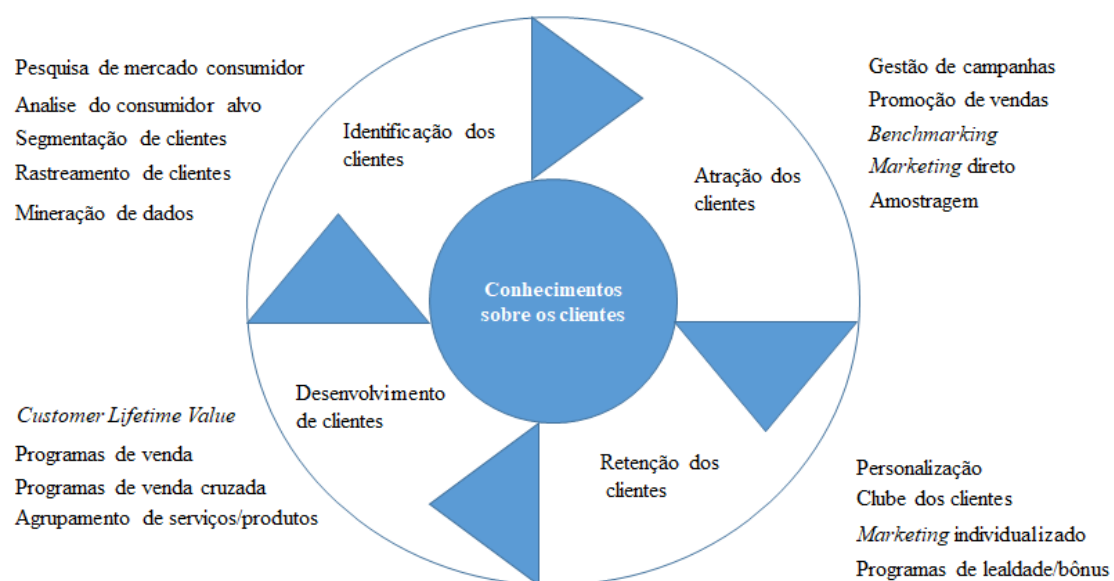
Fonte: Adaptado de Kracklauer; Mills e Seifert (2004).

Os sistemas de CRM podem ser distinguidos em três tipos de categorias: operacional, analítico e colaborativo. O CRM operacional está relacionado a melhorias na eficiência e eficácia dos processos de gestão de cliente, automatizando estes e personalizando a relação com os clientes. Automação de *marketing*, automação de vendas e automação de serviços estão relacionadas a esta classificação de CRM. Por sua vez, o CRM analítico estrutura dados de clientes em informações a respeito destes, identificando comportamentos, necessidades e permitindo análises de negócios, relatórios operacionais e previsões em relação à comportamento de compra e reconhecimento de padrões de compra. Essa categoria incorpora diversas ferramentas analíticas, como a mineração de dados. O CRM colaborativo tem por objetivo otimizar a comunicação multicanal da empresa com os clientes, fornecedores e parceiros (KHODAKARAMI; CHAN, 2014; OLSZAK; BARTUŚ, 2013; PEPPERS; ROGERS, 2016).

As aplicações do CRM analítico incluem a análise de segmentação de clientes, análise de lucratividade de clientes, gestão de campanhas, promoção de vendas, *benchmarking*, comunicação personalizada com clientes, valor da vida útil do cliente (*Customer Lifetime Value* ou CLV), programas de lealdade e bônus para os clientes, dentre outros (KRACKLAUER; MILLS; SEIFERT, 2004).

A Figura 6 expõe as possíveis aplicações de CRM analítico em conjunto com as quatro dimensões do ciclo de gestão do CRM.

Figura 6 – Aplicações CRM analítico

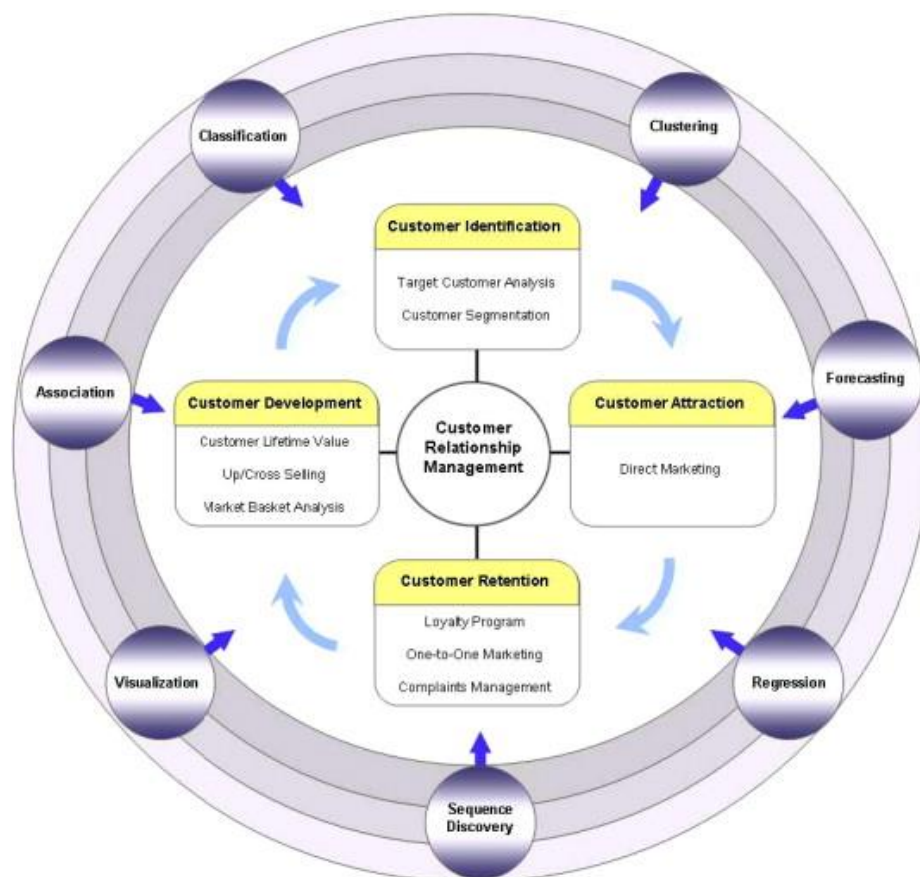


Fonte: Adaptado de Kracklauer; Mills e Seifert (2004).

A mineração de dados desempenha um papel crítico nas aplicações de CRM analítico. Existem várias técnicas de mineração de dados que são utilizadas em aplicações de CRM, tais como árvores de decisão, redes neurais, algoritmos genéricos, aglomeração (*clustering*), árvores de classificação e regressão, regressão logística e regras de associação (NGAI; XIU; CHAU, 2009).

A Figura 7 apresenta a estrutura de classificação que descreve a relação entre as técnicas de mineração de dados e o CRM analítico.

Figura 7 – Classificação de técnicas de mineração de dados no ciclo de gestão de CRM



Fonte: Ngai; Xiu; Chau (2009).

Como observado na Figura 7, a segmentação de clientes é uma das principais funções do CRM analítico e tem como definição o agrupamento de clientes em pequenos grupos ou segmentos mediante características, critérios e valores em comum. A segmentação de clientes permite com que as empresas tenham um maior conhecimento sobre seus clientes, permitindo-as a saber quais são os mais rentáveis, além de projetar e implementar campanhas e programas de *marketing* direcionadas e estratégias de preço para cada segmento de clientes (FLOH et al., 2014; NGAI; XIU; CHAU, 2009).

O objetivo da segmentação de clientes é, segundo Bajaj (2017), reduzir os riscos na decisão de onde, quando, como e para quem um produto ou serviço deve ser comercializado, ademais aumentar a eficiência e eficácia das iniciativas de *marketing*, identificando as metas estratégicas de *marketing* e dividindo os mercados em grupos de consumidores que claramente possuem as mesmas necessidades. Essa abordagem estratégica garante que os esforços de uma organização obtenham um desempenho apropriado ao mesmo tempo em que diminui os custos promocionais com os clientes que não são lucrativos para a companhia.

Uma base de dados com qualidade tem papel fundamental e impacta diretamente no desempenho de uma organização ao desenvolver e implementar estratégias de *marketing* com o uso de dados dos clientes e produtos providos destas bases. Desta forma, é imprescindível dados dos clientes com qualidade para realizar a segmentação destes (ALSHAWI; MISSI; IRANI, 2011; PELTIER; ZAHAY; LEHMANN, 2013).

2.2 MINERAÇÃO DE DADOS (*Data Mining*)

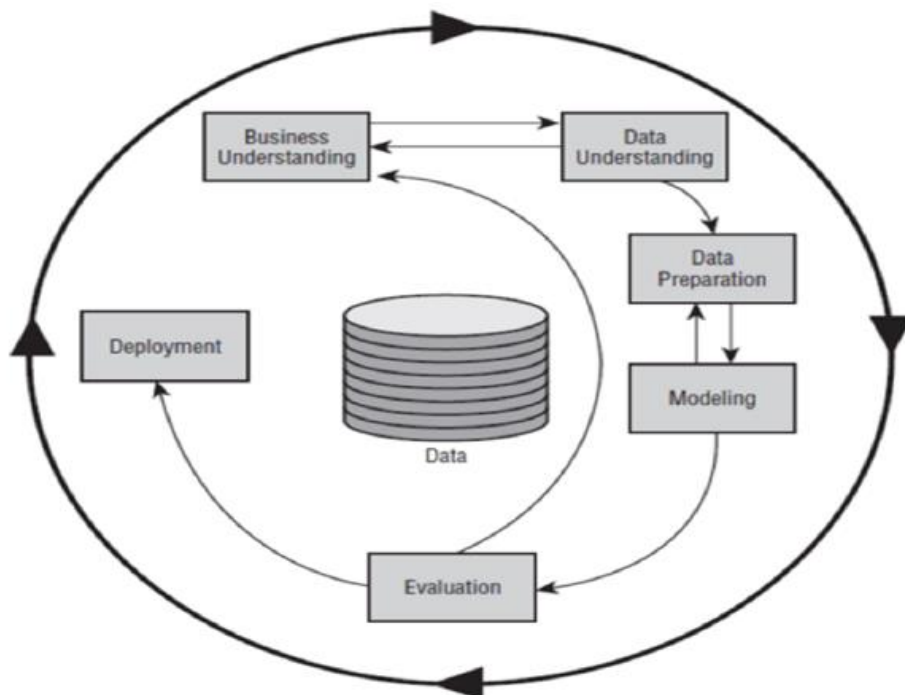
A mineração de dados é definida como o processo de descoberta de padrões e tendências em dados, sendo este processo automático ou semiautomático, levando a descobertas significativas e conhecimentos para obter vantagem competitiva (JIAWEI et al., 2011; WITTEN et al., 2016).

Ainda segundo Jiawei et al. (2011), o processo de descoberta de conhecimento por dados segue o seguinte processo:

- Limpeza de dados: nesta etapa são removidos dados inconsistentes e dados que podem danificar os resultados;
- Integração de dados: é realizada a combinação de diversas fontes de dados;
- Seleção de dados: os dados relevantes para o que se deseja analisar são separados daqueles que não serão pertinentes à análise;
- Transformação de dados: nesta etapa os dados são organizados e consolidados de maneira apropriada para a mineração de dados;
- Mineração de dados: são aplicados métodos para a extração de padrões;
- Avaliação de padrões: os padrões extraídos pela fase de mineração de dados são identificados baseados nos objetivos e interesses da pesquisa;
- Apresentação de conhecimento: etapa onde técnicas de representação visual são utilizados para facilitar a transmissão de conhecimento para os interessados.

Além disso, de acordo com Shafique e Qaiser (2014) o *Cross Industry Standard Process for Data Mining* (CRISP-DM) foi desenvolvido por Daimler Chrysler em 1996, e consiste na forma mais comum de classificar as etapas de descoberta de padrões em análise de dados. O CRISP considera o processo de mineração de dados como uma estratégia geral para solução de problemas e define que um projeto de mineração de dados tem um ciclo de vida que consiste em seis fases, conforme ilustrado na Figura 8.

Figura 8 – Ciclo de vida e etapas do CRISP-DM



Fonte: Shafique e Qaiser (2014).

A primeira fase do CRISP-DM consiste do entendimento do negócio (*business understanding*) onde se busca compreender o negócio onde uma companhia está inserida e o atual contexto econômico do país e setor em que esta se encontra. É uma fase focada na descoberta de fatores importantes, definido critérios de sucesso e os objetivos de negócios (SHAFIQUE; QAISER, 2014).

A fase seguinte, é a compreensão de dados (*data understanding*), onde foco se concentra na coleta de dados, verificado a qualidade, além de explorar estes para obter conhecimento de dados e formar hipóteses para informações ocultas (SHAFIQUE; QAISER, 2014).

