UNIVERSIDADE FEDERAL DE SÃO PAULO
ESCOLA PAULISTA DE POLÍTICA, ECONOMIA E NEGÓCIOS – EPPEN

GUILHERME PANTIGA FIGUEIREDO DE MELO

CONTRIBUIÇÃO DA ABORDAGEM *FUZZY* EM TÉCNICAS DE AGRUPAMENTO: UMA APLICAÇÃO EM CRM

Osasco
2019

UNIVERSIDADE FEDERAL DE SÃO PAULO

ESCOLA PAULISTA DE POLÍTICA, ECONOMIA E NEGÓCIOS – EPPEN

GUILHERME PANTIGA FIGUEIREDO DE MELO

CONTRIBUIÇÃO DA ABORDAGEM *FUZZY* EM TÉCNICAS DE AGRUPAMENTO: UMA APLICAÇÃO EM CRM

Projeto do Trabalho de Conclusão de Curso apresentado ao Curso de Administração da Escola Paulista de Política, Economia e Negócios – EPPEN da Universidade Federal de São Paulo – Unifesp como requisito para obtenção do título de Bacharel em Administração.

Orientador: Prof. Dr. Emerson Gomes dos Santos

Osasco

2019

Ficha catalográfica elaborada pela Biblioteca Unifesp Osasco, CRB-8: 3998,
e Departamento de Tecnologia da Informação Unifesp Osasco,
com os dados fornecidos pelo(a) autor(a)

M528c MELO, Guilherme Pantiga Figueiredo de
Contribuição da abordagem Fuzzy em técnicas de agrupamento: uma aplicação em CRM / Guilherme Pantiga Figueiredo de Melo. - 2019.
48 f. :il.

Trabalho de conclusão de curso (Administração) - Universidade Federal de São Paulo - Escola Paulista de Política, Economia e Negócios, Osasco, 2019.
Orientador: Emerson Gomes dos Santos.

1. CRM (Customer Relationship Management). 2. Ciência de dados. 3. RFM (Recency, Frequency, Monetary). 4. Clusterização. 5. Abordagem fuzzy. I. Santos, Emerson Gomes dos, II. TCC - Unifesp/EPPEN. III. Título.

CDD: 658

Dedico este trabalho a todos que contribuíram direta ou indiretamente em minha formação acadêmica.

AGRADECIMENTOS

Meus sinceros agradecimentos a todos que contribuíram nesta jornada, especialmente:

Minha família, principalmente meus pais, Claudia e Maurício, por me apoiarem e me guiarem em todos os momentos.

A Bruna, pelo companheirismo e paciência.

Ao orientador Prof. Dr. Emerson Gomes dos Santos pela colaboração essencial neste trabalho.

Aos meus amigos pelo incentivo.

RESUMO

Atualmente o cenário em que as organizações estão inseridas é de alta competitividade e adaptar-se é essencial para sua sobrevivência e prosperidade. Neste contexto, inserir o consumidor no âmago da organização e direcionar as atenções e estratégias para ele tem papel fundamental em agregar valor e criar vantagens competitivas. Assim, a área de CRM (*Customer Relationship Management*) dedica-se de maneira perseverante à excelência nas tratativas com o consumidor, desde satisfação, retenção, fidelização e lucratividade até compreender e prever suas necessidades e atuando em comunicações, ofertas, produtos e relacionamentos. Para que seja possível obter êxito da atuação de CRM, é essencial o emprego de tecnologias, inteligência analítica e atuação integrada dentro das organizações. Compreender e entender o cliente neste sentido está pautado nas informações disponíveis sobre ele e atualmente o crescimento de dados vem se tornando cada vez mais presentes, possibilitando uma maior imersão na jornada em busca de satisfazê-lo. Tais dados podem ser desde a idade dos clientes até características de compras como nas variáveis RFM (*recency, frequency, monetary*), ou valor, recência e frequência. Sendo assim, a ciência de dados então é o alicerce que irá pautar essas análises e utilizar as informações dos consumidores de maneira que concretize aplicações dentro das organizações. Dentro da ciência de dados, a modelagem de dados é objeto que sintetiza as características da inteligência analítica, podendo ser de previsão, classificação, clusterização, entre outras. Em especial a clusterização é bastante utilizada pois é eficiente em demonstrar as características dos consumidores e então segmentá-los de forma que sejam possíveis tomar diversas decisões no âmbito organizacional. Para a proposta do trabalho, a ciência de dados foi aplicada ao campo de CRM utilizando variáveis da análise RFM e comparando técnicas de agrupamentos para demonstrar como a abordagem *fuzzy* possui potencial de incrementar os objetivos de CRM.

Palavras chave: CRM; ciência de dados; RFM; clusterização; fuzzy.

ABSTRACT

Currently organizations are allocated in a highly competitive scenario and adaptation is necessary in order to prosper and survive. In this context, having the customers as the main purpose of all actions and strategies in the company has a fundamental role to create competitive advantages and more value to the customers. CRM (Customer Relationship Management) is the area inside organizations devoted to pursuit the best connection with the client in terms of loyalty, satisfaction and monetization by predicting and understanding customer´s needs and taking actions by communications, offers, products and relationships. In order to obtain the major goals in CRM, technology, analytics and integration between areas are essential. Having a deep understanding of customers is based on all the information available and currently this information grows larger each day with advances in technology, making it possible to get truly comprehend the customer. This information has different types, such as age or even the variables in RFM analysis, with that being recency, frequency and monetary. To analyze all this information and transform them into actions for the company, data science is one of the most important tools and it can be used as models such as prediction, classification, clustering and much more. Clustering is gaining notoriety since it is an efficient way to demonstrate customer features and put them in groups that can be used to take decisions in the organizations. For this paper, data science was applied in the field of CRM using RFM features and clustering techniques to demonstrate that fuzzy logic has a great potential to develop goals in CRM.

Keywords: CRM; data science; RFM; clustering; fuzzy.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

CRM - *Customer Relationship Management*

RFM - Recency, frequency and monetary

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO

A importância de inserir o consumidor como núcleo das estratégias organizacionais é fundamental para qualquer organização. O desenvolvimento tecnológico, a disseminação intensa de informações e a constante mudança das interações sociais e culturais refletem diretamente na crescente versatilidade dos hábitos de consumo. A adaptação das organizações para tal contexto torna-se característica essencial em busca de vantagem competitiva.

Para que o consumidor esteja realmente no âmago da organização, é necessário compreender e antecipar suas necessidades, interesses e entendê-lo de uma maneira integral. Para tanto, CRM, sigla do inglês *Customer Relationship Management*, tem o papel crucial de ser um dos alicerces destas funções. Descrever, analisar e incorporar a filosofia da visão centrada no consumidor transita por CRM por meio de práticas, tecnologias e estratégias que colocam o cliente como orientador de todas as atribuições.

O êxito da atuação de CRM no contexto organizacional tange a prosperidade da relação com o consumidor, a fidelização e retenção de clientes e a maior lucratividade. Os indicadores para essas ações passam desde critérios como a satisfação dos clientes ao aumento de vendas de produtos.

Inerente a qualquer que seja a métrica de avaliação, a visão analítica de dados dos consumidores é a base para atingir os objetivos de CRM. Para tanto, a congruência com o desenvolvimento e utilização da área tecnológica é essencial. Com o emprego das tecnologias, a criação de um banco de dados bem estruturado, com boa conectividade, automação e sinergia dos sistemas e processos é um dos caminhos para garantir a captura de diversas informações dos consumidores e assim poder compreendê-lo de uma forma holística. A grande quantidade de dados disponíveis e capturados atualmente são um reflexo dos crescentes avanços tecnológicos e a necessidade de obter o máximo de conhecimento sobre os clientes para criar vantagens competitivas. Não obstante, a extração de valor desses dados para gerar

valor competitivo para empresa torna-se o fator que irá impulsionar e alcançar os objetivos da estratégia de CRM. Nesse contexto, a utilização de técnicas analíticas e a formação de profissionais focados na ciência de dados são imprescindíveis.

Um tipo de análise específica que transcreve os dados como predileções em relação à produtos, lojas e até mesmo proximidades com o poder de compra de cada um dos consumidores é a análise *RFM (recency, frequency, monetary),* ou seja, recência, frequência e valor. Estas variáveis foram inicialmente utilizadas na antiga indústria dos catálogos e em ações de marketing direto para representar e identificar comportamento dos consumidores e aumentar as vendas de forma eficiente e eficaz.

Aliar o uso da tecnologia, dados e capacidade analítica são características fundamentais da ciência de dados, que abrange a aplicação de técnicas tanto nas abordagens quantitativas quanto qualitativas para resolver problemas relevantes dentro da organização e criar modelos que auxiliarão na tomada de decisão. Utilizando os diversos softwares e ferramentas, as abordagens de modelagem de dados são: associação, classificação, clusterização ou agrupamento, previsão, regressão e visualização entre outras. A clusterização em especial trata do processo de agrupamento de objetos em grupos homogêneos. Sua utilização vem ganhando espaço pois é um processo rápido e eficiente de caracterizar e evidenciar as características dos clientes com o intuito de direcionar ações específicas.

Assim, este trabalho pretende apresentar a relevância do emprego da ciência de dados, em especial com o uso de técnicas de agrupamento, para obter êxito dos objetivos de CRM. Para tanto, serão descritas as aplicações e importância do uso de dados do profissional que atua com ciência de dados e, em específico, exemplificar a utilização de técnicas de agrupamento no contexto organizacional com a utilização da análise *RFM* em CRM.

1.1 Objetivos

O objetivo principal deste trabalho é mostrar a importância da análise de dados na tomada de decisão, em especial, com a aplicação de técnicas de agrupamento em Customer Relationship Management (CRM).