A preparação de dados (*data preparation*) é a terceira fase do ciclo do CRISP-DM, que consiste em todas atividades relacionadas para construção final dos dados, ou seja incluem etapas de seleção, limpeza dados, construção, integração e formatação de dados.

Modelagem (*Modelling*) é a fase após a preparação de dados, na qual são aplicadas várias técnicas de modelagem, tendo diferentes parâmetros e modelos aplicados para o mesmo problema de mineração de dados.

Avaliação (*Evaluation*) e desenvolvimento (*Deployment*) são as duas últimas fases e representam as fases de análise dos resultados obtidos pela mineração de dados e os próximos passos são determinados, informando possíveis ações e auxiliando em tomadas de decisões de maneira clara e eficiente.

As técnicas de mineração de dados podem ser divididas em classificação (*classification*), agrupamento (*clustering*), regras de associação (*association rules*) e regressão (*regression*) (PATEL; PATEL, 2014). Além disso, segundo Yoo et al. (2012), de modo geral os algoritmos de mineração de dados podem ser classificados em duas categorias: descritiva que também é conhecida na literatura como aprendizado não supervisionado e preditiva, ou aprendizado supervisionado.

A mineração de dados descritiva agrupa dados e mede a similaridade entre objetos e descobre padrões desconhecidos e relações entre os dados de uma grande base de dados, possibilitando o entendimento destes. Esse tipo de mineração de dados inclui o agrupamento e regras de associação. Por outro lado, a mineração de dados preditiva presume as regras de previsão a partir de dados e aplica estas regras para dados que ainda não foram previstos ou não classificados. A classificação e a regressão fazem parte deste tipo de mineração de dados (YOO et al., 2012).

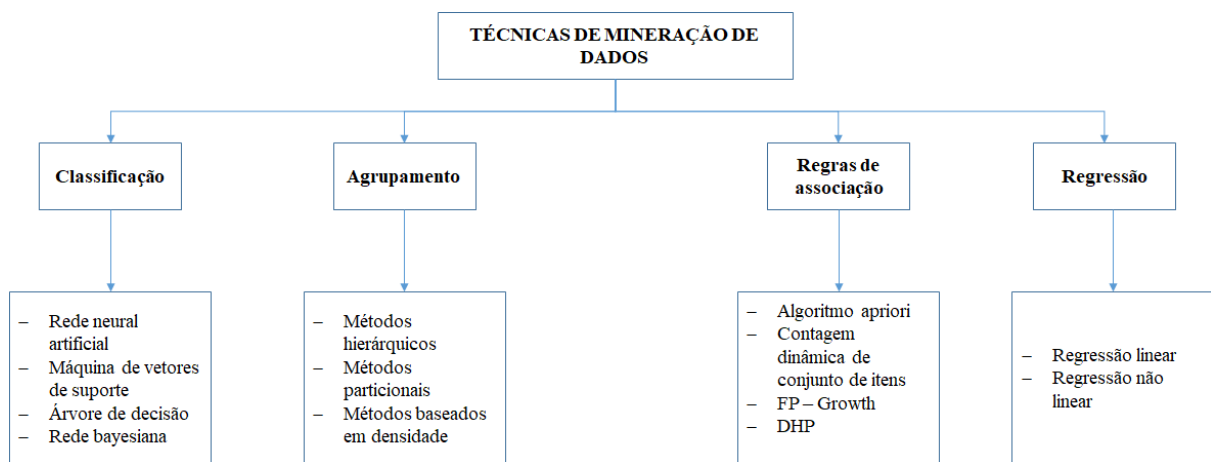
A classificação se baseia e utiliza um modelo para prever as categorias de objetos desconhecidos para distinguir entre objetos de classes diferentes, sendo estas categorias predefinidas, discretas e não ordenadas. A rede neural artificial, a máquina de vetores de suporte e árvore de decisão são técnicas de mineração de dados do tipo classificação (NGAI et al., 2011; PATEL; PATEL, 2014; YOO et al., 2012).

Consoante Ngai et al. (2011), o agrupamento (*clustering*) é utilizado para dividir objetos em grupos significativos (*clusters*), com os objetos de um grupo sendo semelhantes entre si, mas muito diferentes dos objetos de outros grupos. Além disso, o *clustering* também é conhecido como segmentação de dados e é considerado uma vertente de classificação não supervisionada, e tem como técnicas de mineração de dados, modelos hierárquicos, modelos particionais e modelos baseados em densidade (PATEL; PATEL, 2014).

As regras de associação são utilizadas para descobrir elementos que ocorrem simultaneamente de forma repetitiva em um conjunto de dados que consiste em diversas seleções independentes de elementos e descoberta de regras. Já a regressão é uma metodologia estatística usada para revelar a relação de uma ou mais variáveis independentes e uma variável dependente (NGAI et al., 2011; PATEL; PATEL, 2014).

A Figura 9 apresenta as principais técnicas de mineração de dados e os principais métodos que são utilizados nestas categorias.

Figura 9 – Técnicas de mineração de dados

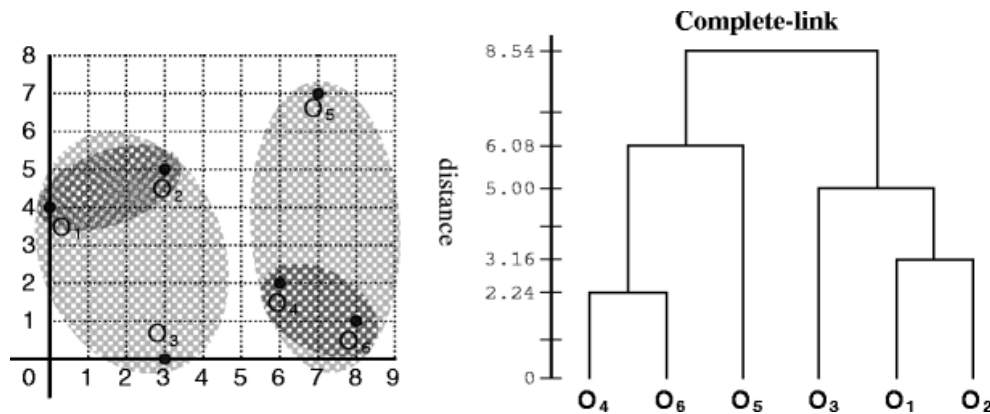


Fonte: Adaptado de Patel; Patel (2014).

Os métodos particionais de *clustering* se caracterizam por utilizar métricas baseadas na distância para agrupar os pontos baseados na sua similaridade, produzindo apenas um nível de particionamento e aglomerados de forma esférica que não se sobrepõem. O algoritmo deste método mais utilizado na literatura é o *K-Means*. Por outro lado, os métodos hierárquicos agrupam os dados mediante diferentes níveis que compõem uma hierarquia, ajudando na visualização de dados e sumarização, e tem como algoritmo mais comum o BIRCH. Diferentemente dos outros dois métodos, os métodos baseados em densidade podem captar arbitrariamente aglomerados com formas, como o em *S-Shape*. Pontos em regiões mais densas são agrupados e os que não estão em regiões tão densas são separados por conta da baixa densidade. O algoritmo mais ordinário é o DBSCAN (AGGARWAL; REDDY, 2013).

A Figura 10 ilustra um exemplo de *clustering* do método hierárquico onde é possível verificar o agrupamento pelo distanciamento de pontos.

Figura 10 – Exemplo de agrupamento



Fonte: Yoo et al. (2012).

O algoritmo *K-Means* tem sua popularidade atribuída a simplicidade e facilidade de implementar, sendo que quase todos os *softwares* de mineração de dados possuem a implementação deste algoritmo e de modo geral os aspectos, como a inicialização, a função de distância, dentre outros, deste algoritmo podem ser modificados. Entretanto, o *K-Means* apenas consegue detectar aglomerados compactos e hiperesféricos que estão bem separados (CELEBI; KINGRAVI; VELA, 2013). O conceito do algoritmo é utilizar K como um parâmetro, dividir um número X de objetos dentro de um número K de *clusters* para criar semelhança relativamente alta ou baixa dentre estes (BASWADE; NALWADE, 2013).

De acordo com Baswade e Nalwade (2013), as etapas do método *K-Means* são as seguintes:

- Inicialmente tem como entrada o número K de grupos que se deseja e o conjunto de dados;
- Definir qualquer objeto X como centroide;
- Encontrar a distância de todos os objetos daqueles centroides, menos a distância que está no centro dos centroides;
- Obter os centroides que estão nestes grupos;
- Repetir as últimas duas etapas até que o valor dos centroides sejam o mesmo;
- Dar o resultado dos K grupos e seus respectivos centroides.

Um centroide é conceitualmente o ponto central de um *cluster*, e pode ser definido de diversas maneiras, como a média dos objetos atribuídos ao *cluster* (JIAWEI et al., 2011). Consoante Wang e Su (2011), o critério de erro quadrado para agrupamento é $E = \sum_{i=1}^K \sum_{j=1}^{n_i} ||x_{i,j} - m_i||^2$, onde $x_{i,j}$ é objeto j da classe i, m_i o centro da classe i e n_i é o número de objetos da classe i. Desta forma o algoritmo segue as seguintes etapas:

- Definir os objetos a serem agrupados $\{x_1, x_2, \dots, x_n\}$ e o número de *clusters* k ;
- Arbitrariamente selecionar k objetos como centros de *clusters* $(m_1, m_2, \dots, m_k)$;
- Calcular a distância entre cada objeto x_i e cada centro dos *clusters*, a seguir atribuir cada objeto para o *cluster* mais próximo. A expressão para calcular a distância d é a seguinte: $d = \sqrt{\sum_{j=1}^d (X_{i1} - m_{j1})^2}$, $i = 1, \dots, n$; $j = 1, \dots, k$; $d(x_i, m_j)$ é a distância entre o dado i e o *cluster* j ;
- Calcular a média dos objetos em cada *cluster* conforme os novos centros dos *clusters*, $m_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}$, $i = 1, 2, \dots, k$; N_i é o número atual de amostras do *cluster* i ;
- Repetir os últimos dois passos até a função E convergir e retornar $(m_1, m_2, \dots, m_k)$.

Os algoritmos de agrupamento podem ser difíceis de avaliar sem informações prévias ou suposições sobre os dados, para suprir este fato existem alguns índices que tem por objetivo mensurar a qualidade de um dado agrupamento. O método mais conhecido na literatura é o de largura da silhueta (*Silhouette Width*). Cada *cluster* é representado por uma silhueta que é baseada na comparação de sua rigidez e separação. Esta silhueta demonstra quais objetos foram colocados de forma correta nos *clusters* e quais estão meramente entre os estes, sendo que a largura média da silhueta representa a relação ideal onde se maximiza a distância entre grupos e minimiza a distância entre os conjuntos. Ela é calculada pela seguinte expressão: $s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$, onde i é cada *cluster*, $s(i)$ é a silhueta média de cada *cluster*, $a(i)$ é a distância média entre i e todas outras entidades no qual o *cluster* i pertence e $b(i)$ é o valor mínimo da média das distâncias entre i e todas as outras entidades em cada outro *cluster* (BURNEY; TARIQ, 2014; KODINARIYA; MAKWANA, 2013).

A Tabela 1 apresenta a interpretação dos resultados de silhueta média onde observações com um valor maior de $s(i)$ representam um melhor agrupamento e valores mais próximos a 0 representam que a observação não é válida entre os *clusters*.

Tabela 1 – Interpretação de resultados da silhueta média

Silhueta Média	Interpretação
0,71 - 1,00	Separação Excelente
0,51 - 0,70	Separação Rasoável
0,26 - 0,50	Separação Fraca
$\leq 0,25$	Separação Ruim

Fonte: Adaptado de Burney; Tariq (2014).

Técnicas de suporte e modelos para auxílio a tomadas de decisões de *marketing* são muito importantes desde a criação de sistemas de apoio a decisão e muitos domínios do *marketing* como preço, desenvolvimento de novos produtos, publicidade e segmentação de clientes se beneficiaram dessas técnicas e métodos. A mineração de dados é uma ferramenta valiosa nesse contexto e possui diversas técnicas incluindo análises estatísticas e algoritmos de aprendizado de máquina. Dentre estas técnicas a análise RFM tem sido utilizada há muito tempo em decisões de *marketing* e é reconhecida como uma técnica de mineração de dados útil para a segmentação de clientes e modelos de resposta, estando entre as técnicas mais utilizadas para segmentação e modelagem produtivas pelos profissionais do ramo (OLSON; CHAE, 2012).

De acordo com Cho et al. (2012), RFM são siglas para *recency*, *frequency*, e *monetary*. A recência significa “como recentemente um cliente tem comprado?”, a frequência “com qual frequência ele compra?” e monetário “quanto ele gasta nas compras?”. Muitos estudos comprovaram que estas três variáveis são importantes para prever futuras respostas dos clientes para potenciais esforços de *marketing*. Algumas das limitações da análise RFM foram reconhecidas pela literatura e outras variáveis foram adicionadas ao método de acordo com o contexto em que a pesquisa foi realizada e que teriam impacto relevante no estudo, como, por exemplo, variáveis sócio demográficas e psicográficas (OLSON; CHAE, 2012).

2.3 TOMADA DE DECISÃO MULTICRITÉRIO

A análise de decisão multicritério é uma ferramenta para auxílio em tomadas de decisões complexas envolvendo fatores qualitativos e quantitativos e seus métodos são utilizados para avaliar e tomar decisões dentre as melhores opções disponíveis, a fim de escolher os critérios mais relevantes (AZIZ; SOROOSHIAN; MAHMUD, 2016; MARDANI; JUSOH; ZAVADSKAS, 2015).

Os problemas relativos às tomadas de decisões são de forma geral incertos de várias maneiras. A falta de informação acarreta em um estado pouco claro sobre como será o futuro de um determinado sistema e essa incerteza foi abordada usando teorias de probabilidade e estatística. Porém, em diversas situações reais, para avaliação, julgamento e decisão, a linguagem natural é frequentemente empregada para articular os pensamentos e as percepções subjetivas, não tendo um significado claro e bem definido. Como resultado, se a subjetividade é utilizada para definir um conjunto ao qual os objetivos pertencem ou não pertencem, estes ficarão confusos. Além disso, quando indivíduos estão julgando um evento, mesmo utilizando as mesmas palavras, eles podem divergir significativamente, uma vez que cada um deles possui uma percepção diferente acerca do contexto (MARDANI; JUSOH; ZAVADSKAS, 2015).

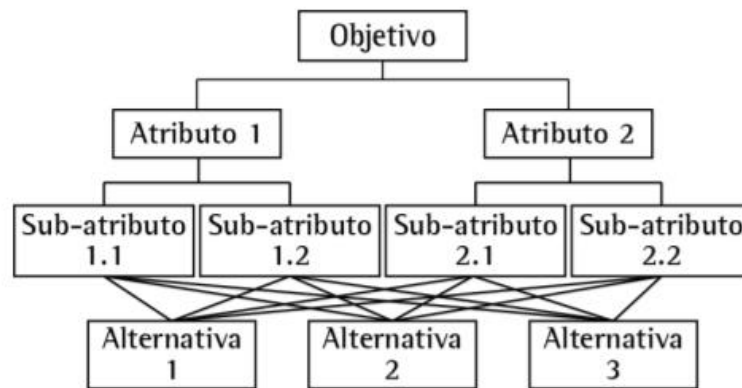
Segundo Aziz, Sorooshian e Mahmud (2016), existem dois tipos de análise de decisão multicritério: métodos de decisão multiobjetivo e métodos de decisão multiatributo. O primeiro consiste em uma abordagem que utiliza técnicas de otimização matemática e envolve principalmente procedimentos de análise relacionados ao processo de *design* de cálculo. Por outro lado, o segundo consistem em uma abordagem desenvolvida para selecionar a melhor opção dentre várias alternativas, o método *Analytical Hierarchy Process* (AHP) é um método de decisão multiatributo.

Ainda consoante aos mesmo autores, o método AHP foi criada na década de 1970 por Saaty e é composto por técnicas que são adequadas para a classificação (*ranking*) de problemas críticos de gerenciamento. Este método é amplamente utilizado na tomada de decisões em grupos em vários setores, como negócios, governo, indústria, educação, saúde e outros. Além disso, o método permite teste de consistência de julgamento, permitindo verificar e reduzir as inconsistências nas opiniões e julgamentos subjetivos.

O AHP tem a vantagem de permitir uma estrutura hierárquica dos critérios, que fornece ao usuário um melhor foco em critérios e subcritérios específicos ao alocar os pesos. Essa abordagem é importante, porque uma estrutura diferente pode levar a uma classificação final diferente. Na implementação dele, uma técnica de comparação pareada é usada para derivar as prioridades para os critérios em termos de sua importância para atingir o objetivo. Portanto, o AHP é baseado em três princípios: decomposição, julgamento comparativo e síntese de prioridades (OUMA; TATEISHI, 2014).

Ao organizar e avaliar as alternativas com relação a uma hierarquia de atributos multifacetados conforme apresentado na Figura 11, o AHP fornece uma ferramenta eficaz para o auxílio em tomadas de decisões de problemas complexos (ALVES; ALVES, 2015).

Figura 11 – Estrutura hierárquica do método AHP



Fonte: Alves; Alves (2015).

Ao utilizar o AHP para modelar um problema, é necessário a estrutura hierárquica ou de rede para representar o problema, assim como comparações pareadas para estabelecer relações dentro da estrutura. A escala fundamental representa os valores e intensidades dos julgamentos, conforme Tabela 2 (SAATY; VARGAS, 2012).

Tabela 2 – Escala fundamental AHP

Intensidade	Definição	Explicação
1	Importância igual	Duas atividades contribuem igualmente para o objetivo
2	Fraco	Experiência e julgamento favorecem levemente uma atividade sobre a outra
3	Importância moderada	
4	Moderado +	Experiência e julgamento favorecem fortemente uma atividade sobre a outra
5	Importância forte	
6	Forte +	Experiência e julgamento favorecem muito fortemente uma atividade sobre a outra
7	Importância muito forte	
8	Muito forte +	A evidência favorecendo uma atividade sobre a outra é a maior possível ordem de afirmação
9	Importância extremamente forte	

Fonte: Adaptado de Saaty; Vargas (2012).

Ainda segundo Saaty e Vargas (2012), a inconsistência é um elemento inerente ao ser humano e portanto deve haver um parâmetro para mensurar a tolerância de aceitação dentre os julgamentos estabelecidos para os critérios. Este parâmetro é o índice de inconsistência (IC) e ele pode ser calculado pela expressão: $IC = |(\lambda_{m\acute{a}x} - n)| / (n - 1)$, onde n representa a ordem da matriz e $\lambda_{m\acute{a}x}$ o maior autovalor da matriz de comparações. Para resultados de $IC \leq 0,1$, as comparações são validas, porém caso este não seja os casos, deve-se identificar o julgamento mais inconsistente da matriz de decisão, determinar um intervalo de valores para

possivelmente alterar deste julgamento inconsistente e solicitar ao tomador de decisão que reconsidere o julgamento para um valor aceitável (OUMA; TATEISHI, 2014).

A razão de consistência (RC), proposto por (SAATY; VARGAS, 2012), permite que o usuário concluir se as avaliações finais são suficientemente consistentes e é obtida pela equação: $RC = IC/IR$, onde IR é o índice randômico obtido de uma matriz recíproca de mesma ordem que a matriz de julgamento. A Tabela 3 apresenta os valores do índice randômico para um determinada matriz de ordem n.

Tabela 3 – Índice randômico por ordem de matriz

N	1	2	3	4	5	6	7	8	9	10
IR	0	0	0,58	0,9	1,12	1,24	1,32	1,41	1,45	1,49

Fonte: Adaptado de Saaty; Vargas (2012).

3 MÉTODO E CLASSIFICAÇÃO DE PESQUISA

A pesquisa realizada neste trabalho possui como característica natureza aplicada, uma vez que explora conhecimentos da literatura e aplica em uma situação específica prática, tendo uma finalidade e objetivos específicos (MORABITO et al., 2018).

Segundo o mesmo autor, em relação à abordagem, este trabalho se classifica como descritivo, pois analisa e correlaciona aspectos, ou variáveis, que envolvem fatos sem alterá-los, buscando apenas descrever o fenômeno que está ocorrendo. Além disso, pelo fato dos fenômenos serem descritos a partir de números, obtidos de diversas fontes de dados, que buscam observar e traduzir uma realidade, esta pesquisa se classifica como qualitativa.

Por se tratar de uma pesquisa onde há participação do pesquisador na situação a ser investigada, disfrutando do conhecimento acadêmico em combinação com aplicações práticas, a pesquisa tem o procedimento de pesquisa-ação.

Na Figura 12 é apresentado uma síntese do método e classificação de pesquisa.

Figura 12 – Método e classificação de pesquisa

Natureza	Objetivos	Abordagem	Procedimento
Básica	Exploratória	Qualitativa	Bibliográfica
Aplicada	Descritiva	Quantitativa	Documental
	Explicativa	Combinada	Estudo de caso
			Expost Facto
			Experimental
			Modelagem e/ou simulação
			Levantamento (<i>Survey</i>)
			Participante
			Pesquisa-ação

Fonte: Próprio Autor (2018)

4 DESENVOLVIMENTO E RESULTADOS

4.1 DESCRIÇÃO DA EMPRESA E DA ÁREA

A empresa em que o projeto foi realizado é uma multinacional americana no setor de eletrodomésticos de linha branca fundada em 1911 e detentora de marcas conhecidas mundialmente. A companhia possui dois públicos-alvo: clientes B2B (*Business to Business*), nos quais a venda ocorre para clientes varejistas e estes vendem posteriormente para o consumidor final, e B2C (*Business to Consumer*), nos quais a venda ocorre de maneira direta para o consumidor final pelos próprios canais de venda online desta empresa.

Dentre as linhas de produtos que constituem o setor de eletrodomésticos de linha branca, esta empresa atua principalmente em seis categorias: Refrigeração (RE), Cocção (FG), *Air Treatment* (AT), Lavanderia (LR), Lava-louças (LL) e Micro-ondas (MWO). Essas categorias podem ser descritas da seguinte forma:

- RE faz referência a eletrodomésticos destinados à refrigeração, assim como geladeiras, freezers verticais e horizontais;
- FG engloba eletrodomésticos destinados à cocção, assim como fogões, fogões embutidos, fornos, fornos embutidos e *cooktops*;
- AT diz respeito a eletrodomésticos destinados ao tratamento do ar, como por exemplo, aparelhos de ar condicionado e climatizadores;
- LR, LL e MWO, se referem à eletrodomésticos relacionados à lavanderia, lava-louças e micro-ondas, respectivamente.

Os clientes varejistas cuja atividade comercial inclui a revenda de produtos que recebem de fabricantes de produtos às vezes recebem concessões ou incentivos destas empresas manufactureiras. Esses incentivos de fornecedor podem ser de vários tipos de crédito, abatimento, desconto de preço ou dinheiro que o fornecedor paga ao varejista que vende seu produto ou serviço. A empresa em questão deste projeto está inserida neste contexto e estes incentivos representam as condicionalidades comerciais que esta possui com seus clientes, seja por acordos, contratos ou ações que se enquadram dentro do seu planejamento comercial.

O setor desta companhia no qual o projeto de pesquisa foi desenvolvido é a área de gestão de rentabilidade, a qual é responsável pela gestão das condicionalidades comerciais dos clientes e monitoramento da competitividade do varejo, visando alavancar o volume e receita desta empresa garantindo a paridade de condições entre os clientes.

O trabalho se desenvolveu em conjunto com a equipe de operação desta área, cujo objetivo é segmentar os clientes em pequenos grupos gerenciáveis, utilizando determinados critérios considerados relevantes para a área, que possam ser facilmente identificados e que auxiliem em tomadas de decisões para alavancar o volume e receita da companhia alvo do projeto. Neste contexto foi proposto para a empresa uma abordagem de segmentação de clientes varejistas (B2B) utilizando o algoritmo de agrupamento *K-Means* e o método AHP para encontrar a importância relativa desses grupos usando esta técnica de tomada de decisões com múltiplos critérios.

4.2 DETERMINAÇÃO DAS CARACTERÍSTICAS DE AVALIAÇÃO DOS CLIENTES

Os clientes são avaliados de acordo com suas diferentes características em relação ao setor da companhia. Os tipos de dados mantidos sobre os clientes variam entre os setores e até mesmo entre empresas. Por exemplo, setores como varejo, telecomunicações e serviços bancários possuem um rico banco de dados sobre os clientes e se preocupam com diferentes características dos clientes de acordo com as suas políticas de CRM.

Por outro lado, setores como os de bens de consumo duráveis, como os de linha branca, que possuem um longo tempo de compra, têm informações limitadas sobre os clientes, fato que traduz que suas avaliações sejam diferentes de outros setores e torna tanto a compreensão de negócio tanto a análise do setor vitais para a segmentação do mercado e de clientes.

Para determinar as características para avaliar e segmentar os clientes neste projeto foi realizada revisão abrangente da literatura e além disso, a técnica de *brainstorming* foi aplicada para um conjunto de especialistas da área de gestão de rentabilidade e de planejamento comercial da companhia. A questão específica que o *brainstorming* tinha por objetivo responder era a seguinte: “O que você considera como importante para a caracterização de um cliente?”.