Os objetivos específicos são:

1) Levantar a relevância do profissional e da área de ciência de dados no panorama organizacional, com foco específico em CRM.
2) Avaliar a contribuição da técnica de agrupamento *fuzzy*, uma extensão da análise tradicional, para auxiliar no processo decisório na área de CRM.

1.2 Justificativa

O dinamismo do cenário em que as empresas estão inseridas é incessante e adaptar-se é essencial para sua sobrevivência e prosperidade. Assim, a competitividade tornou-se mais acirrada e presente no cotidiano de qualquer empresa. Além disso, altamente alavancado pelo desenvolvimento de tecnologias e junto com as alterações no mercado, o consumidor também passou por diversas mudanças nos seus hábitos de consumo, na interação social e nas relações com as empresas. Dado que diversas estratégias organizacionais colocam o consumidor como objeto central e mais valioso para organização, a gestão do relacionamento da empresa com o consumidor tornou-se fundamental. Nesse contexto, o papel de CRM foi amplamente aprofundado e desenvolvido, com sua importância colocada em altos patamares nas grandes organizações. Para que os objetivos de CRM sejam alcançados, é essencial o emprego de tecnologias e aplicação de técnicas de ciência de dados na tomada de decisão.

Uma das técnicas relevantes em CRM visa a formação de grupos, em específico a análise de agrupamento *fuzzy*, que tem a finalidade pautada no agrupamento de objetos similares nos mesmos grupos, sendo que objetos dissimilares devem pertencer a grupos diferentes. Uma das características mais relevantes do *fuzzy* em comparação

com outras técnicas de agrupamentos tradicionais, como o *k-means,* é a possibilidade de determinar o grau de pertencimento dos objetos em relação a cada grupo (SOLTANI & NAVIMIPOUR, 2016).

Dado que a proposta do algoritmo *fuzzy* proporciona uma nova ótica de agrupamento comparada aos métodos convencionais, a aplicação da técnica em áreas específicas como CRM trazem um novo panorama na tomada de decisão no âmbito organizacional. Portanto, a comparação da técnica de agrupamento *fuzzy* com a técnica convencional *k-means* torna-se relevante para expor as vantagens e desvantagens da aplicação dessas ferramentas sob a óptica de tomada de decisão em CRM.

CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA

O referencial teórico abordará três tópicos a serem fundamentados, o primeiro será CRM levantando as suas características gerais, seu desenvolvimento e relevância no contexto organizacional. O segundo levantará os conceitos relativos à ciência de dados em CRM. Por fim, uma seção sobre as Técnicas de Agrupamento evidenciando seus atributos e abrindo caminho para uma aplicação específica com foco em CRM.

2.1 Customer Relationship Management (CRM)

2.1.1 Histórico

A partir de um breve resgate histórico, ressaltar como as interações dos produtores e consumidores influenciaram na concepção de *Customer Relationship Management* e sua relevância no âmbito organizacional, principalmente em relação às vantagens competitivas propostas a partir da visão do consumidor como ponto central da organização (HOSSEINI et. al, 2010).

Inicialmente, entre os séculos XIX e começo do século XX, o cliente era considerado o âmago da organização. A relação entre a produção comercial e o consumidor era muito próxima, os produtos e as vendas eram feitas de forma próxima, individualista, baseados em relações em que o comerciante conhecia as predileções, necessidades e rotinas de seus compradores. Já em meados do século XX, essa relação foi modificada com a introdução da produção e comercialização em massa, deslocando o foco no cliente para um foco no produto, cujas estratégias organizacionais buscavam resultados expressivos em produção.

A transição substancial que colocou novamente o consumidor como núcleo das estratégias organizacionais veio com o desenvolvimento tecnológico, a globalização e maior dinamismo do mercado competitivo, no final do século XX e início do século XXI. Com critérios de fabricação, preços e qualidades equiparáveis na esfera da produção e oferta de bens e serviços, obter uma visão holística sobre o consumidor tornou-se o

critério fundamental para obter vantagem competitiva. Sendo assim, a importância de CRM nas abordagens das principais estratégias organizacionais tornou-se cada vez mais relevante (CORREIA et al, 2005).

2.1.2 Conceitos e características

O planejamento de estratégias de negócios e tomada de decisões empresariais são fundamentalmente baseadas em conhecimentos e relações com os clientes. A partir da interação com as informações que obtidas dos clientes e a maior proximidade com este consumidor, é possível antever seus desejos, aliando sua satisfação e lealdade à empresa com a possibilidade de criar um vínculo que pode potencializar a próxima compra, resultando em um planejamento futuro para o negócio e maior rentabilidade. Nesse contexto, as estratégias pautadas em CRM são imprescindíveis (PEPPER & ROGERS, 2004).

Por mais que a importância de CRM no âmbito de estratégias organizacionais tenha se tornado extremamente reconhecida, não existe uma definição concreta e definitiva de CRM (LING & YEN, 2001, apud NGAIL, 2008). Para Swift (2001, p. 12, apud NGAIL, 2008), CRM pode ser definido como uma abordagem empresarial que visa entender e influenciar o comportamento do cliente a partir de comunicações relevantes, com o intuito de desenvolver a aquisição, retenção, lealdade e, finalmente, a lucratividade do cliente.

Já na visão de Kincaid (2003, p. 41, apud NGAIL, 2008), CRM é o uso estratégico da informação, processos, tecnologia e pessoas com o objetivo de gerenciar a relação do cliente com a organização em todo ciclo de vida do cliente. Para Parvatiyar e Sheth (2001, p. 5, apud NGAIL, 2008) CRM é concernente à estratégia e ao processo e estratégia de adquirir, reter e criar relações sustentáveis com consumidores para agregar valor para empresa e para o próprio consumidor. Isto envolve a integração das funções de marketing, vendas, serviços ao consumidor e cadeia de suprimentos dentro das organizações, objetivando maior eficácia e efetividade na agregação de valor ao consumidor.

Adicionalmente, do ponto de vista empresarial, a possibilidade de influenciar o comportamento do cliente a partir de comunicações mais assertivas e contextualizadas transforma CRM em uma ferramenta poderosa em questões de retenção e lealdade de clientes e na lucratividade da empresa (SWIFT, 2001 apud MONTEIRO, 2015).

Para Pepper & Rogers (2004, p. 59), as aplicações e o termo CRM pode ser definido como:

> CRM é uma estratégia de negócio voltada ao entendimento e à antecipação das necessidades dos clientes atuais e potenciais de uma empresa. Do ponto de vista tecnológico, CRM envolve capturar os dados do cliente ao longo de toda a empresa, consolidar todos os dados capturados interna e externamente em um banco de dados central, analisar os dados consolidados, distribuir resultados dessa análise aos vários pontos de contato com o cliente e usar essa informação ao interagir com o cliente através de qualquer ponto de contato com a empresa.

Dado a contextualização, fica evidente que os atributos de referência de CRM, como a identificação de necessidades do cliente e a manutenção de sua satisfação, são considerados um diferencial competitivo para as empresas no cenário globalizado (ALVES, E. et al, 2014, p.9).

Segundo Xu e Walton (2005, apud BARRETTO, 2007) a gestão do relacionamento com o cliente reflete em vantagens competitivas à empresa porque traduz o objetivo de foco no cliente em conhecimento organizacional.

Para Brown (2000, apud HOSSEINI et. al, 2009) CRM é a chave de uma estratégia competitiva, focando nas necessidades dos clientes e na integração do mesmo com a organização. De acordo com o estudo de Feinberg e Kadam (2002, apud HOSSEINI, 2009), as receitas aumentam entre 25-80% quando a retenção de clientes aumenta em 5 pontos percentuais.

Berry e Linoff (2004) destacam que a combinação da estratégia de CRM com tecnologia e ciência de dados possibilita diversas aplicações comerciais com alto impacto em vantagens competitivas, tais como em sistemas de recomendações de produtos, *cross selling*, sugestões de cupons e até mesmo na redução do risco de crédito em instituições financeiras.

Adicionalmente, as organizações que utilizam uma estratégia de CRM bem desenvolvida e consolidada têm apurado seus benefícios a partir do aumento no número de clientes e do *feedback* positivo em relação à fidelização e satisfação dos consumidores (CORREIA et al, 2005).Sendo assim, a tecnologia é um dos maiores aliados em qualquer aplicação de CRM, tendo um papel essencial em estratégias bem sucedidas. Essa agregação facilita o fluxo de comunicação com os clientes e potencializa o processo de análise de dados, auxiliando na identificação e compreensão das necessidades dos consumidores, transformando dados em conhecimento e por fim em ações estratégicas. (JOSIASSEN et al., 2014 apud SOLTANI e NAVIMIPOUR, 2016).

2.1.3 Análise *RFM*

Como aponta Baier (2002, apud RAJEH et. al, 2014), o estudo e análise das variáveis RFM (*recency, frequency* e *monetary*), ou seja, recência, frequência e valor são utilizadas por muitas décadas em estudos de marketing direto para representar e identificar comportamento dos consumidores. As informações são caracterizadas por: a recência remete à proximidade, medida em dias ou meses, da data da compra em relação à data estipulada na análise. A frequência demonstra o número de transações realizadas no período avaliado. Por fim, o valor é o preço pago nas transações. Com tais informações, é possível extrair diversas informações dos clientes, como predileções em relação a produtos, lojas e até mesmo obter proximidades com o poder de compra de cada um dos consumidores.