Os membros do grupo responderam esta questão e, em seguida, usando suas respostas e a pesquisa bibliográfica, as características de avaliação dos clientes foram definidas, caracterizando o método RFM adaptado. As características de *recency* e *frequency*, que representam, respectivamente, a última data que o cliente realizou a compra dentro de um determinado período e com que frequência de compras em um determinado período, não foram consideradas, uma vez que a empresa em questão trabalha com planos e vendas mensais para os clientes e dessa forma são características que não auxiliam no processo de diferenciação entre os clientes e, conseqüentemente, na segmentação destes.

Com isso, as características determinadas foram: margem de contribuição unitária; margem de contribuição unitária percentual; condicionalidade comercial percentual; cobertura de estoque e potencial de mercado.

A margem de contribuição unitária (MCU) representa o montante restante de receita após dedução de custos fixos e custos diretos tendo como base toda produção e capacidade produtiva em um determinado intervalo de tempo, representando desta forma a rentabilidade de um cliente para a companhia. Por sua vez, a margem de contribuição percentual (MCUP) considera a relação de MCU sobre a receita líquida que o cliente gerou, desta forma são obtidos índices de rentabilidade com uma maior paridade entre os clientes.

A condicionalidade comercial percentual (CCP) representa o percentual de todo o montante de incentivos que um cliente recebe sobre a receita de produtos gerada por este em determinado período, como por exemplo, desconto de preços e verbas comerciais para ações de *sell in*, ou seja, ações para auxiliar o varejista a introduzir e faturar um produto e *sell out*, ou seja, ações que ajudam o cliente a realizar ações promocionais para o cliente final e desta forma diminuir o seu estoque.

O índice de cobertura de estoque (CE) é utilizado para medir o tempo que o estoque, em um determinado período, consegue suportar demandas futuras sem haver necessidade de suprimento. Ou seja, representa o tempo que o produto leva para sair do estoque e é obtido pela divisão do estoque atual pela quantidade média de vendas

O potencial de mercado (PM) é a diferença entre a quantidade de certas categorias de produtos que um cliente comprou no mercado e a quantidade destas mesmas categorias que este cliente comprou de uma companhia, representando assim o volume de produtos que determinado cliente poderia comprar de uma companhia mas está comprando de empresas concorrentes no mercado.

4.3 OBTENÇÃO E PREPARAÇÃO DE DADOS

Neste trabalho, os dados foram extraídos do sistema ERP da companhia. O período a ser analisado foi considerado de julho de 2017 até julho de 2018, pois é necessário dados mais recentes para realizar a segmentação de clientes com o intuito de tomadas de decisões no presente, já que diversos fatores externos podem influenciar em dados antigos, como por exemplo, a situação econômica em que o país estava em anos anteriores, além de taxas de impostos, juros, etc.

Como resultado foram obtidos dados de 142 clientes, onde dentre estes 23 foram julgados como inapropriados para uso e foram ignorados. Os dados obtidos pelo sistema ERP posteriormente foram transferidos para o *Microsoft Office Excel*[®]. Os nomes dos clientes foram ocultados de acordo com a política de privacidade da companhia. Os dados podem ser vistos no Apêndice A.

Para objetivos de agrupamento, os dados dos clientes para todos os critérios foram normalizados entre 0 e 1 conforme a normalização min-max que funciona verificando quanto maior é o valor da célula em relação ao menor valor e escalando essa diferença dentro do intervalo. O objetivo da normalização é criar uma escala comum para as características e torná-las comensuradas. A seguinte equação foi utilizada para a normalização: $n_{ij} = \frac{x_{ij} - X_j^-}{X_j^* - X_j^-}$, onde X_{ij} representa o valor original do campo, n_{ij} o valor normalizado do campo, X_j^* o melhor valor da característica j e X_j^- o pior valor da característica j. Os dados normalizados podem ser vistos no Apêndice B, onde, respectivamente, NMCU, NMCUP, NCCP, NCE e NPM representam a normalização de MCU, MCUP, CCP, CE e PM.

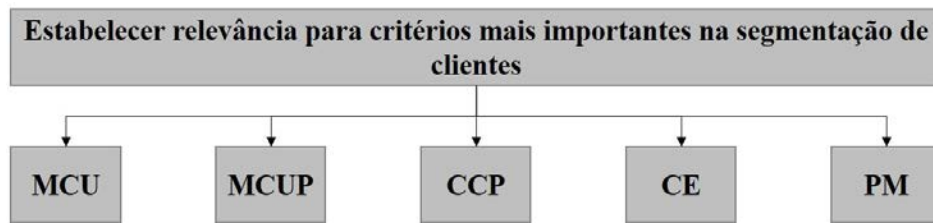
4.4 MODELAGEM MÉTODO AHP

A partir da definição dos critérios a serem utilizados para avaliação e segmentação dos clientes, o método AHP foi utilizado para determinar pesos de importância para cada uma das características, possibilitando a obtenção da relação dos clientes mais importantes mediante estes critérios.

Os julgamentos par a par dos critérios foram obtidos, devido à familiaridade com o assunto, com o apoio do gestor da área de gestão de rentabilidade da empresa do projeto e o software utilizado para análise e consolidação dos dados foi o *Microsoft Office Excel*[®].

As linhas e colunas representam os critérios a serem avaliados e, neste caso, as linhas foram consideradas como prioridades sobre as colunas, ou seja para a linha MCU em comparação com a coluna CCP apresentou-se o valor 5 da escala Saaty, o que significa que a margem de contribuição unitária é fortemente preferida em relação à condicionalidade comercial percentual. A Figura 13 representa a estrutura hierárquica do AHP e a Tabela 4 apresenta os resultados dos julgamentos par a par. Além disso, nesta tabela, a coluna “Auto Vetor” foi obtida pelo cálculo da média aritmética de cada linha, e os resultados da coluna “Normalizado” foram determinados através da divisão do Auto Vetor da linha pela soma total de todos auto vetores.

Figura 13 – Estrutura hierárquica AHP



Fonte: Próprio Autor (2018)

Tabela 4 – Matriz de comparação de critérios

MATRIZ DE CRITÉRIOS	MCU	MCUP	CCP	CE	PM	Auto vetor (AV)	Normalizado
MCU	1	3	5	7	3	3,80	0,4171
MCUP	1/3	1	3	7	3	2,87	0,3147
CCP	1/5	1/3	1	1	1/3	0,57	0,0629
CE	1/7	1/7	1/5	1	1/5	0,34	0,0370
PM	1/3	1/3	1	5	1	1,53	0,1683
Total (Soma)	2,01	4,81	10,20	21,00	7,53	9,11	1,0000

Fonte: Próprio Autor (2018)

Para verificar validação da matriz de comparação dos critérios, os quais são cinco ($n=5$), pela razão de consistência, primeiramente foi realizado o cálculo do máximo autovalor ($\lambda_{máx}$), obtido pela soma produto da linha “Total (Soma)” pelo seu valor correspondente na coluna “Normalizado”, em segunda foi obtido o valor do índice de consistência e por último a razão de consistência por meio das equações obtidas do referencial teórico.

A Tabela 5 mostra os respectivos valores para o cálculo de consistência.

Tabela 5 – Cálculo de consistência

N	5
IR	1,12
$\lambda_{máx}$	5,038442
IC	0,009610
RC	0,008581

Fonte: Próprio Autor (2018)

Conforme pode ser analisado pela Tabela 5, $IC \leq 0,1$ e $RC < 0,1$. Desta forma, pode-se considerar as comparações da matriz de critérios consistentes.

Pelo fato das alternativas representarem todos os clientes obtidos na coleta de dados, ou seja, 119 alternativas e realizar o processo de comparação par a par para cada alternativa se

tornar inviável para este caso, foi utilizado os dados normalizados das características dos clientes.

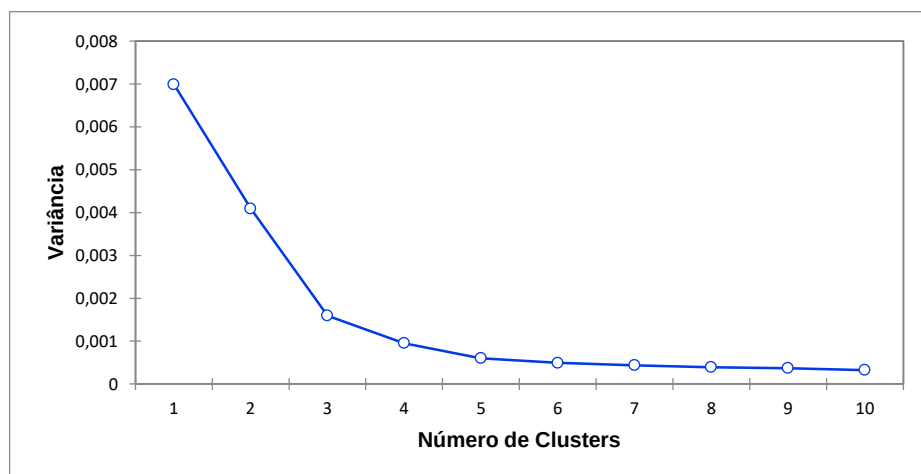
A consolidação dos resultados foi feita a partir do somatório dos produtos dos pesos dos critérios pela normalização de cada critério para cada cliente, obtendo a pontuação geral (PG) de cada cliente, estabelecendo um *ranking* entre estes e os classificando conforme sua importância para a companhia. Os dados referentes à essa classificação podem ser visualizados no Apêndice C e estes foram utilizados como entrada de dados para o algoritmo *K-Means* no *plugin XLSTAT* para Excel.

4.5 APLICAÇÃO MÉTODO K-MEANS E RESULTADOS

Na realização do método de *K-Means* é necessário estabelecer o número *k* de agrupamentos a serem realizados por este. Ao realizar simulação deste algoritmo no XLSTAT foi determinado um número de *clusters* variável de um até dez e para escolher o *k* ideal foi utilizado o método *Elbow* que é um método para interpretação e validação da consistência dentro da análise de um *cluster*, projetado para encontrar o número apropriado de *clusters* de um conjunto de dados onde é examinado a porcentagem de variância explicada em função do número de *clusters* e deve-se escolher um número de *clusters* para que a adição de outro *cluster* não forneça uma modelagem muito melhor dos dados, ou seja, onde o ganho marginal irá diminuir, formando um ângulo no gráfico e desta forma caracterizando um “cotovelo” que será o número ideal de *clusters*.

A Figura 14 ilustra a aplicação do método *Elbow* para obtenção do número ideal de *clusters*.

Figura 14 – Método *Elbow*



Fonte: Próprio Autor (2018)

Ao analisar a Figura 13, o número ideal de *clusters*, segundo o método *Elbow*, é $k = 4$.

Após a determinação do número de *clusters*, foi realizado a aplicação do método *K-Means* com este número de *clusters* para o agrupamento dos clientes.

O algoritmo do método inicia designando k registros iniciais aleatoriamente como centroides de agrupamento, ou seja, centros iniciais, e atribui cada registro ao seu agrupamento mais próximo. À medida que novos registros são adicionados aos *clusters*, os centroides do *cluster* são recalculados para refletir para seus novos membros. Em seguida, os casos são retribuídos para os *clusters* ajustados. Por fim, os centroides dos *clusters* finais são obtidos após este processo iterativo. A Tabela 6 apresenta os centroides de *cluster* iniciais e a Tabela 7 indica os centroides de *cluster* finais e seu respectivo *ranking*.

Tabela 6 – Centroides *cluster* iniciais

Cluster	PG
1	0,200
2	0,220
3	0,209
4	0,226

Fonte: Próprio Autor (2018)

Tabela 7 – Centroides *cluster* finais e *ranking*

Cluster	PG	Ranking
1	0,263	2
2	0,184	3
3	0,095	4
4	0,652	1

Fonte: Próprio Autor (2018)

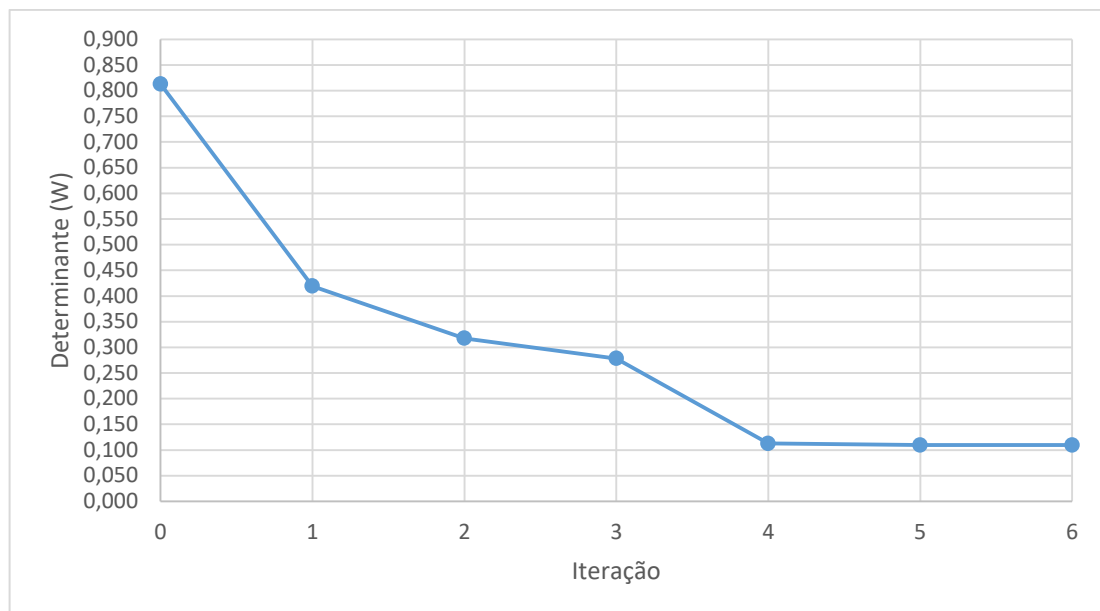
Se considerarmos os centroides do *cluster* final, pode-se afirmar que o *cluster* quatro tem a pontuação geral mais alta (0,625) e por este motivo é o *cluster* mais importante entre os *clusters* e ficando com a primeira posição no *ranking*. De forma análoga, o *cluster* três é o que possui menor pontuação entre os *clusters* (0,095) e obtêm a pior posição no *ranking*.

É sabido que o método *K-Means* é um processo iterativo e o seu processo teve término após a sua quinta iteração, assim como pode ser visto na Figura 15 em conjunto com a Tabela 8 que representa o sumário de iterações do algoritmo até atingir o valor mínimo de seu determinante que foi $W = 0,110$.

Tabela 8 – Sumário de iterações do algoritmo *K-Means*

Iteração	Determinante (W)
0	0,813
1	0,419
2	0,318
3	0,278
4	0,113
5	0,110
6	0,110

Fonte: Próprio Autor (2018)

Figura 15 – Número de iterações do algoritmo *K-Means*

Fonte: Próprio Autor (2018)

A Tabela 9 apresenta os valores de quantidades de clientes em cada *cluster*, a distância mínima, média e a distância máxima para o centroide. No primeiro *cluster* foram assinalados 45 clientes, no segundo 60, no terceiro 12 e no quarto 2. O Apêndice D apresenta a relação de quais clientes foram designados para cada *cluster*.

Tabela 9 – Resultados do algoritmo *K-Means* para os quatro *clusters*

Cluster	1	2	3	4
Quantidade de Clientes	45	60	12	2
Distância mínima para centroide	0,001	0,001	0,008	0,114
Distância média para centroide	0,026	0,018	0,028	0,114
Distância máxima para centroide	0,082	0,043	0,072	0,114

Fonte: Próprio Autor (2018)

Por sua vez, a Tabela 10 indica os valores referentes às distâncias entre os centroides dos *clusters*. A distância máxima ocorre entre o *cluster* três e o *cluster* quatro (0,557), fato que indica que há maior dissimilaridade entre estes dois *clusters*. Esse índice comprova o *ranking* entre os dois grupos, uma vez que o *cluster* quatro é o mais importante e o três o menos importante.

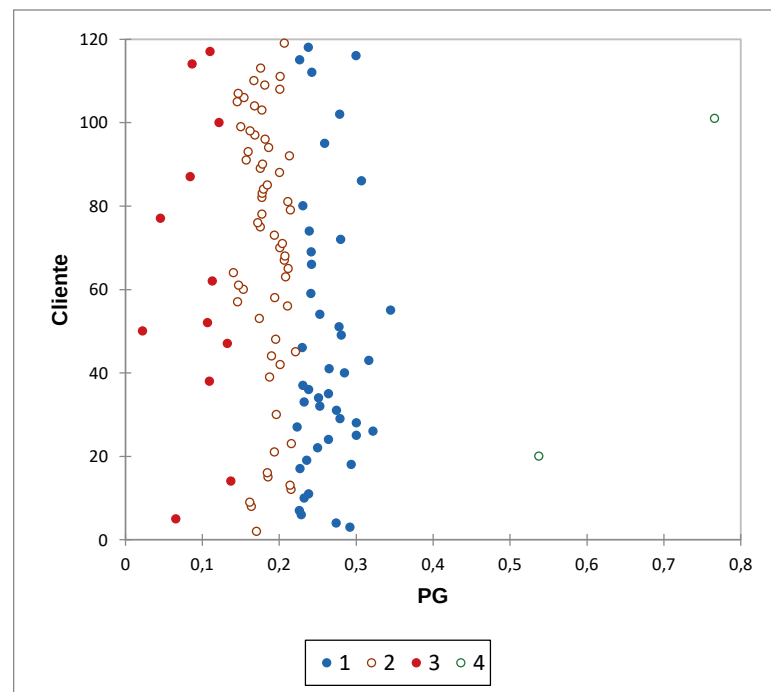
Tabela 10 – Distância entre centroides dos *clusters*

Cluster	1	2	3	4
1	0	0,079	0,168	0,389
2	0,079	0	0,089	0,468
3	0,168	0,089	0	0,557
4	0,389	0,468	0,557	0

Fonte: Próprio Autor (2018)

O gráfico de dispersão do agrupamento dos clientes em *clusters* de acordo com a sua pontuação geral pode ser visualizado na Figura 16.

Figura 16 – Gráfico de dispersão da segmentação de clientes



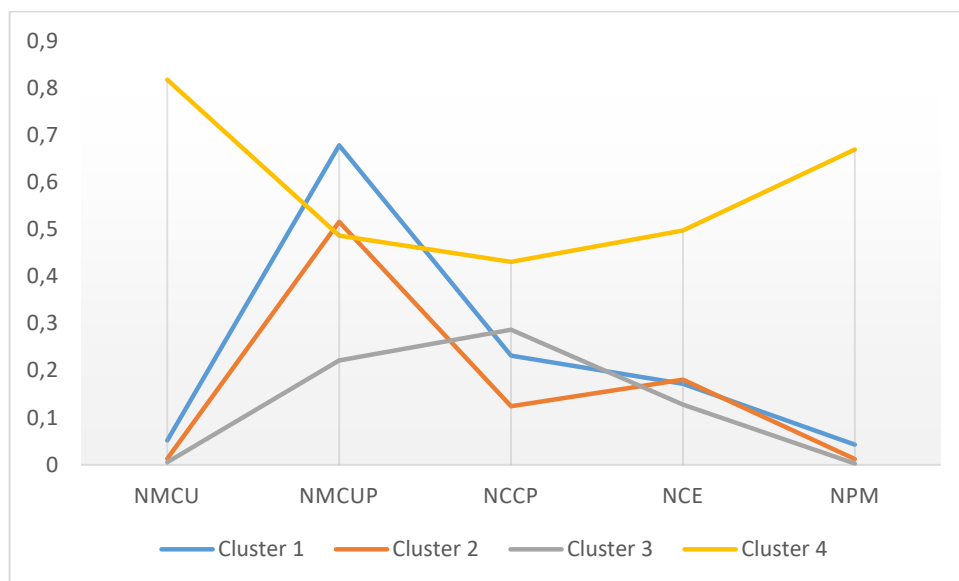
Fonte: Próprio Autor (2018)

Por sua vez, a Figura 17 ilustra a avaliação dos grupos sob cada característica e que possibilita de maneira visual estabelecer direcionais estratégicos de *marketing* e vendas para

cada *cluster*. É possível analisar que existe uma grande lacuna entre o *cluster* quatro e os outros *clusters* em relação às características NMCU, NCCP, NCE e NPM, fato que traduz ser o *cluster* de maior importância e de maior rentabilidade e potencial entre todos os clientes para a companhia, embora apresenta alto índice de condicionalidade comercial e elevada cobertura de estoque. Por sua vez os clientes do *cluster* 1 apresentam maior rentabilidade do que os *clusters* dois e três, além de um potencial de mercado maior e uma condicionalidade comercial menor que o *cluster* três, e um potencial direcional estratégico para a companhia para aumentar sua rentabilidade como um todo seria investir nas condicionalidades comerciais deste grupo para alcançar melhores resultados em termos de rentabilidade.

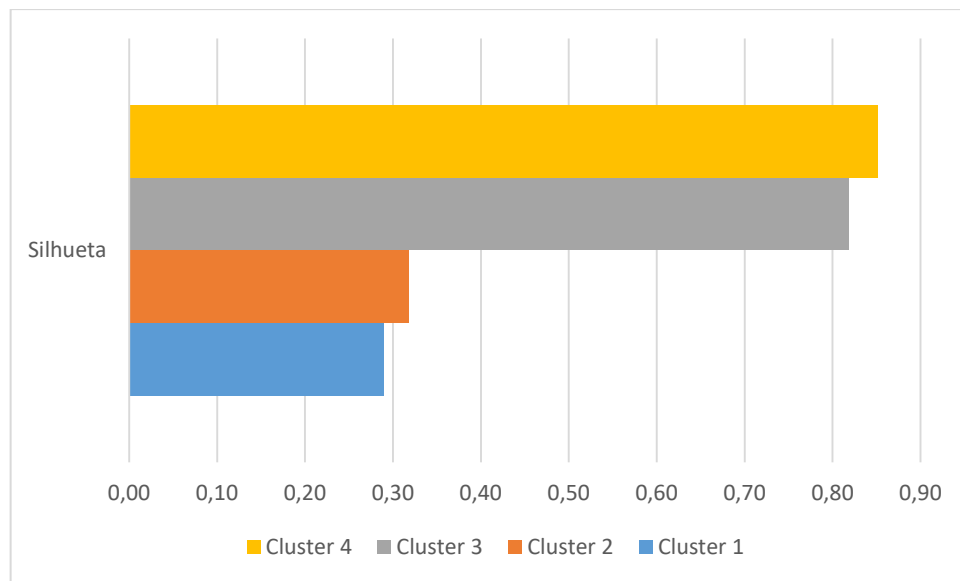
Além disso, é possível notar que o grupo de clientes do grupo três apresenta os piores índices para todas as características com exceção de cobertura de estoque, onde ele consegue ter uma maior facilidade para retirar seus itens do estoque, porém possuem uma baixa rentabilidade que é ampliada devido sua alta condicionalidade comercial, a qual poderia diminuir para melhorar o índice de NMCU.

Figura 17 – Média de características para cada *cluster*



Fonte: Próprio Autor (2018)

Para verificar a validação da segmentação de clientes, foi utilizado o método da silhueta e a Figura 18 traduz os índices de cada silhueta (*s*) para cada *cluster*. A segmentação dos *clusters* 4 e 3 podem ser classificadas como excelentes por possuírem índices maiores que 0,71 e a segmentação dos *cluster* 1 e 2 podem ser classificadas como fracas por possuírem índices entre 0,26 e 0,50.

Figura 18 – Silhueta de cada *cluster*

Fonte: Próprio Autor (2018)

De forma geral, pode-se concluir dos resultados que os clientes do *cluster 4* representam os clientes mais importante para a companhia, uma vez que estes correspondem aos clientes que tem maior rentabilidade e potencial de mercado para crescimento em termos de volume de vendas para estes, a boa relação com os clientes deve ser mantida e priorizada em relação aos outros clientes.

Já o *cluster 1* simboliza os clientes que possuem o segundo maior potencial de mercado e representam a segunda maior rentabilidade da companhia, além de serem clientes que compram um excelente mix de produtos da companhia, uma vez que o seu índice de DCMU é o maior entre todos os *clusters*. Para este grupo de clientes, estratégias como o aumento de suas condicionalidade pode traduzir em uma alavanca para os termos de volume e receita que estes clientes geram.

O *cluster 2*, apesar de possuir uma menor rentabilidade e potencial de mercado que os clientes do grupo 1 podem apresentar, de forma análoga, direcionais estratégicos similares aos desse grupo, porém em menor intensidade devido sua menor representatividade.

Por fim, o *cluster 3* é constituído do grupo de clientes que retêm os piores índices em termos de rentabilidade e potencial de mercado, além de um péssimo mix de produtos e elevada condicionalidade comercial. São clientes onde o investimento deveria ser menor devido ao retorno que estes trazem para a companhia.

5 CONSIDERAÇÕES FINAIS

As companhias de um modo geral possuem um número grande de clientes, podendo gerar complexidades e dificuldades na gestão de uma base tão vasta de clientes. Saber o valor de um cliente é extremamente importante para direcionais estratégicos de *marketing* e vendas. Além disso, estabelecer estratégias específicas para cada cliente é difícil de gerir e requer uma demanda de muito tempo e investimento. Por esta razão a segmentação de clientes em grupos menores auxilia neste processo de reconhecimento de valor do cliente e em direcionais estratégicos de acordo com a similaridade entre estes.

A mineração de dados ajuda na classificação e identificação de clientes, além de prever seus comportamentos. Uma das áreas mais comuns de aplicação de mineração de dados para CRM é a segmentação de clientes.