Inicialmente utilizado pela indústria de catálogos, RFM foi descrito como um método de segmentação para aumentar vendas de uma maneira relativamente simples, combinando as variáveis ou utilizando-as separadamente para uma base de clientes já existente. Entretanto, com o desenvolvimento de tal análise, suas aplicações foram crescendo e pesquisas de Rud (2001) demonstram que a empresa SAM (Southern Area Merchants) estava disposta a aumentar sua base de clientes e utilizou a análise de RFM para criar a maneira mais eficiente e eficaz. A partir da análise das variáveis

separadamente, a organização percebeu que a recência era o fator mais importante para mensuração da métrica de penetração da marca, ou seja, com a variável foi possível identificar os clientes que mais faziam compras e aceitavam as propostas de marketing direto. Sendo assim, foi realizado outro estudo olhando especificamente a variável de idade combinada com a variável de recência, com o intuito de criar grupos de novos clientes para ação de marketing direto. Por fim, foi possível determinar, a partir da análise RFM, a melhor idade para categorizar e criar os melhores grupos para ação obter novos clientes.

Tendo como base a retenção e fidelização de clientes nos objetivos centrais de CRM, Liu & Shih (2005, apud HOSSEINI et al, 2009) destacam que as variáveis *RFM* são parâmetros utilizados para classificar os clientes em termos de maior importância para organização em termos direcionamento de estratégia de vendas, marketing, comunicação, entre outras.

2.2 Ciência de dados

A ciência de dados é frequentemente associada a vários temas que envolvem as áreas de estatística e computação, tais como, *big data*, mineração de dados, aprendizado de máquina, inteligência artificial e *advanced analytics*. Dado a crescente relevância desses temas no cenário acadêmico e organizacional pela solução de problemas de tomada de decisão nos mais diversos contextos e, em específico, em CRM, que é a área foco de aplicação das técnicas neste trabalho (CAO, 2017).

A disponibilidade de cada vez mais dados no cenário globalizado impulsionou a chamada era do *big data*. Tal termo refere-se a grande volumes de dados, como por exemplo os 2 bilhões de vídeos assistidos por dia no Youtube ou, em 2011, foram feitas mais de 5 bilhões de buscas no Google por dia. Com as novas tecnologias de hardware e software, as organizações estão cada vez mais capacitadas em coletar e analisar esses dados, sendo possível processar e interpretar seus resultados (DAVENPORT; HO-KIM, 2014).

Por mais desafiador que seja a adaptação a esta era revolucionária, as oportunidades e possibilidades de inovação de mudanças em modelos de negócios são abrangentes. Esse novo panorama não se limita simplesmente ao uso dos dados, mas sim aos aspectos de transformação e criação de novos conhecimentos possíveis a partir da utilização e exploração desses dados (CAO, 2017).

Para Waller e Fawcett (2013), a ciência de dados abrange a aplicação de técnicas tanto nas abordagens quantitativas quanto qualitativas para resolver problemas relevantes dentro da organização e criar modelos preditivos que auxiliarão na tomada de decisão. Sendo a extração de conhecimento e informações a partir de dados é a base fundamental para contextualizar a ciência de dados a partir da combinação de coleta, manipulação, visualização e gestão de grandes quantidades de dados (PAIXÃO et al, 2015).

A relevância da ciência de dados também deve ser notada pela grande abrangência de áreas em que pode auxiliar no desenvolvimento e tomada de decisão, como marketing, vendas, suporte aos clientes e até mesmo em advocacia e astronomia. Entretanto, a utilização das informações depende dos tipos de dados disponíveis, do objetivo a ser alcançado e do conhecimento do indivíduo (BERRY et al, 2004).

O objetivo de entender cada cliente individualmente e com isso ajustar decisões referentes ao valor de mantê-los é extremamente relevante para as organizações. Para garantir tais objetivos, o uso extenso de dados é colocado como orientação fundamental no processo de notar o que os clientes estão fazendo, lembrar o que eles fazem e fizeram ao longo do tempo, aprender com essas memórias e por fim atuar com base nessas informações com o intuito de rentabilização (BERRY; LINOFF, 2004).

Para Correia (2005), a manipulação desses dados e sua transformação em informação relevante é um dos delimitadores de critérios de sucesso para área de CRM pois a captura de dados são fatores que irão criar os alicerces das visões analíticas sobre os consumidores. Para que a estruturação e transformação desses dados ocorra, é necessário alinhar o papel fundamental de ferramentas tecnológicas e softwares

especializados ao conhecimento dos profissionais envolvidos nas análises de CRM (JUNIOR 2011, apud MONTEIRO, 2015).

Já o cientista de dados, ou seja, o profissional da ciência de dados, necessita ter a perspectiva dos diversos problemas de tomada de decisão pela óptica dos dados. Sendo assim, o profundo conhecimento de técnicas analíticas, sejam elas matemáticas ou estatísticas, devem estar alinhados com o conhecimento em relação aos dados disponíveis, às tecnologias disponíveis e aos tipos de desafios que serão encarados (WALLER E FAWCETT, 2013).

Além do conhecimento profundo sobre a análise de dados, os modelos aplicados e as possíveis conclusões, a apresentação dos resultados também possui um papel importantíssimo a ser considerado para o cientista de dados, pois quanto mais clara for a demonstração das conclusões, maior a probabilidade do direcionamento que a análise propôs guie decisões e ações assertivas. Sendo assim, o profissional também deve-se atentar à questão de “contar a história dos dados”. Para tanto, a utilização de gráficos dinâmicos, aproximação das técnicas com o público a partir da gamificação são exemplos de como alcançar tal objetivo (DAVENPORT & HO-KIM, 2014).

Podemos então destacar que a ciência de dados aplicada no âmbito organizacional é relevante. De acordo com o estudo de McAfee e Brynjolfsson (2012, apud Waller e Fawcett, 2013), quanto mais as empresas se denominam tomadoras de decisão guiadas por dados, melhor o seu desempenho em áreas financeiras e operacionais. Adicionalmente, um terço das melhores empresas com essas características são, em média, 5% mais produtivas e 6% mais lucrativas que seus competidores.

Quando o assunto é melhoria na tomada de decisão, Davenport & Ho-kim (2014) destaca que pesquisas sobre como as organizações conseguiram melhorar 57 decisões diferentes, a ciência analítica estava em primeiro lugar como caminho para tal aprimoramento. Também nesse contexto, o tópico “melhores dados” apareceu em terceiro lugar na classificação de importância. Suas pesquisas destacam como a empresa Microsoft utiliza da inteligência analítica a partir da ciência de dados para

customizar as ofertas do seu motor de busca do “Bing”. Disposta de variáveis como idade, localização e histórico de buscas, a Microsoft criou modelos preditivos que otimizam a oferta de busca para cada cliente. Em relatos, os grandes *stakeholders* descreveram tal método como bastante eficaz para aumentar vendas.

A empresa Google utilizou a ciência de dados para traços de empregados com maior inclinação à alta performance. Diferente do que a intuição dos analistas indicava, os fatores como altas notas escolares e avaliações boas de entrevistas não estavam sendo bons previsores do desempenho. Para mitigar tais problemas, foi levantado diversas questões e a partir de correlações muitas vezes inesperadas, como o alto desempenho estar ligado com criação de empresas sem fins lucrativos, foi possível aumentar muito a assertividade dos modelos para previsão de alta performance (DAVENPORT; HO-KIM, 2014).

2.3 Técnicas de agrupamentos

As técnicas de ciência de dados ganharam espaço e popularidade pois possibilitam estabelecer relações entre os consumidores e as empresas, auxiliando no processo de tomada de decisão e solidificando posições estratégicas de vantagem competitiva, em especial com foco em CRM (BAHARI & ELAYIDOM, 2015).

Para Turban, Aronson, Liang e Sharda (2007, p.305, apud NGAI et. al, 2008) ciência de dados trata principalmente de modelos matemáticos, estatísticos ou de inteligência artificial e aprendizado de máquina para extrair e identificar informações relevantes e obter conhecimentos a partir de grandes bases de dados. Essas técnicas, segundo o seu objetivo, podem executar um ou mais dos seguintes tipos de modelagem de dados: associação, classificação, clusterização ou agrupamento, previsão, regressão e visualização entre outras.

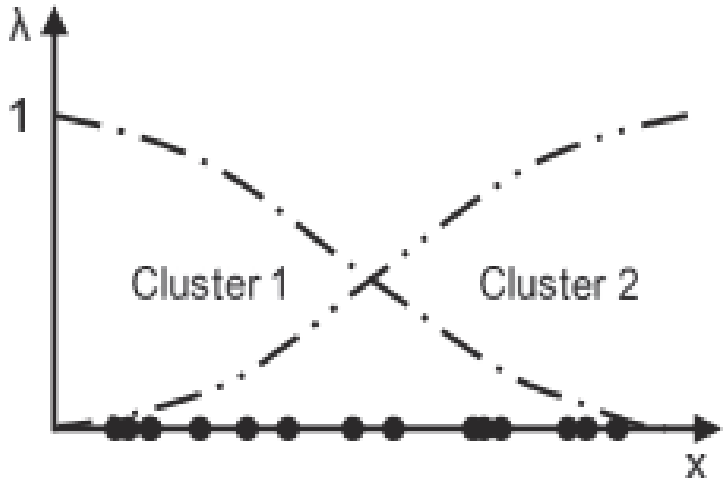
Partindo para um olhar mais profundo sobre a clusterização, ela trata do processo de agrupamento de objetos em grupos homogêneos. Um cluster pode ser definido como a união de dados ou objetos que são similares dentro de seu próprio

cluster e dissimilares aos objetos em outros clusters (Han & Kamber, 2001 apud NGAI et. al, 2008).

Para Kucukkancabas (2007, apud RAJEH et. al, 2014), uma das maneiras mais rápidas para se construir uma estratégia de CRM bem sucedida é dividindo os consumidores em segmentos ou grupos para que seja possível identificar os clientes mais rentáveis. Adicionalmente, separar os consumidores em grupos significativos e homogêneos baseados em suas características é uma ferramenta que possibilita que as organizações construam diferentes estratégias, sempre guiados pelos atributos dos consumidores.