O foco deste trabalho foi verificar a segmentação de clientes determinando os critérios importantes para caracterizá-los utilizando como base o método adaptado de RFM, estabelecendo índices de relevância entre estes por meio do método AHP e agrupando os clientes em grupos pelo método de agrupamento *K-Means*. Os objetivos gerais e específicos foram atingidos, dessa modo possibilitando a conclusão dos benefícios em se realizar a segmentação de clientes para auxílio em direcionais estratégicos que uma dada companhia pode definir mediante a sua realidade no mercado, conforme exposto no Capítulo 4.

Os fatores mais críticos deste projeto são os dados coletados e o julgamento na aplicação do método AHP. Como eles influenciam diretamente os resultados do *ranking* final e no agrupamento pelo método *K-Means*, a pessoa que foi escolhida para julgar os critérios tem que ter um vasto conhecimento do processo analisado e do cenário onde a companhia em que se deseja realizar um processo semelhante está inserido, pois ela é responsável por expressar em números a realidade e a importância dos critérios avaliados. Caso contrário, o *ranking* e resultados finais podem refletir um resultado distorcido e indesejado.

Apesar de ter sido aplicado para um problema específico de uma empresa do setor de linha branca, este projeto pode ser replicado para qualquer pesquisa onde se tem por objetivo realizar o agrupamento de objetos dado certos critérios pré-estabelecidos. Além disso, na prática, a base de clientes, avaliação das características e suas importâncias são subjetivas e devem mudar de acordo com o tempo e o cenário em que a companhia está inserida.

Para futuras pesquisas seria interessante a aplicação de outros métodos de agrupamento de dados e comparação entre estes para um determinado projeto de segmentação de clientes e

utilizar mais de um julgador durante a aplicação do método AHP, de forma a obter um resultado mais neutro e imparcial.

REFERÊNCIAS

- AGGARWAL, Charu C.; REDDY, Chandan K. (Ed.). **Data clustering: algorithms and applications**. Boca Raton: CRC press, 2013.
- ALSHAWI, Sarmad; MISSI, Farouk; IRANI, Zahir. Organisational, technical and data quality factors in CRM adoption: SMEs perspective. **Industrial Marketing Management**, London, v. 40, n. 3, p. 376-383, 2011.
- ALVES, José Roberto Xavier; ALVES, João Murta. Definição de localidade para instalação industrial com o apoio do método de análise hierárquica (AHP). **Production Journal**, São José dos Campos, v. 25, n. 1, p. 13-26, 2015.
- AZIZ, Nor Filianie; SOROOSHIAN, Shahryar; MAHMUD, Fatimah. MCDM-AHP method in decision makings. **ARNP Journal of Engineering and Applied Sciences**, Pekan, v. 11, n. 11, p. 7217-7220, 2016.
- BAJAJ, A. A study on market segmentation of samsung electronics ltd. with special references to mobile phones. **International Journal of Advance Research, Ideas and Innovations in Technology**, Naya Raipur, v. 3, n. 5, p. 367–372, 2017.
- BASWADE, Anand M.; NALWADE, Prakash S. Selection of initial centroids for k-means algorithm. **IJCSMC**, Nanded, v. 2, n. 7, p. 161-164, 2013.
- BURNEY, SM Aqil; TARIQ, Humera. K-means cluster analysis for image segmentation. **International Journal of Computer Applications**, Carachi, v. 96, n. 4, 2014.
- CELEBI, M. Emre; KINGRAVI, Hassan A.; VELA, Patricio A. A comparative study of efficient initialization methods for the k-means clustering algorithm. **Expert systems with applications**, Shaveport, v. 40, n. 1, p. 200-210, 2013.
- CHO, Young Sung et al. Implementation of Personalized recommendation System using k-means Clustering of Item Category based on RFM. In: INTERNATIONAL CONFERENCE ON MANAGEMENT OF INNOVATION AND TECHNOLOGY (ICMIT), 2012, Sanur Bali. **Proceedings...** Sanur Bali: IEEE, p. 378-383, 2012.
- FLOH, Arne et al. Customer segmentation using unobserved heterogeneity in the perceived-value–loyalty–intentions link. **Journal of Business Research**, Amsterdam, v. 67, n. 5, p. 974-982, 2014.
- HAN, Jiawei; PEI, Jian; KAMBER, Micheline. **Data mining: concepts and techniques**. Waltham: Elsevier, 2011.
- KHAJVAND, Mahboubeh et al. Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study. **Procedia Computer Science**, Amsterdam, v. 3, p. 57-63, 2011.

KHODAKARAMI, Farnoosh; CHAN, Yolande E. Exploring the role of customer relationship management (CRM) systems in customer knowledge creation. **Information and Management**, Amsterdam, v. 51, n. 1, p. 27-42, 2014.

KODINARIYA, Trupti M.; MAKWANA, Prashant R. Review on determining number of Cluster in K-Means Clustering. **International Journal**, Bavnagar, v. 1, n. 6, p. 90-95, 2013.

KRACKLAUER, Alexander H.; MILLS, D. Quinn; SEIFERT, Dirk. Customer management as the origin of collaborative customer relationship management. In: _____. **Collaborative customer relationship management**. Berlim: Springer, 2004. p. 3-6.

MARDANI, Abbas; JUSOH, Ahmad; ZAVADSKAS, Edmundas Kazimieras. Fuzzy multiple criteria decision-making techniques and applications—two decades review from 1994 to 2014. **Expert systems with Applications**, Skudai, v. 42, n. 8, p. 4126-4148, 2015.

MENDE, Martin; BOLTON, Ruth N.; BITNER, Mary Jo. Decoding customer–firm relationships: how attachment styles help explain customers' preferences for closeness, repurchase intentions, and changes in relationship breadth. **Journal of Marketing Research**, Chicago, v. 50, n. 1, p. 125-142, 2013.

MOHAMMADHOSSEIN, Nastaran; ZAKARIA, Nor Hidayati. CRM benefits for customers: literature review (2005-2012). **International Journal of Engineering Research and Applications**, Skudai, v. 2, n. 6, p. 1578-1586, 2012.

MORABITO, R. et al. **Metodologia de pesquisa em engenharia de produção e gestão de operações**. Rio de Janeiro: Elsevier Brasil, 2018.

NGAI, Eric WT et al. The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. **Decision support systems**, Hong Kong, v. 50, n. 3, p. 559-569, 2011.

NGAI, Eric WT; XIU, Li; CHAU, Dorothy CK. Application of data mining techniques in customer relationship management: a literature review and classification. **Expert systems with applications**, Hong Kong, v. 36, n. 2, p. 2592-2602, 2009.

NGUYEN, Bang; MUTUM, Dilip S. A review of customer relationship management: successes, advances, pitfalls and futures. **Business Process Management Journal**, Oxford, v. 18, n. 3, p. 400-419, 2012.

OLSON, David L.; CHAE, Bongsug Kevin. Direct marketing decision support through predictive customer response modeling. **Decision Support Systems**, Lincoln, v. 54, n. 1, p. 443-451, 2012.

OLSZAK, Celina. Multi-agent framework for social customer relationship management systems. In: PROCEEDINGS OF THE INFORMING SCIENCE AND INFORMATION TECHNOLOGY EDUCATION CONFERENCE, 2013, Katowice. **Proceedings...** Katowice: Informing Science Institute, p. 367-387, 2013.

OUMA, Yashon; TATEISHI, Ryutaro. Urban flood vulnerability and risk mapping using integrated multi-parametric AHP and GIS: methodological overview and case study assessment. **Water**, Eldoret, v. 6, n. 6, p. 1515-1545, 2014.

PATEL, Hetal; PATEL, Dharmendra. A brief survey of data mining techniques applied to agricultural data. **International Journal of Computer Applications**, Changa, v. 95, n. 9, 2014.

PELTIER, James W.; ZAHAY, Debra; LEHMANN, Donald R. Organizational learning and CRM success: a model for linking organizational practices, customer data quality, and performance. **Journal of Interactive Marketing**, Whitewater, v. 27, n. 1, p. 1-13, 2013.

PEPPERS, Don; ROGERS, Martha. **Managing customer experience and relationships: a strategic framework**. Hoboken: John Wiley and Sons, 2016.

SAATY, Thomas L.; VARGAS, Luis G. **Models, methods, concepts and applications of the analytic hierarchy process**. New York: Springer Science and Business Media, 2012.

ŠEBJAN, Urban; BOBEK, Samo; TOMINC, Polona. Factors influencing attitudes towards the use of CRM's analytical tools in organizations. **Organizacija**, Maribor, v. 49, n. 1, p. 28-41, 2016.

SHAFIQUE, Umair; QAISER, Haseeb. A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). **International Journal of Innovation and Scientific Research**, Gujrat, v. 12, n. 1, p. 217-222, 2014.

SHAON, SM Kamrul Islam; RAHMAN, Hasebur. A theoretical review of CRM effects on customer satisfaction and loyalty. **Central European Business Review**, Prague, v. 4, n. 1, p. 23, 2015.

CALIXTA, Mauro; AFONSO, Tarcisio; LOCATELLI, Ronaldo Lamounier. Segmentação de mercado, diferenciação de produtos e a perspectiva da antropologia do consumo. **Revista Gestão e Tecnologia**, Pedro Leopoldo, v. 11, n. 1, p. 106-122, 2011.

TRAMARICO, Claudemir Leif et al. Analytic hierarchy process and supply chain management: a bibliometric study. **Procedia Computer Science**, v. 55, p. 441-450, 2015.

TSENG, Shu-Mei; WU, Pin-Hong. The impact of customer knowledge and customer relationship management on service quality. **International journal of quality and service sciences**, Kaohsiung, v. 6, n. 1, p. 77-96, 2014.

VIANA, Joana Coelho; ALENCAR, Luciana Hazin. Metodologias para seleção de fornecedores: uma revisão da literatura. **Production**, v. 22, n. 4, p. 625-636, 2012.

VINODH, S.; JOY, Dino. Structural equation modelling of lean manufacturing practices. **International Journal of Production Research**, Tiruchirappalli, v. 50, n. 6, p. 1598-1607, 2012.

WANG, Juntao; SU, Xiaolong. An improved K-Means clustering algorithm. In: 2011 IEEE 3RD INTERNATIONAL CONFERENCE ON COMMUNICATION SOFTWARE AND NETWORKS, 2011, Xi'an. **Proceedings...** Xi'an: IEEE, p. 44-46, 2011.

WITTEN, Ian H. et al. **Data mining**: practical machine learning tools and techniques. Cambridge: Morgan Kaufmann, 2016.

YOO, Illhoi et al. Data mining in healthcare and biomedicine: a survey of the literature. **Journal of medical systems**, Basel, v. 36, n. 4, p. 2431-2448, 2012.

APÊNDICE A – Dados primários

Cliente/Critério	MCU	MCUP	CCP	CE	PM
CLIENTE 1	15.368.854,84	33,45	0,12	234,78	135.444,87
CLIENTE 2	802.873,72	29,23	0,02	58,20	3.605,55
CLIENTE 3	13.441.835,17	34,46	0,06	281,03	115.141,74
CLIENTE 4	2.000.657,49	38,49	0,02	194,95	12.321,84
CLIENTE 5	817.051,43	17,49	0,07	58,62	8.419,90
CLIENTE 6	15.040.140,05	28,05	0,08	133,09	93.921,96
CLIENTE 7	3.027.533,28	32,37	0,05	174,38	41.382,30
CLIENTE 8	976.057,56	26,44	0,03	462,93	13.551,63
CLIENTE 9	2.503.240,82	25,22	0,08	144,85	56.961,67
CLIENTE 10	2.678.458,21	29,55	0,15	170,37	62.555,82
CLIENTE 11	2.237.588,21	33,68	0,05	178,01	37.454,84
CLIENTE 12	976.887,37	32,08	0,01	503,99	1.269,19
CLIENTE 13	3.882.855,96	30,07	0,06	145,59	89.139,00
CLIENTE 14	720.253,05	22,27	0,14	50,15	12.506,82
CLIENTE 15	2.375.850,96	27,46	0,08	234,87	37.856,40
CLIENTE 16	1.656.963,35	28,84	0,04	322,41	2.860,42
CLIENTE 17	12.646.672,45	29,69	0,05	56,82	99.231,44
CLIENTE 18	13.853.803,61	35,77	0,02	198,17	152.124,12
CLIENTE 19	13.330.982,94	27,40	0,09	177,56	205.166,80
CLIENTE 20	95.848.669,27	29,38	0,07	782,48	524.652,10
CLIENTE 21	1.688.966,69	29,32	0,07	112,02	10.013,78
CLIENTE 22	4.594.351,29	32,01	0,05	446,59	134.851,25
CLIENTE 23	433.098,99	30,02	0,02	946,63	9.993,12
CLIENTE 24	3.772.366,69	34,51	0,05	556,57	21.678,83
CLIENTE 25	1.124.093,77	41,25	0,03	145,77	9.006,88
CLIENTE 26	493.064,73	44,21	-	194,95	2.398,71
CLIENTE 27	4.459.054,16	30,80	0,07	221,34	55.712,07
CLIENTE 28	34.025.397,28	28,49	0,04	192,96	318.736,10
CLIENTE 29	4.098.587,39	36,43	0,06	310,42	30.006,81
CLIENTE 30	3.750.565,40	27,21	0,10	215,58	62.163,00
CLIENTE 31	9.168.003,85	36,15	0,04	57,99	41.244,48
CLIENTE 32	5.081.441,17	34,16	0,05	295,98	18.728,71
CLIENTE 33	19.415.768,47	26,70	0,09	76,68	114.982,61
CLIENTE 34	634.211,15	37,07	0,02	106,65	5.885,44
CLIENTE 35	5.663.169,45	35,91	0,04	100,15	31.425,98
CLIENTE 36	4.007.851,46	33,47	0,06	59,83	27.658,72
CLIENTE 37	4.659.372,35	30,22	0,08	437,24	58.962,14
CLIENTE 38	254.334,97	17,89	0,20	50,26	4.540,48
CLIENTE 39	816.408,86	28,52	0,08	192,77	5.544,64
CLIENTE 40	10.856.989,16	36,72	0,02	67,15	80.119,04