Neste trabalho se propõe a apresentar a técnica de agrupamento *fuzzy*, a qual os objetos são designados ao clusters com nível de grau de pertencimento, tal característica difere o *fuzzy* de algoritmos tradicionais, como o k-means, em que cada objeto pertença exclusivamente à um cluster. Sendo assim, o objetivo da técnica é de encontrar os grupos que minimizem o peso da soma das distâncias euclidianas entre os objetos e a centróide de cada cluster (PETERS et al, 2012). O gráfico a seguir ilustra o pertencimento dos objetos aos clusters, dada a ideia proposta com uso da técnica de agrupamento *fuzzy*.

Figura 1 - Representação do grau de pertencimento na técnica *fuzzy c-means*



Fonte: G. Peters et al./International Journal of Approximate Reasoning 54 (2012, p. 307-322)

CAPÍTULO 3 – ASPECTOS METODOLÓGICOS

3.1 Caracterização da pesquisa

Com base nas considerações iniciais, foi definido para esse trabalho o caráter exploratório quantitativo. De acordo com Gil (2002), trata-se de uma pesquisa de âmbito exploratória aquela que visa o aprimoramento de ideias ou a descoberta de intuições. O atributo que define a questão quantitativa da pesquisa é o fator primordial da análise de dados observados de natureza quantitativa, que retorna a avaliação de interações entre variáveis e utiliza técnicas estatísticas para análises. Dado isso, para alcançar o objetivo do estudo serão utilizados dados secundários com intuito de realizar as análises que colaborem para o desenvolvimento da pesquisa.

3.2 Plano de análise dos dados

Com o intuito de exemplificar a aplicação de ciência de dados no âmbito organizacional, em específico em CRM, o trabalho utilizou uma base de dados secundários, ou seja, informações já coletadas e públicas referentes às transações online em uma empresa de varejo. A base contém informações da data da transação, descrição do item, identificação do comprador, valor da compra e a identificação do item de compra. A base de dados está disponível para consulta em: https://archive.ics.uci.edu/ml/datasets/Online+Retail#.

Com a disposição dos dados, foi possível criar as variáveis de recência, frequência e valor. Dado a relevância da análise dessas variáveis para os objetivos de CRM, essas informações foram tratadas para aplicação de técnicas de agrupamento tradicionais (k-means) e de agrupamento *fuzzy*, para avaliação da sua contribuição.

Para exploração, tratamentos, criação de modelos e análise de dados foi utilizado o software livre de análise de dados Anaconda, que possibilita o uso das linguagens de programação Python e R para exploração de pacotes para análise de

dados e criação de modelos estatísticos. As ferramentas são livres e estão disponíveis para *download* em:

https://www.anaconda.com/distribution/
https://www.python.org/downloads/
https://cran.r-project.org/bin/windows/base/

As linguagens de programação livre ganharam grande importância no cenário atual pois permitem a democratização do acesso à programação e principalmente por serem ferramentas que permitem a aplicação de técnicas de ciência de dados para a tomada de decisão. Adicionalmente, cada vez mais há cursos disponíveis em diversas plataformas online e presenciais para aprendizado do seu uso (CAO, 2017).

A partir dessa base de dados e do uso das ferramentas de linguagem de programação, foi feita a análise exploratória dos dados e a aplicação das técnicas de agrupamento *k-means* e *fuzzy c-means*. Por fim, foi feita a comparação das duas técnicas evidenciando as vantagens e desvantagens entre ambas sob uma ótica de tomada de decisão específica em CRM.

CAPÍTULO 4 - APRESENTAÇÃO DE RESULTADOS

A análise da base de dados a seguir foi realizada em 3 passos primordiais, no primeiro foi feita a estruturação da base que consistiu na manipulação e exploração dos dados, que irá garantir a preparação das informações para as aplicações das técnicas. Em segundo lugar, houve a aplicação das técnicas de agrupamento na base estruturada, com o intuito de observar as características dos dados e direcionar a análise para os objetivos de CRM. Por fim, foram extraídas as conclusões da execução das técnicas de agrupamentos e feita a comparação sob a óptica dos impactos em CRM.

4.1 Estruturação dos dados

A disposição da base de dados não permite a aplicação imediata de técnicas de agrupamento, portanto a manipulação dos dados é essencial para dar continuidade nas análises. A figura 2 mostra as primeiras linhas da base com a estrutura inicial:

Figura 2 – Estrutura inicial com as primeiras linhas da base de dados

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	2010-12-01 08:26:00	7.65	17850.0	United Kingdom
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	2010-12-01 08:26:00	4.25	17850.0	United Kingdom
7	536366	22633	HAND WARMER UNION JACK	6	2010-12-01 08:28:00	1.85	17850.0	United Kingdom
8	536366	22632	HAND WARMER RED POLKA DOT	6	2010-12-01 08:28:00	1.85	17850.0	United Kingdom
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	2010-12-01 08:34:00	1.69	13047.0	United Kingdom
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	2010-12-01 08:34:00	2.10	13047.0	United Kingdom
11	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	2010-12-01 08:34:00	2.10	13047.0	United Kingdom
12	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	2010-12-01 08:34:00	3.75	13047.0	United Kingdom

Fonte: Elaboração Própria

A manipulação inicial das informações consiste na criação das variáveis de frequência, recência e valor conforme seção de fundamentação. A figura 3 apresenta o código da programação Python para criação de tais variáveis e a mudança na base de dados após criação e tratamentos das variáveis (valor, recência e frequência).

Figura 3 – Tratamento e criação das variáveis *RFM*

```
# Transformar variáveis de tempo para criação da recência
datetime_object = datetime.strptime('2012-01-01', '%Y-%m-%d')
df_t['maxdate'] = datetime_object

# Criando as funções agregadas para criação da quantidade total de transações e de valores no período
tryme = df_t.groupby("CustomerID", as_index=False).agg(
    {"maxdate": np.max, "Date": np.max, "InvoiceNo": lambda x: x.nunique(), "UnitPrice": np.sum})

# Renomeando variável para frequência e valor
tryme = tryme.rename(index=str, columns={"InvoiceNo": "frequencia","UnitPrice": "valor"})

# Convertendo variáveis para criação da recência
tryme['date_object_c']= tryme['maxdate'].astype(str)
tryme['max_date_c']= tryme['Date'].astype(str)
d1 = tryme['date_object_c'].apply(lambda x: datetime.strptime(x, '%Y-%m-%d'))
d2 = tryme['max_date_c'].apply(lambda x: datetime.strptime(x, '%Y-%m-%d'))
delta = d1 - d2
d3 = delta.dt.days
tryme['recencia'] = d3

# Criando base final para clusterizar
X = tryme[['CustomerID','valor','recencia','frequencia']]
# Amostra da base tratada:
X.head()
```

	CustomerID	valor	recencia	frequencia
0	12346.0	1.04	348	1
1	12347.0	481.21	25	7
2	12348.0	178.71	98	4
3	12349.0	605.10	41	1
4	12350.0	65.30	333	1

Fonte: Elaboração Própria

O código de programação apresentado na figura 3 segue os seguintes passos:

1) Criar as variáveis que irão determinar o período a ser analisado para a variável de recência, separadas por dia, mês e ano;
2) Fazer a soma das quantidades de transações e valores gastos no período, para cada um dos consumidores;

3) Renomear as variáveis de frequência e valor criadas no passo 2;
4) Realizar o cálculo da diferença entre a última data de transação do cliente e o período máximo estipulado para criar a variável de recência, medida em quantidade de dias;
5) Manter as variáveis em uma base de dados com os consumidores, valores, frequências e recência;
6) Demonstrar os primeiros registros da base para validação das informações.

Na tabela 1, foi realizada a análise descritiva com a qual observa-se a partir da média e o desvio padrão que as variáveis apresentam grandes diferenças em relação à sua escala. Como a técnica de agrupamento procura maximizar a similaridade dos objetos dentro de um mesmo cluster, é fundamental que as variáveis utilizadas estejam em uma escala em que sejam equiparáveis, com o intuito de que não exista grande discrepância em relação às suas importâncias na formação dos grupos, ou seja, a influência de cada variável deve ser equiparável.

Tabela 1 – Análise descritiva das variáveis

	valor	recencia	frequencia
contagem	4275	4275	4275
média	233.18	116.21	3.66
desvio padrão	306.13	100.11	3.90
mínimo	0	23	1
25%	505	41	1
50%	121	74	2
75%	285	167	4
máximo	2354	396	27

Fonte: Elaboração Própria

Dada a necessidade da padronização das variáveis, visto que as diferenças nas distribuições das variáveis fariam com que as informações não fossem consideradas com mesma importância na formação dos grupos, foi considerada a normalização

dessas variáveis. A padronização escolhida foi realizada com o objetivo de que todas as variáveis ficassem com média zero (0) e o desvio padrão igual a um (1). A tabela 2 apresenta a nova distribuição das variáveis após este cálculo

Tabela 2 – Análise descritiva das variáveis normalizadas

	valor	recencia	frequencia
contagem	4,275	4,275	4,275
média	0.000	0.000	0.000
desvio padrão	1.000	1.000	1.000
mínimo	-0.762	-0.931	-0.681
25%	-0.596	-0.751	-0.681
50%	-0.356	-0.422	-0.425
75%	0.162	0.507	0.086
máximo	6.932	2.795	5.970

Fonte: Elaboração Própria

Após o tratamento dos dados, a próxima seção apresenta a aplicação das técnicas de agrupamento tradicionais (k-means e fuzzy).

4.2 Aplicação das técnicas de agrupamento

Nesta seção é apresentado os resultados da aplicação das duas técnicas de agrupamento. Inicialmente, foi feita a padronização das variáveis, visto que as distribuições das variáveis não são homogêneas. Com a padronização todas as informações são consideradas com a mesma importância na formação dos grupos.

4.2.1 Aplicação da técnica de agrupamento K means

Após a padronização das variáveis, a aplicação das técnicas de agrupamento é possível. A primeira aplicação foi da técnica *k-means.* Entre muitos modelos de agrupamento, o algoritmo *K-means* é uma das técnicas mais utilizadas. O “K” em seu nome refere-se à propriedade fixa de número de clusters, definidos nos termos da proximidade das variáveis. O número de variáveis independentes utilizadas na clusterização pode variar segundo os interesses do estudo (BERRY; LINOFF, 2004).

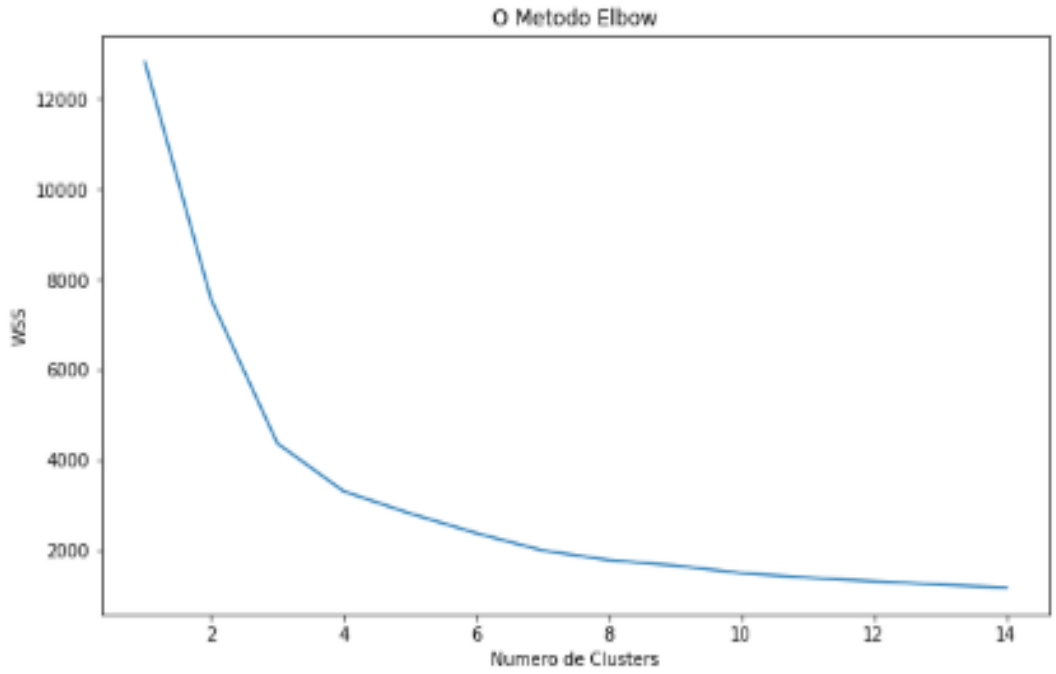
Na publicação de J. B. MacQueen (1967, apud Berry e Linoff, 2004), são definidos três passos para o algoritmo *K-means.* Inicialmente, são selecionados *K* pontos aleatórios dentro do conjunto de variáveis, esses pontos também são chamados de sementes. O segundo passo consiste em determinar, os pontos mais próximos a cada semente. A proximidade é dada a partir do cálculo da distância entre as sementes e os objetos. Adicionalmente, tais pontos irão estabelecer as fronteiras iniciais dos clusters. Por fim, o terceiro passo baseia-se no cálculo das centróides de cada cluster. Para tal objetivo, é feito o cálculo do valor médio para cada dimensão dos objetos dentro do cluster. Sendo assim, as centróides tornam-se as sementes para a próxima iteração do algoritmo, então o passo dois repete-se, designando cada objeto para a centróide do cluster mais próximo. O processo continua até que as fronteiras dos clusters parem de se modificar.

Assim, um dos principais elementos da modelagem é determinar o número de clusters *k*. Dentre as diversas maneiras de encontrar o número de cluster, foi utilizado o método *elbow*, ou cotovelo, que direciona o número de *k* utilizando como parâmetro a soma dos erros ao quadrado. Dado a tendência de quanto maior o número de *k*, menor a soma ao quadrado dos erros, o intuito da técnica é visualmente demonstrar o número de clusters em que o aumento de *k* não retorna mudanças significantes para a técnica. O processo é exploratório e como mostra a figura 4, entre 4 e 6 cluster as mudanças do parâmetro são cada vez menores, direcionando os testes na base de dados.

Figura 4 – Método *Elbow* para definição do número de clusters

```
# Método Elbow para determinar melhor número de K
from sklearn.cluster import KMeans
wcss = []
num_k = 15

for i in range(1, num_k):
    kmeans = KMeans(n_clusters = i, init = 'random')
    kmeans.fit(scaled_df)
#     print(i,kmeans.inertia_)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, num_k), wcss)
plt.title('O Metodo Elbow')
plt.xlabel('Numero de Clusters')
plt.ylabel('WSS') #within cluster sum of squares
plt.show()
```



Fonte: Elaboração Própria

O código apresentado na figura 4 segue os seguintes passos:

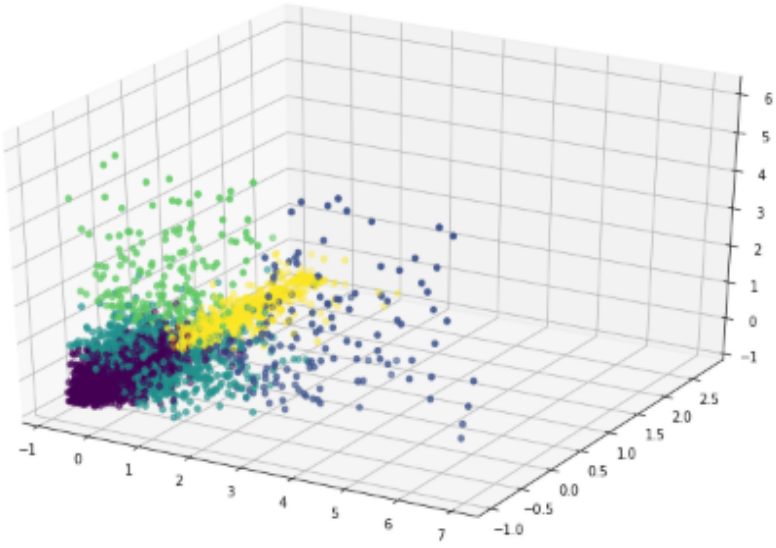
1) Mapear a biblioteca que contém o algoritmo *k-means;*
2) Criar o conjunto vazio “wss” que irá aglutinar as informações das somas das distâncias dos clusters;
3) Definir o número de “k” clusters a serem testados no modelo. Essa informação deverá constar no objeto “num_k”;

4) Criar um *looping* que fará a interação do número de “k” definidos no passo 3;
5) Aplicar a técnica na base de dados partindo do teste inicial de 1 cluster até o número de clusters definidos no passo 3. Cada vez que o modelo treinar com determinado número de “k”, a informação da soma das distâncias será salvo no conjunto “wss”, que será usado para compor o gráfico;
6) Inserir a distribuição no gráfico para visualmente iniciar o processo de exploração da quantidade ideal de “k” para o modelo final.

Após análise exploratória, a técnica foi aplicada e o melhor resultado alinhado aos objetivos de CRM teve o parâmetros de k = 5, ou seja, 5 clusters distintos. Como representado na figura 5, foi possível distribuir os clientes em cluster distintos de acordo com as variáveis:

Figura 5 – Resultados com a técnica de agrupamento *k-means com 5 grupos*

```
# Initializing KMeans
kmeans = KMeans(n_clusters=5, init='random')
# Fitting with inputs
kmeans = kmeans.fit(testando)
# Predicting the clusters
y = kmeans.predict(testando)
# Getting the cluster centers
C = kmeans.cluster_centers_
# Getting the distance
distance = kmeans.fit_transform(testando)
# Getting the labels
labels = kmeans.labels_
# PLotando
fig = plt.figure()
ax = Axes3D(fig)
ax.scatter(testando[:, 0], testando[:, 1], testando[:, 2], c=y)
ax.scatter(C[:, 0], C[:, 1], C[:, 2], marker='', c='#050505', s=1000)
plt.show()
```



Fonte: Elaboração Própria

O código apresentado na figura 5 segue os seguintes passos:

1) Definir os parâmetros do algoritmo *k-means*, sendo o número de “k” e a semente de inicialização;
2) Treinar e aplicar o modelo na base de dados previamente tratada;
3) Determinar as centróides dos clusters, as distâncias e as variáveis para visualização no gráfico;
4) Aplicar no gráfico determinando o domínio das variáveis, as centróides, as cores e formatação.