CLIENTE 41	1.078.731,70	37,77	0,02	176,99	17.944,00
CLIENTE 42	4.262.964,50	28,85	0,03	336,53	98.648,82
CLIENTE 43	413.639,77	43,70	-	209,22	227,73
CLIENTE 44	1.851.921,88	27,22	0,12	109,95	37.846,04
CLIENTE 45	896.859,86	33,31	0,03	67,37	50.973,97
CLIENTE 46	3.176.623,76	31,82	0,07	248,12	32.707,43
CLIENTE 47	736.286,34	22,28	0,10	209,72	11.154,19
CLIENTE 48	708.622,66	31,33	0,01	183,22	4.439,15
CLIENTE 49	4.587.702,07	38,83	0,02	84,37	22.788,14
CLIENTE 50	1.906.253,78	13,40	0,03	209,22	2.430,65
CLIENTE 51	3.615.086,34	36,42	0,05	96,32	132.061,21
CLIENTE 52	505.448,26	23,12	-	209,22	215,78
CLIENTE 53	6.228.195,08	27,50	0,02	209,22	6.419,89
CLIENTE 54	2.055.598,50	35,87	0,04	125,53	24.762,89
CLIENTE 55	42.333.398,09	28,43	0,09	347,95	316.152,71
CLIENTE 56	4.920.345,08	28,82	0,08	275,46	44.875,60
CLIENTE 57	103.329,10	27,13	0,00	194,95	335,77
CLIENTE 58	1.328.056,73	30,99	0,02	157,20	5.527,50
CLIENTE 59	139.238,08	36,36	0,00	209,22	21,92
CLIENTE 60	10.509,29	27,89	0,00	194,95	142,11
CLIENTE 61	418.578,42	26,85	0,01	209,22	957,01
CLIENTE 62	285.747,39	20,78	0,10	197,63	3.081,27
CLIENTE 63	639.185,65	31,04	0,06	197,63	5.227,47
CLIENTE 64	404.567,33	26,17	0,01	209,22	827,25
CLIENTE 65	381.193,00	33,48	-	197,63	613,73
CLIENTE 66	9.670.866,97	32,82	0,02	194,95	54.589,33
CLIENTE 67	21.927.302,89	24,17	0,04	67,91	205.115,00
CLIENTE 68	1.441.901,70	32,25	0,02	194,95	5.102,21
CLIENTE 69	362.943,69	36,38	0,00	194,95	909,35
CLIENTE 70	3.704.463,96	31,03	0,01	194,95	10.289,87
CLIENTE 71	7.040.359,28	29,78	0,01	194,95	89.901,00
CLIENTE 72	10.660.838,17	35,50	0,03	194,95	100.700,67
CLIENTE 73	1.415.674,95	31,20	0,01	194,95	5.227,07
CLIENTE 74	4.195.529,49	34,55	0,01	194,95	21.944,67
CLIENTE 75	3.033.724,51	28,78	0,01	194,95	19.565,50
CLIENTE 76	760.464,21	29,04	0,01	194,95	3.841,82
CLIENTE 77	93.914,04	17,26	-	194,95	4.157,53
CLIENTE 78	4.306.563,12	28,62	0,01	194,95	4.544,84
CLIENTE 79	5.520.677,76	30,74	0,04	148,61	40.898,34
CLIENTE 80	4.933.140,51	32,42	0,06	69,49	41.594,13
CLIENTE 81	4.729.486,84	30,51	0,02	194,95	101.335,83

CLIENTE 82	1.021.115,34	29,66	0,01	194,95	4.171,95
CLIENTE 83	1.018.673,06	29,71	0,01	194,95	3.783,97
CLIENTE 84	1.169.265,00	29,82	0,01	194,95	4.288,61
CLIENTE 85	1.104.931,97	30,34	0,01	194,95	4.015,19
CLIENTE 86	465.328,76	41,90	0,02	280,75	2.657,65
CLIENTE 87	986.472,24	20,50	0,01	194,95	4.157,53
CLIENTE 88	1.287.384,19	31,89	0,01	194,95	3.977,90
CLIENTE 89	1.723.717,57	29,35	0,01	194,95	4.424,37
CLIENTE 90	747.478,56	28,13	0,06	194,95	3.722,49
CLIENTE 91	916.519,95	27,64	0,01	194,95	3.781,30
CLIENTE 92	424.148,25	33,58	0,00	194,95	1.284,39
CLIENTE 93	1.075.781,05	27,90	0,01	194,95	4.099,84
CLIENTE 94	935.090,53	30,62	0,01	194,95	3.018,99
CLIENTE 95	4.277.277,45	35,56	-	446,59	45.931,56
CLIENTE 96	667.029,50	30,25	0,01	194,95	2.434,88
CLIENTE 97	159.495,37	29,05	0,01	194,95	1.011,43
CLIENTE 98	986.402,81	28,19	0,01	194,95	3.022,97
CLIENTE 99	772.840,75	27,02	0,01	194,95	3.184,70
CLIENTE 100	839.786,63	24,28	0,01	194,95	4.192,86
CLIENTE 101	151.058.480,05	27,42	0,10	209,35	1.548.453,00
CLIENTE 102	2.167.006,05	38,93	0,02	194,95	6.022,00
CLIENTE 103	679.959,57	29,86	0,01	194,95	2.508,78
CLIENTE 104	781.735,53	28,77	0,01	194,95	3.404,44
CLIENTE 105	951.816,09	26,56	0,01	194,95	2.842,22
CLIENTE 106	425.431,34	27,32	0,02	194,95	1.769,91
CLIENTE 107	895.942,41	26,69	0,01	194,95	3.603,41
CLIENTE 108	853.395,40	32,01	0,01	194,95	2.561,54
CLIENTE 109	563.716,48	30,03	0,01	194,95	1.875,56
CLIENTE 110	143.790,26	29,17	0,00	194,95	371,75
CLIENTE 111	315.628,97	32,44	0,00	194,95	1.008,80
CLIENTE 112	24.965.193,79	28,19	0,05	111,68	33.644,58
CLIENTE 113	126.670,55	28,09	0,06	194,95	4.909,57
CLIENTE 114	1.609.449,55	20,30	0,02	209,22	142,85
CLIENTE 115	40.695,05	35,06	0,00	194,95	89,42
CLIENTE 116	24.629.532,19	29,83	0,08	446,59	219.625,05
CLIENTE 117	711.906,83	23,13	0,01	194,95	2.748,87
CLIENTE 118	1.900.252,00	31,81	0,13	111,68	22.150,77
CLIENTE 119	179.426,13	33,01	0,00	194,95	246,29

APÊNDICE B – Dados normalizados

Cliente/Critério	NMCU	NMCUP	NCCP	NCE	NPM
CLIENTE 1	0,1017	0,6505	0,6114	0,2059	0,0875
CLIENTE 2	0,0052	0,5137	0,1010	0,0090	0,0023
CLIENTE 3	0,0889	0,6834	0,2927	0,2575	0,0743
CLIENTE 4	0,0132	0,8144	0,0879	0,1615	0,0079
CLIENTE 5	0,0053	0,1327	0,3343	0,0095	0,0054
CLIENTE 6	0,0995	0,4753	0,3892	0,0925	0,0606
CLIENTE 7	0,0200	0,6157	0,2382	0,1386	0,0267
CLIENTE 8	0,0064	0,4232	0,1561	0,4605	0,0087
CLIENTE 9	0,0165	0,3834	0,3905	0,1056	0,0368
CLIENTE 10	0,0177	0,5240	0,7775	0,1341	0,0404
CLIENTE 11	0,0147	0,6581	0,2570	0,1426	0,0242
CLIENTE 12	0,0064	0,6061	0,0566	0,5062	0,0008
CLIENTE 13	0,0256	0,5409	0,3227	0,1065	0,0576
CLIENTE 14	0,0047	0,2879	0,6937	-	0,0081
CLIENTE 15	0,0157	0,4561	0,3848	0,2061	0,0244
CLIENTE 16	0,0109	0,5009	0,1835	0,3037	0,0018
CLIENTE 17	0,0837	0,5286	0,2453	0,0074	0,0641
CLIENTE 18	0,0916	0,7258	0,0788	0,1651	0,0982
CLIENTE 19	0,0882	0,4543	0,4603	0,1421	0,1325
CLIENTE 20	0,6345	0,5186	0,3631	0,8169	0,3388
CLIENTE 21	0,0111	0,5166	0,3707	0,0690	0,0065
CLIENTE 22	0,0303	0,6037	0,2653	0,4422	0,0871
CLIENTE 23	0,0028	0,5394	0,1145	1,0000	0,0064
CLIENTE 24	0,0249	0,6850	0,2457	0,5649	0,0140
CLIENTE 25	0,0074	0,9037	0,1304	0,1067	0,0058
CLIENTE 26	0,0032	1,0000	-	0,1615	0,0015
CLIENTE 27	0,0295	0,5646	0,3321	0,1910	0,0360
CLIENTE 28	0,2252	0,4895	0,1935	0,1593	0,2058
CLIENTE 29	0,0271	0,7474	0,3027	0,2903	0,0194
CLIENTE 30	0,0248	0,4482	0,5069	0,1845	0,0401
CLIENTE 31	0,0606	0,7382	0,1984	0,0088	0,0266
CLIENTE 32	0,0336	0,6735	0,2441	0,2742	0,0121
CLIENTE 33	0,1285	0,4315	0,4732	0,0296	0,0742
CLIENTE 34	0,0041	0,7681	0,0863	0,0630	0,0038
CLIENTE 35	0,0374	0,7306	0,2128	0,0558	0,0203
CLIENTE 36	0,0265	0,6512	0,3078	0,0108	0,0178
CLIENTE 37	0,0308	0,5456	0,3873	0,4318	0,0381
CLIENTE 38	0,0016	0,1455	1,0000	0,0001	0,0029
CLIENTE 39	0,0053	0,4905	0,3969	0,1591	0,0036
CLIENTE 40	0,0718	0,7566	0,1256	0,0190	0,0517

CLIENTE 41	0,0071	0,7908	0,1007	0,1415	0,0116
CLIENTE 42	0,0282	0,5015	0,1553	0,3194	0,0637
CLIENTE 43	0,0027	0,9832	-	0,1774	0,0001
CLIENTE 44	0,0122	0,4485	0,5966	0,0667	0,0244
CLIENTE 45	0,0059	0,6460	0,1550	0,0192	0,0329
CLIENTE 46	0,0210	0,5976	0,3457	0,2208	0,0211
CLIENTE 47	0,0048	0,2882	0,5189	0,1780	0,0072
CLIENTE 48	0,0046	0,5818	0,0761	0,1484	0,0029
CLIENTE 49	0,0303	0,8253	0,0786	0,0382	0,0147
CLIENTE 50	0,0126	-	0,1676	0,1774	0,0016
CLIENTE 51	0,0239	0,7469	0,2711	0,0515	0,0853
CLIENTE 52	0,0033	0,3153	-	0,1774	0,0001
CLIENTE 53	0,0412	0,4575	0,1013	0,1774	0,0041
CLIENTE 54	0,0135	0,7290	0,1984	0,0841	0,0160
CLIENTE 55	0,2802	0,4877	0,4480	0,3322	0,2042
CLIENTE 56	0,0325	0,5004	0,4119	0,2513	0,0290
CLIENTE 57	0,0006	0,4454	0,0005	0,1615	0,0002
CLIENTE 58	0,0087	0,5708	0,0983	0,1194	0,0036
CLIENTE 59	0,0009	0,7451	0,0054	0,1774	-
CLIENTE 60	-	0,4702	0,0011	0,1615	0,0001
CLIENTE 61	0,0027	0,4363	0,0421	0,1774	0,0006
CLIENTE 62	0,0018	0,2392	0,4925	0,1645	0,0020
CLIENTE 63	0,0042	0,5723	0,3208	0,1645	0,0034
CLIENTE 64	0,0026	0,4145	0,0452	0,1774	0,0005
CLIENTE 65	0,0025	0,6515	-	0,1645	0,0004
CLIENTE 66	0,0640	0,6301	0,0877	0,1615	0,0352
CLIENTE 67	0,1451	0,3495	0,2152	0,0198	0,1325
CLIENTE 68	0,0095	0,6118	0,0803	0,1615	0,0033
CLIENTE 69	0,0023	0,7456	0,0060	0,1615	0,0006
CLIENTE 70	0,0245	0,5722	0,0605	0,1615	0,0066
CLIENTE 71	0,0465	0,5315	0,0360	0,1615	0,0580
CLIENTE 72	0,0705	0,7170	0,1328	0,1615	0,0650
CLIENTE 73	0,0093	0,5776	0,0341	0,1615	0,0034
CLIENTE 74	0,0277	0,6864	0,0599	0,1615	0,0142
CLIENTE 75	0,0200	0,4991	0,0405	0,1615	0,0126
CLIENTE 76	0,0050	0,5074	0,0734	0,1615	0,0025
CLIENTE 77	0,0006	0,1251	-	0,1615	0,0027
CLIENTE 78	0,0284	0,4937	0,0710	0,1615	0,0029
CLIENTE 79	0,0365	0,5627	0,2287	0,1098	0,0264
CLIENTE 80	0,0326	0,6172	0,2853	0,0216	0,0268
CLIENTE 81	0,0312	0,5551	0,1117	0,1615	0,0654