Por fim, a partir dos 5 grupos formados foi elaborada a tabela 3 com as médias. A análise de cada variável nos clusters representa a característica geral dos clientes dentro do agrupamento, como apresenta a tabela 3. Observa-se que cada cluster possui diferenças relevantes em relação às variáveis da métrica *RFM*

Tabela 3 – Média das variáveis *RFM* dos indivíduos nos 5 clusters obtidos no *k-means*

cluster	valor	frequencia	recencia
0	88.44	1.42	276.61
1	434.26	5.97	54.86
2	111.26	2.12	75.72
3	584.62	14.82	40.04
4	1432.34	12.4	42.96

Fonte: Elaboração Própria

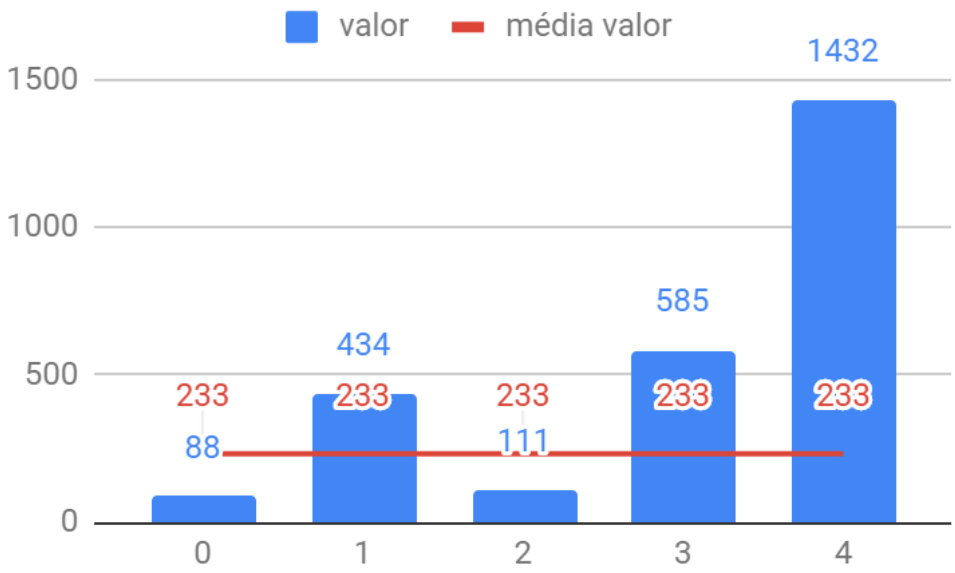
Com tais dados, é possível extrair diversas informações e aplicações possíveis em CRM. Como por exemplo, dado o objetivo de CRM de retenção, percebemos que o cluster “3” representa os clientes que compram com maior frequência na loja, além de também serem os clientes que, em média, comparecem na loja em datas mais recentes à análise. Para tal cluster, a possibilidade de comunicações e ofertas são abrangentes. Outro exemplo de cluster ligado aos objetivos de lucratividade do cliente em CRM está presente no cluster “4”, cujo valor médio da compra é o maior dentre todos os outros agrupamentos, o que remete ao cliente ter uma renda maior e direciona campanhas específicas para sua abordagem.

Usando as variáveis da análise *RFM* como parâmetros de distinção dos clusters, podemos classificar cada agrupamento de acordo com sua relevância em comparação à base total de clientes.

A partir da análise da variável de valor, é possível classificar os clientes comparando com a média de valor gasto em todas as transações no período. A figura 6 apresenta a classificação utilizando a variável valor como métrica de avaliação, sendo assim, é possível observar a relevância do cluster “4”, que gasta em média 614% a

mais que a média geral dos clientes, portanto é possível denominá-los como os clientes mais rentáveis para a organização. Dentro da estratégia de CRM, o objetivo fundamental da lucratividade coloca em evidência a importância deste cluster.

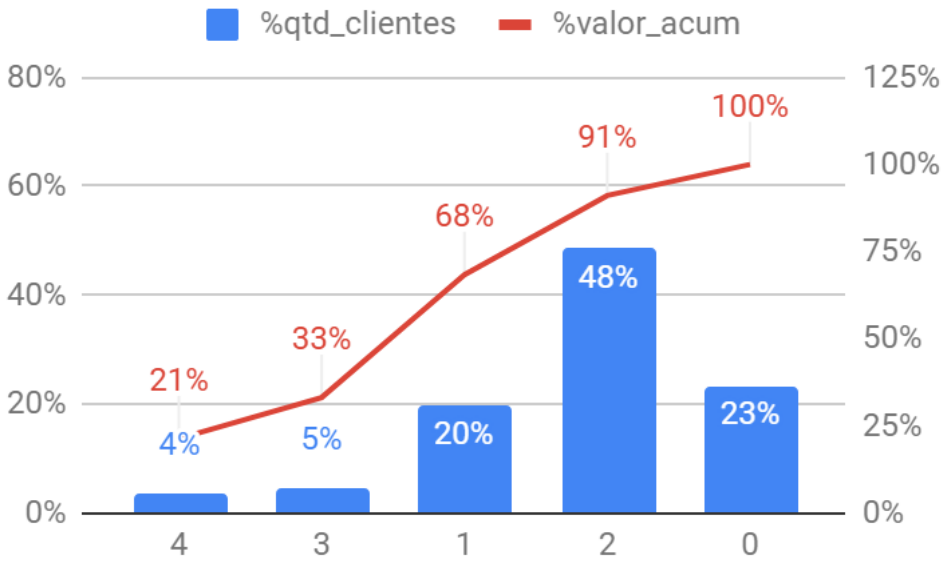
Figura 6 – Classificação da lucratividade dos clusters utilizando a variável valor



Fonte: Elaboração Própria

Adicionalmente, como apresentado na figura 7, ao reorganizar os cluster pela sua classificação de importância em relação à média de valor gasto nas transações, temos que 68% do total de gastos é representado por 29% do total de clientes, ou seja, caso a atuação fosse focada somente nos 29% dos clientes contidos nos clusters “4”, “3” e “1”, já seria possível obter grande parte dos lucros da organização. Quando expandimos essa visão para os clientes que mais gastam em média, que seriam os consumidores do cluster “4”, temos que 4% de todos os clientes garantem 21% dos lucros para a organização, deixando ainda mais em evidência a importância deste cluster e sua classificação em relação à variável de valor.

Figura 7 – Reorganização dos cluster pelo valor médio



Fonte: Elaboração Própria

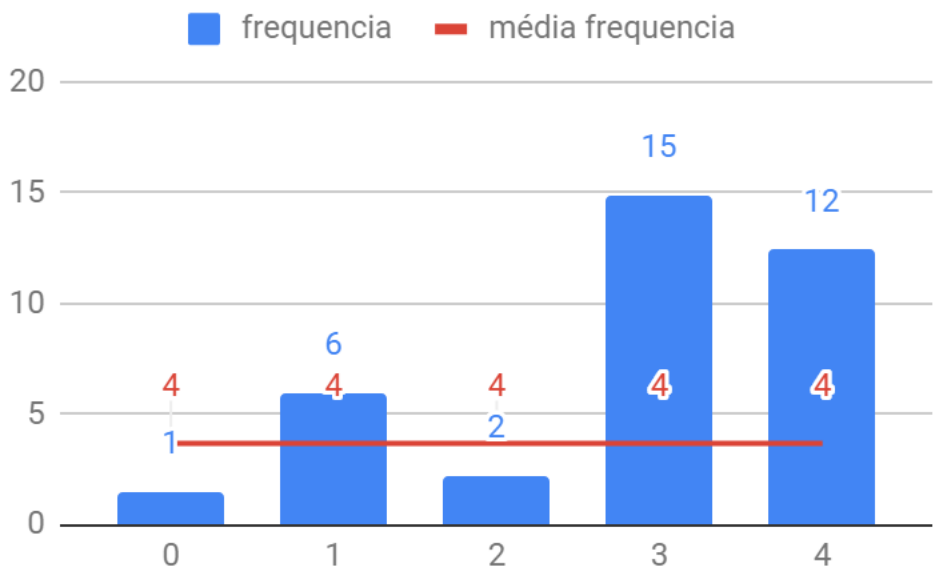
Sob a ótica da variável de recência, clientes que comparecem à organização com maior proximidade em relação ao período estipulado estão em um contato maior com o cotidiano e as mudanças dentro da empresa, ou seja, a lembrança da compra, da experiência e de toda interação feita na transação está presente na memória do cliente. Sendo assim, é importante notar que quanto menor a recência, mais próximo do período estipulado foi feita a transação. A figura 8 apresenta a classificação utilizando a recência como métrica de avaliação, e é possível observar a relevância do cluster “3”, que sua recência é em média 34% menor que a média geral dos clientes, portanto é possível denominá-los como os clientes com as transações mais recentes.

Figura 8 – Classificação da recência dos clusters

Fonte: Elaboração Própria

A análise da variável de frequência representa a quantidade de transações que o cliente realizou no período estipulado. A maior frequência de compras é um indicativo de um objetivo fundamental de CRM, que é a de fidelização de clientes, ou seja, quanto maior a frequência que o consumidor transaciona na organização, mais familiarizado e integrado ele está com a organização. A figura 9 apresenta a classificação a partir da comparação da frequência média dos cluster em relação à média geral dos consumidores, sendo assim, é possível observar o destaque do cluster “3”, dado que sua frequência de transações é em média 405% maior que a média geral dos clientes:

Figura 9 – Classificação da frequência dos clusters



Fonte: Elaboração Própria

Por fim, a partir da análise dos dados, é possível classificar os clusters em relação à sua importância tendo em vista as variáveis de RFM. Para a análise, todas as variáveis possuem o mesmo peso, portanto a tabela 4 representa a classificação final dos clusters utilizando como métrica de avaliação a menor somatória das variáveis de RFM:

Tabela 4 – Classificação dos clusters obtidos no *k-means* após análise das variáveis de RFM

	valor		recencia		frequencia		classificação final	
cluster	média	rank	média	rank	média	rank	soma	rank
0	88	5	277	5	1	5	15	5
1	434	3	55	3	6	3	9	3
2	111	4	76	4	2	4	12	4
3	585	2	40	1	15	1	4	1
4	1,432	1	43	2	12	2	5	2

Fonte: Elaboração Própria

Por fim, a classificação final dos cluster é representada por: cluster “3”, cluster “4”, cluster “1”, cluster “2” e por fim o cluster “0”. Sendo assim, o cluster mais relevante para os objetivos de CRM é representado na técnica como o cluster “3”, pois a combinação da classificação das variáveis de valor, frequência e recência evidenciam a melhor oportunidade de fidelização, rentabilidade e comunicação com esses clientes, constituindo o maior potencial para construção de uma relação sustentável e mais rentável a longo prazo. Em contrapartida, o cluster “0” representa o menor potencial de negócios para a organização, ou seja, o investimento neste tipo de cliente irá trazer o menor resultado em comparação com os outros consumidores.