CLIENTE 82	0,0067	0,5275	0,0458	0,1615	0,0027
CLIENTE 83	0,0067	0,5292	0,0408	0,1615	0,0024
CLIENTE 84	0,0077	0,5329	0,0434	0,1615	0,0028
CLIENTE 85	0,0072	0,5497	0,0434	0,1615	0,0026
CLIENTE 86	0,0030	0,9251	0,0806	0,2572	0,0017
CLIENTE 87	0,0065	0,2303	0,0497	0,1615	0,0027
CLIENTE 88	0,0085	0,6001	0,0314	0,1615	0,0026
CLIENTE 89	0,0113	0,5175	0,0280	0,1615	0,0028
CLIENTE 90	0,0049	0,4781	0,3158	0,1615	0,0024
CLIENTE 91	0,0060	0,4621	0,0502	0,1615	0,0024
CLIENTE 92	0,0027	0,6548	0,0049	0,1615	0,0008
CLIENTE 93	0,0071	0,4705	0,0399	0,1615	0,0026
CLIENTE 94	0,0061	0,5589	0,0342	0,1615	0,0019
CLIENTE 95	0,0282	0,7190	-	0,4422	0,0296
CLIENTE 96	0,0043	0,5467	0,0323	0,1615	0,0016
CLIENTE 97	0,0010	0,5077	0,0401	0,1615	0,0006
CLIENTE 98	0,0065	0,4798	0,0443	0,1615	0,0019
CLIENTE 99	0,0050	0,4418	0,0462	0,1615	0,0020
CLIENTE 100	0,0055	0,3529	0,0422	0,1615	0,0027
CLIENTE 101	1,0000	0,4548	0,4984	0,1776	1,0000
CLIENTE 102	0,0143	0,8285	0,0945	0,1615	0,0039
CLIENTE 103	0,0044	0,5342	0,0289	0,1615	0,0016
CLIENTE 104	0,0051	0,4989	0,0481	0,1615	0,0022
CLIENTE 105	0,0062	0,4270	0,0408	0,1615	0,0018
CLIENTE 106	0,0027	0,4516	0,0811	0,1615	0,0011
CLIENTE 107	0,0059	0,4312	0,0443	0,1615	0,0023
CLIENTE 108	0,0056	0,6037	0,0434	0,1615	0,0016
CLIENTE 109	0,0037	0,5395	0,0712	0,1615	0,0012
CLIENTE 110	0,0009	0,5119	0,0013	0,1615	0,0002
CLIENTE 111	0,0020	0,6178	0,0045	0,1615	0,0006
CLIENTE 112	0,1652	0,4799	0,2699	0,0686	0,0217
CLIENTE 113	0,0008	0,4768	0,3106	0,1615	0,0032
CLIENTE 114	0,0106	0,2238	0,0921	0,1774	0,0001
CLIENTE 115	0,0002	0,7027	0,0002	0,1615	0,0000
CLIENTE 116	0,1630	0,5333	0,3839	0,4422	0,1418
CLIENTE 117	0,0046	0,3157	0,0494	0,1615	0,0018
CLIENTE 118	0,0125	0,5973	0,6388	0,0686	0,0143
CLIENTE 119	0,0011	0,6364	0,0072	0,1615	0,0001

APÊNDICE C – Pontuação geral e *ranking*

Cliente/Critério	PG	Ranking
CLIENTE 1	0,3079	6
CLIENTE 2	0,1709	91
CLIENTE 3	0,2926	12
CLIENTE 4	0,2746	20
CLIENTE 5	0,0663	117
CLIENTE 6	0,2292	43
CLIENTE 7	0,2267	46
CLIENTE 8	0,1642	95
CLIENTE 9	0,1622	97
CLIENTE 10	0,2329	38
CLIENTE 11	0,2387	34
CLIENTE 12	0,2158	50
CLIENTE 13	0,2148	52
CLIENTE 14	0,1376	108
CLIENTE 15	0,1860	75
CLIENTE 16	0,1852	76
CLIENTE 17	0,2277	44
CLIENTE 18	0,2942	11
CLIENTE 19	0,2363	37
CLIENTE 20	0,5379	2
CLIENTE 21	0,1942	71
CLIENTE 22	0,2503	28
CLIENTE 23	0,2162	49
CLIENTE 24	0,2646	22
CLIENTE 25	0,3006	9
CLIENTE 26	0,3222	4
CLIENTE 27	0,2239	47
CLIENTE 28	0,3007	8
CLIENTE 29	0,2795	16
CLIENTE 30	0,1968	67
CLIENTE 31	0,2749	19
CLIENTE 32	0,2535	25
CLIENTE 33	0,2327	39
CLIENTE 34	0,2518	27
CLIENTE 35	0,2644	23
CLIENTE 36	0,2387	35
CLIENTE 37	0,2313	40
CLIENTE 38	0,1099	113
CLIENTE 39	0,1880	73
CLIENTE 40	0,2853	13

CLIENTE 41	0,2653	21
CLIENTE 42	0,2018	62
CLIENTE 43	0,3171	5
CLIENTE 44	0,1903	72
CLIENTE 45	0,2217	48
CLIENTE 46	0,2303	42
CLIENTE 47	0,1331	109
CLIENTE 48	0,1957	68
CLIENTE 49	0,2812	14
CLIENTE 50	0,0226	119
CLIENTE 51	0,2783	18
CLIENTE 52	0,1072	114
CLIENTE 53	0,1747	89
CLIENTE 54	0,2533	26
CLIENTE 55	0,3452	3
CLIENTE 56	0,2111	56
CLIENTE 57	0,1464	105
CLIENTE 58	0,1944	69
CLIENTE 59	0,2417	32
CLIENTE 60	0,1540	101
CLIENTE 61	0,1477	103
CLIENTE 62	0,1135	111
CLIENTE 63	0,2087	57
CLIENTE 64	0,1410	107
CLIENTE 65	0,2122	54
CLIENTE 66	0,2424	30
CLIENTE 67	0,2071	60
CLIENTE 68	0,2081	58
CLIENTE 69	0,2420	31
CLIENTE 70	0,2012	65
CLIENTE 71	0,2047	61
CLIENTE 72	0,2803	15
CLIENTE 73	0,1943	70
CLIENTE 74	0,2397	33
CLIENTE 75	0,1761	87
CLIENTE 76	0,1728	90
CLIENTE 77	0,0460	118
CLIENTE 78	0,1782	83
CLIENTE 79	0,2152	51
CLIENTE 80	0,2311	41
CLIENTE 81	0,2117	55

CLIENTE 82	0,1781	84
CLIENTE 83	0,1782	82
CLIENTE 84	0,1800	80
CLIENTE 85	0,1851	77
CLIENTE 86	0,3072	7
CLIENTE 87	0,0847	116
CLIENTE 88	0,2007	66
CLIENTE 89	0,1758	88
CLIENTE 90	0,1787	81
CLIENTE 91	0,1575	99
CLIENTE 92	0,2136	53
CLIENTE 93	0,1599	98
CLIENTE 94	0,1869	74
CLIENTE 95	0,2594	24
CLIENTE 96	0,1821	78
CLIENTE 97	0,1688	92
CLIENTE 98	0,1628	96
CLIENTE 99	0,1504	102
CLIENTE 100	0,1224	110
CLIENTE 101	0,7665	1
CLIENTE 102	0,2792	17
CLIENTE 103	0,1780	85
CLIENTE 104	0,1685	93
CLIENTE 105	0,1458	106
CLIENTE 106	0,1545	100
CLIENTE 107	0,1473	104
CLIENTE 108	0,2013	64
CLIENTE 109	0,1819	79
CLIENTE 110	0,1675	94
CLIENTE 111	0,2016	63
CLIENTE 112	0,2431	29
CLIENTE 113	0,1764	86
CLIENTE 114	0,0872	115
CLIENTE 115	0,2272	45
CLIENTE 116	0,3002	10
CLIENTE 117	0,1106	112
CLIENTE 118	0,2383	36
CLIENTE 119	0,2072	59

APÊNDICE D – Dados clientes por *cluster*

Ciente	Cluster
CLIENTE 1	1
CLIENTE 2	2
CLIENTE 3	1
CLIENTE 4	1
CLIENTE 5	3
CLIENTE 6	1
CLIENTE 7	1
CLIENTE 8	2
CLIENTE 9	2
CLIENTE 10	1
CLIENTE 11	1
CLIENTE 12	2
CLIENTE 13	2
CLIENTE 14	3
CLIENTE 15	2
CLIENTE 16	2
CLIENTE 17	1
CLIENTE 18	1
CLIENTE 19	1
CLIENTE 20	4
CLIENTE 21	2
CLIENTE 22	1
CLIENTE 23	2
CLIENTE 24	1
CLIENTE 25	1
CLIENTE 26	1
CLIENTE 27	1
CLIENTE 28	1
CLIENTE 29	1
CLIENTE 30	2
CLIENTE 31	1
CLIENTE 32	1
CLIENTE 33	1
CLIENTE 34	1
CLIENTE 35	1
CLIENTE 36	1
CLIENTE 37	1
CLIENTE 38	3
CLIENTE 39	2
CLIENTE 40	1

CLIENTE 41	1
CLIENTE 42	2
CLIENTE 43	1
CLIENTE 44	2
CLIENTE 45	2
CLIENTE 46	1
CLIENTE 47	3
CLIENTE 48	2
CLIENTE 49	1
CLIENTE 50	3
CLIENTE 51	1
CLIENTE 52	3
CLIENTE 53	2
CLIENTE 54	1
CLIENTE 55	1
CLIENTE 56	2
CLIENTE 57	2
CLIENTE 58	2
CLIENTE 59	1
CLIENTE 60	2
CLIENTE 61	2
CLIENTE 62	3
CLIENTE 63	2
CLIENTE 64	2
CLIENTE 65	2
CLIENTE 66	1
CLIENTE 67	2
CLIENTE 68	2
CLIENTE 69	1
CLIENTE 70	2
CLIENTE 71	2
CLIENTE 72	1
CLIENTE 73	2
CLIENTE 74	1
CLIENTE 75	2
CLIENTE 76	2
CLIENTE 77	3
CLIENTE 78	2
CLIENTE 79	2
CLIENTE 80	1
CLIENTE 81	2

CLIENTE 82	2
CLIENTE 83	2
CLIENTE 84	2
CLIENTE 85	2
CLIENTE 86	1
CLIENTE 87	3
CLIENTE 88	2
CLIENTE 89	2
CLIENTE 90	2
CLIENTE 91	2
CLIENTE 92	2
CLIENTE 93	2
CLIENTE 94	2
CLIENTE 95	1
CLIENTE 96	2
CLIENTE 97	2
CLIENTE 98	2
CLIENTE 99	2
CLIENTE 100	3
CLIENTE 101	4
CLIENTE 102	1
CLIENTE 103	2
CLIENTE 104	2
CLIENTE 105	2
CLIENTE 106	2
CLIENTE 107	2
CLIENTE 108	2
CLIENTE 109	2
CLIENTE 110	2
CLIENTE 111	2
CLIENTE 112	1
CLIENTE 113	2
CLIENTE 114	3
CLIENTE 115	1
CLIENTE 116	1
CLIENTE 117	3
CLIENTE 118	1
CLIENTE 119	2