4.2.2 Aplicação da técnica de agrupamento Fuzzy

Também definida como técnica de agrupamento, foi escolhida para aplicação a clusterização pautada na técnica de *fuzzy c-means.* A técnica tem como diferença essencial em relação às técnicas convencionais como a *k-means* pois ela permite que os objetos tenham um grau de pertencimento ao referido “K”, ou seja, ao cluster. A vantagem de tal característica é de proporcionar a oportunidade de flexibilizar a classificação dos objetos em clusters dado o seu grau de pertencimento (Al-Augby et al, 2014). Assim, um elemento pode ser representado com um grau de pertencer aos grupos.

Para aplicação da técnica e sua comparação com a abordagem do *k-means,* o número de clusters também foi definido em 5. A figura 10 demonstra que para cada observação, existe um grau de pertencimento à todos os clusters

Figura 10 – Técnica de agrupamento *fuzzy c-means*

```
# Treinando o modelo
cntr, u_orig, u0, d, jm, p, fpc = fuzz.cluster.cmeans(
    testando.T, 5, 2, error=0.0005, maxiter=1000)
# Aplicando na base
u_pred, u0_pred, d_pred, jm_pred, p_pred, fpc_pred = fuzz.cmeans_predict(
    testando.T, cntr, 2, error=0.0005,maxiter=1000,init=None,seed=None)

# Cada coluna representa 1 dos 5 clusters, com grau de pertencimento
u_pred.T
```

```
array([[ 0.00846156,  0.02163293,  0.04884995,  0.88418067,  0.03687489],
       [ 0.02001445,  0.68959773,  0.06454208,  0.02261386,  0.20323188],
       [ 0.00573806,  0.03806573,  0.27981253,  0.03090195,  0.64548172],
       ...,
       [ 0.00628908,  0.03084014,  0.70036621,  0.02673544,  0.23576913],
       [ 0.98661994,  0.00697993,  0.00199825,  0.00156621,  0.00283568],
       [ 0.0034948 ,  0.01972672,  0.67166616,  0.01736582,  0.28774649]])
```

Fonte: Elaboração Própria

O código apresentado na figura 10 segue os seguintes passos:

1) Treinar a técnica *fuzzy c-means* definindo a base a ser utilizada, o número de “k” clusters, o vetor exponencial, o critério de erro para parar as iterações e por fim o número máximo de iterações;
2) Inserir a localização das centróides treinadas no modelo para utilização na base e determinar os mesmos parâmetros do modelo treinado.
3) Por fim, garantir que para cada linha treinada existam 5 variáveis, cada uma contendo um grau de pertencimento para cada cluster. No total, a soma das 5 colunas deve somar 1.

Dado as propriedades fundamentais da técnica fuzzy, o grau de pertencimento a cada um dos cluster possibilita ampliar o panorama da aplicação das técnicas de agrupamentos que dependem da estratégia de abordagem ao cliente e necessidade de classificação dentro da organização, ou seja, dado tal flexibilidade é possível criar muitos cenários pautados no valor de corte ao grau de pertencimento. Por exemplo, a tabela 5 apresenta a definição do cluster com maior grau de pertencimento para cada

um dos consumidores, ou seja, caso o maior grau de pertencimento seja relativo à coluna do cluster 3, o consumidor será classificado no cluster 3. Observa-se que cada cluster possui diferenças relevantes em relação às variáveis da métrica *RFM*:

Tabela 5 – Média das variáveis *RFM* dos indivíduos nos 5 clusters obtidos no *fuzzy c-means*

cluster	valor	frequência	recência
1	84.1	1.7	84.0
2	614.7	8.8	49.3
3	260.0	4.4	58.9
4	87.4	1.4	281.9
5	1,291.9	16.6	35.7

Fonte: Elaboração Própria

Todas as análises feitas com base na clusterização do *k-means* são aplicáveis para os agrupamentos retirados do maior grau de pertencimento da técnica *fuzzy c-means*. Entretanto, a flexibilidade da técnica possibilita que sejam marcados os clusters para todos os graus de pertencimento, tendo seu limite estipulado ao número de “k”. Como exemplo, a tabela 6 representa os 5 primeiros dados dos consumidores, para as variáveis de “grau de pertencimento ao clusters”, é possível observar o resultado da aplicação da técnica *fuzzy c-means*. Tendo cada um dos graus de pertencimento, as variáveis de classificação do cluster são criadas a partir da regra de maior número dentre as 5 variáveis, ou seja, caso o consumidor possua o maior grau de pertencimento representado no cluster “4”, o primeiro cluster (cluster_fuzz_1) que ele será classificado será o cluster “4”, caso seu segundo maior grau de pertencimento seja representado no cluster “1”, seu segundo cluster (cluster_fuzz_2) será marcado como cluster “1”:

Tabela 6 – Representação dos dados obtidos com a aplicação da técnica *fuzzy c-means*

ID	grau de pertencimento ao cluster					classificação do cluster, 1 sendo o maior e 5 o menor				
	1	2	3	4	5	cluster_fuzz_1	cluster_fuzz_2	cluster_fuzz_3	cluster_fuzz_4	cluster_fuzz_5
1	0.049	0.022	0.037	0.884	0.008	4	1	3	2	5
2	0.065	0.690	0.203	0.023	0.020	2	3	1	4	5
3	0.280	0.038	0.645	0.031	0.006	3	1	2	4	5
4	0.266	0.214	0.390	0.089	0.041	3	1	2	4	5
5	0.028	0.012	0.021	0.934	0.005	4	1	3	2	5

Fonte: Elaboração Própria

Sendo assim, o espectro da análise de cada agrupamento é ampliado, como pode ser observado na tabela 7, a matriz cruzada do cluster de maior grau de pertencimento demarcado no eixo vertical (cluster_fuzz_1) e o segundo maior grau de pertencimento demarcado no eixo horizontal (cluster_fuzz_2) possibilita a ampliação de 5 agrupamentos para 11. Para tanto, o segundo maior grau de pertencimento remete à segunda menor distância do consumidor em relação aos clusters, determinando certo nível de similaridade com este segundo grupo de classificação. Como exemplificado na tabela, quando os consumidores possuem seu cluster de maior grau de pertencimento representado pelo cluster “1”, 92,5% possui o segundo cluster com maior grau de pertencimento representado pelo cluster “3”:

Tabela 7 – Matriz de migração de clusters dado grau de pertencimento

	cluster_fuzz_2					
cluster_fuzz_1	1	2	3	4	5	Grand Total
1	0.0%	0.0%	92.5%	7.5%	0.0%	100.0%
2	0.0%	0.0%	78.1%	0.0%	21.9%	100.0%
3	77.4%	22.0%	0.0%	0.6%	0.0%	100.0%
4	97.5%	0.3%	2.2%	0.0%	0.0%	100.0%
5	0.0%	100.0%	0.0%	0.0%	0.0%	100.0%

Fonte: Elaboração Própria

Dessa forma, a partir das novas combinações, a tabela 8 demonstra a nova disposição dos clusters em relação aos valores das variáveis de *RFM:*

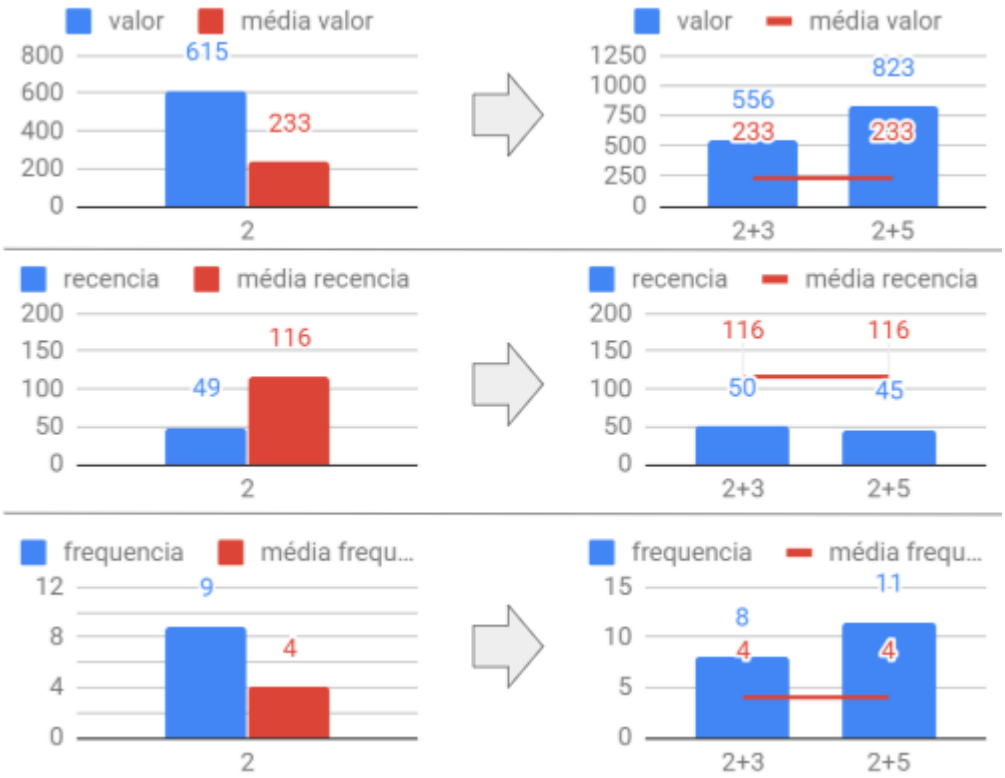
Tabela 8 – Média das variáveis *RFM* dos indivíduos nos 11 clusters obtidos no *fuzzy c-means*

cluster_fuzz_1	cluster_fuzz_2	valor	frequencia	recencia
1	**3**	85.5	1.7	77.0
1	**4**	66.8	1.5	170.0
2	**3**	556.0	8.1	50.3
2	**5**	823.5	11.4	45.4
3	**1**	229.1	4.0	58.5
3	**2**	366.2	5.8	56.2
3	**4**	364.1	4.8	203.8
4	**1**	79.1	1.4	282.9
4	**2**	605.0	6.3	317.3
4	**3**	382.7	3.8	232.3
5	**3**	1,291.9	16.6	35.7

Fonte: Elaboração Própria

Ao comparar todos os consumidores da tabela 5 marcados com o maior grau de pertencimento referente ao cluster “2”, vemos que a expansão do segundo maior grau de pertencimento demonstrado na tabela 8 possibilita um detalhamento maior dos clientes, sendo possível direcionar novas ações para a população deste cluster. Para tanto, a figura 11 representa tal mudança em comparação com a média total dos clientes, ou seja, a transição da população total do cluster de maior grau de pertencimento para a combinação com o segundo cluster de maior pertencimento:

Figura 11 – Combinação dos clusters dado o primeiro e segundo maior grau de pertencimento



Fonte: Elaboração Própria

Com isso, é possível obter maior detalhamento das características dos clientes com a mesma quantidade inicial de clusters obtidos na técnica *k-means*. Tal propriedade de entender não somente as características principais de um cluster, mas também a similaridade mais próxima com outros clusters possibilitam direcionamentos com mais fundamentos e confiabilidade para guiar e expandir a atuação de estratégias de CRM dentro das organizações, além de flexibilizar e aumentar a assertividade da tomada de decisão.

4.3 Conclusões

A utilização das técnicas de agrupamento contribuem para direcionar à excelência aos objetivos da área de CRM. A aplicação deixa em evidência características similares e dissimilares entre consumidores e com isso orientam as estratégias para maior rentabilidade, fidelização e satisfação do cliente, a partir da construção de um relacionamento sustentável, comunicações eficientes e otimização da gestão de relacionamento com o consumidor.

Ao comparar técnicas tradicionais como *k-means* à aplicação de técnicas como o *fuzzy c-means* é possível observar a garantia de um incremento da atuação dentro das estratégias de CRM dado a maior flexibilidade de utilização, ampliada pelas características do grau de pertencimento aos clusters, evidenciando maior granularidade e detalhamento dos consumidores e aperfeiçoando os objetivos de CRM.

Entretanto, é importante notar que quanto mais abrangente for a classificação do grau de pertencimento, ou seja, quanto mais combinações de agrupamentos, menor a quantidade de clientes por cluster e menor será a potencialização da utilização das técnicas de agrupamento, pois o objetivo de classificação à objetos similares torna-se cada vez menos significativo.

CAPÍTULO 5 - CONSIDERAÇÕES FINAIS

Tendo em vista o ambiente altamente competitivo em que as empresas estão inseridas, colocar o cliente como objeto central de estratégias organizacionais é fundamental para garantir vantagem competitiva. A inteligência anlítica obtida a partir da manipulação de dados sobre este cliente é essencial para tal objetivo e o desenvolvimento tecnológico está cada vez mais proporcionando acesso aos dados e grande volumes de informação, não obstante, entender, analisar e compreender essas informações é o diferencial que irá permitir ter a visão holística do consumidor e criar estratégias de ação e tomada de decisão no âmbito organizacional. Tais estratégias visam excelência em satisfação, retenção e lucratividade de clientes, para isso a área de CRM juntamente com a ciência de dados possibilita atingir tais objetivos a partir de análises e modelos. Em especial, os modelos de agrupamentos são eficazes em demonstrar as características dos consumidores em grupos similares e então criar insumos para atuação dentro do escopo de CRM.

Dentro dos modelos de agrupamento, comparar o impacto de técnicas tradicionais como o *k-means* com outras técnicas como o *fuzzy c-means* torna-se significativo para área de CRM a partir da nova ótica de utilização das técnicas em aplicações na organização. Dado que a proposta do algoritmo *fuzzy* proporciona uma nova ótica de agrupamento comparada aos métodos convencionais pois tem a característica de garantir o grau de pertencimento do objeto aos clusters, a aplicação da técnica em áreas específicas como CRM trazem um novo panorama na tomada de decisão no âmbito organizacional. A granularidade e flexibilidade obtida a partir da aplicação da técnica *fuzzy c-means* é demonstrada a partir da potencialização do agrupamento dado a proximidade com os outros clusters, evidenciado pelo grau de pertencimento. Portanto, a comparação da técnica de agrupamento *fuzzy* com a técnica convencional *k-means* torna-se relevante para expor as vantagens e desvantagens da aplicação dessas ferramentas para a busca da excelência dos objetivos de CRM.

REFERÊNCIAS

AL-AUGBY, Salam et al. A Comparison Of K-Means And Fuzzy C-Means Clustering Methods For A Sample Of Gulf Cooperation Council Stock Markets. **Folia Oeconomica Stetinensia**, [s.l.], v. 14, n. 2, p.19-36, 1 dez. 2014. Walter de Gruyter GmbH. http://dx.doi.org/10.1515/foli-2015-0001. Acesso em: 18 abr., 2019.

BAHARI, T. Femina; ELAYIDOM, M. Sudheep. An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour. **Procedia Computer Science**, [s.l.], v. 46, p.725-731, 2015. Elsevier BV. http://dx.doi.org/10.1016/j.procs.2015.02.136. Acesso em: 21 jan., 2019.

BERRY, Michael J. A.; LINOFF, Gordon S.. **Data Mining Techniques.** 2. ed. Indianopolis, Indiana, Eua: Wiley Publishing, Inc., 2004. 621 p. 349 - 381.

CAO, Longbing. Data Science. **Acm Computing Surveys**, [s.l.], v. 50, n. 3, p.1-42, 29 jun. 2017. Association for Computing Machinery (ACM). http://dx.doi.org/10.1145/3076253. Acesso em: 26 fev., 2019.

CORREIA, Christiane de Miranda e S., et al. **CRM nas Organizações**, Belo Horizonte, v. 6, n.1, p. 45-58, jul. 2005.

DAVENPORT, Thomas H.; HO-KIM, Jin. **Dados demais!: como desenvolver habilidades analíticas para resolver problemas complexos, reduzir riscos e decidir melhor.** Rio de Janeiro: Elsevier, 2014. 240 p.

GIL, Antônio Carlos, 1946 - Como elaborar projetos de pesquisa/Antônio Carlos Gil. - 4. ed. - São Paulo : Atlas, 2002.

GROUP, Peppers & Rogers. **CRM SERIES MARKETING 1 MARKETING 1 TO 1®.** 2004. Disponível em: <http://docplayer.com.br/431753-Crm-series-marketing-1-to-o-1-3-a-edicao-revista-e-ampliada-r-ferreira-de-araujo-202-10o-andar-05428-000-sao-paulo-sp-tel-55-11-3097-7610.html>. Acesso em: 06 dez. 2018.

HOSSEINI, Seyed Mohammad Seyed; MALEKI, Anahita; GHOLAMIAN, Mohammad Reza. Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. **Expert Systems With Applications**, [s.l.], v. 37, n. 7, p.5259-5264, jul. 2010. Elsevier BV. http://dx.doi.org/10.1016/j.eswa.2009.12.070. Acesso em: 30 abr. 2019.

NGAI, E.w.t.; XIU, Li; CHAU, D.c.k.. Application of data mining techniques in customer relationship management: A literature review and classification. **Expert Systems With Applications**, [s.l.], v. 36, n. 2, p.2592-2602, mar. 2008. Elsevier BV. http://dx.doi.org/10.1016/j.eswa.2008.02.021. Acesso em: 4 fev., 2019.

PAIXÃO, Alexandre de Oliveira et al. De Business Intelligence a Data Science: um estudo comparativo entre áreas de conhecimento relacionadas. In: CONGRESSO INTEGRADO DE TECNOLOGIA DA INFORMAÇÃO, 8., 2015, Rio de Janeiro. **Congresso Integrado de Tecnologia da Informação.** Rio de Janeiro, RJ: Citi, 2015. p. 40 - 48.

PETERS, Georg et al. Soft clustering – Fuzzy and rough approaches and their extensions and derivatives. **International Journal Of Approximate Reasoning**, [s.l.], v. 54, n. 2, p.307-322, fev. 2013. Elsevier BV. http://dx.doi.org/10.1016/j.ijar.2012.10.003. Acesso em: 04 fev., 2019.

RAJEH, Shohreh Mirzaiean et al. A Model for Customer Segmentation Based On Loyalty Using Data Minig Approach and Fuzzy Concept in Iranian Bank. **International Journal Of Business And Behavioral Sciences.** Ghazvin, Iran, p. 118-136. set. 2014. Disponível em: <https://pdfs.semanticscholar.org/0832/f0380bbe6d0a1ecfb1dab3af6cd904c7382c.pdf>. Acesso em: 01 abr. 2019

RUD, Olivia Parr. **Data Mining Cookbook:** Modeling Data for Marketing, Risk, and Customer Relationship Management. New York: John Wiley & Sins, Inc., 2001. 429 p.

SOLTANI, Zeynab; NAVIMIPOUR, Nima Jafari. Customer relationship management mechanisms: A systematic review of the state of the art literature and recommendations for future research. **Computers In Human Behavior**, [s.l.], v. 61, p.667-688, ago. 2016. Elsevier BV. http://dx.doi.org/10.1016/j.chb.2016.03.008.Acesso em: 18 abr., 2019.

WALLER, Matthew A.; FAWCETT, Stanley E.. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. **Journal Of Business Logistics.** Fayetteville, Ar, EUA, p. 77-84. mar. 2013.