



Alice Linhares Amigo

**Modelos de clientes de alto valor e impactos de uma
ação de relacionamento no aumento de valor do
cliente: uma contribuição para a teoria de valor do cliente.**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Administração de Empresas do Departamento de Administração da PUC-Rio.

Orientador: Prof. Jorge Brantes Ferreira

Rio de Janeiro
Abril de 2020



Alice Linhares Amigo

**Modelos de clientes de alto valor e impactos de uma
ação de relacionamento no aumento de valor do
cliente: uma contribuição para a teoria de valor do cliente.**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Administração de Empresas da PUC-Rio. Aprovada pela Comissão Examinadora abaixo.

Prof. Jorge Brantes Ferreira

Orientador

Departamento de Administração de Empresas – PUC-Rio

Profa. Ângela da Rocha

Departamento de Administração de Empresas – PUC-Rio

Prof. Angilberto Sabino de Freitas

UNIGRANRIO

Rio de Janeiro, 29 de abril de 2020

Todos os Direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autoria e do orientador.

Alice Linhares Amigo

Graduou-se em Comunicação Social – Publicidade e Propaganda na PUC-Rio em 2008. Cursou MBA em Gestão de Negócios na Fundação Dom Cabral, com conclusão em 2012. É professora convidada da ESPM, ministrando cursos de curta duração na área de varejo.

Ficha Catalográfica

Amigo, Alice Linhares

Modelos de clientes de alto valor e impactos de uma ação de relacionamento no aumento de valor do cliente : uma contribuição para a teoria de valor do cliente / Alice Linhares Amigo ; orientador: Jorge Brantes Ferreira. – 2020.

84 f. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Administração, 2020.

Inclui bibliografia

1. Administração – Teses. 2. Valor do cliente. 3. . Ação de relacionamento. 4. Varejo. 5. Modelos de clientes. 6. *Machine Learning*. I. Ferreira, Jorge Brantes. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Administração. III. Título.

CDD: 658

Para meu pai, José Carlos Vilar Amigo,
que sempre me inspirou na sua carreira
profissional e acadêmica e me incentivou
a correr atrás dos meus objetivos.

Agradecimentos

Ao meu orientador Professor Jorge Brantes Ferreira pelo estímulo, ensinamentos e parceria para a realização deste trabalho.

Ao CCPG e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

A todos os meus professores, que me ensinaram tanto durante o período de Mestrado.

Aos meus pais, pela educação, atenção, carinho e incentivo.

Aos meus colegas da PUC-Rio.

Aos professores que participaram da Comissão examinadora.

A todos os professores e funcionários do Departamento de Administração pelos ensinamentos e pela ajuda.

A todos os amigos e familiares que, de uma forma ou de outra, me estimularam ou me ajudaram.

A todos os funcionários da rede varejista parceira neste trabalho.

"O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001"

Resumo

Amigo, Alice Linhares; Ferreira, Jorge Brantes. Modelos de clientes de alto valor e impactos de uma ação de relacionamento no aumento de valor do cliente: uma contribuição para a teoria de valor do cliente. Rio de Janeiro, 2020. 84p. Dissertação de Mestrado – Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro.

Em um contexto mercadológico de orientação para o mercado e não mais para o produto, torna-se fundamental que as empresas conheçam seus clientes, entreguem valor a eles e maximizem o valor dos clientes para a empresa. O valor do cliente está relacionado ao valor econômico do relacionamento do cliente com a empresa. Nesta relação, empresas podem obter informações em relação ao comportamento dos clientes e construir bases de dados valiosas que ajudem a compreender o perfil de cada um e mensurar o valor dos mesmos. A partir de uma base de dados de clientes de uma grande empresa de varejo no setor supermercadista, esse trabalho tem como objetivo propor um modelo de classificação de clientes de alto valor e analisar a resposta destes e dos demais clientes a uma ação de relacionamento de *cross-selling*, compreendendo o impacto desta ação no aumento de valor do cliente. O trabalho contribui com o entendimento do perfil de clientes com maior e menor propensão de resposta a uma ação de CRM e a consequente geração de valor à base de clientes, sendo útil a praticantes e acadêmicos. A partir de um modelo de dados secundários extraídos de bancos de dados de um supermercado brasileiro, aplicou-se modelos *machine learning* de classificação em uma base de mais de 1.500 clientes para encontrar modelos que expliquem clientes de alto valor.

Palavras-Chave

Valor do cliente; ação de relacionamento; varejo; modelos de clientes; *machine learning*.

Abstract

Amigo, Alice Linhares; Ferreira, Jorge Brantes (Advisor). High Value Customer Models and impacts of relationship action on increasing customer value: a contribution to customer Value Theory. Rio de Janeiro, 2020. 84p. Dissertação de Mestrado – Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro.

In a market context oriented towards the customer and no longer towards the product, it is essential that companies know their customers, deliver value to them and maximize the value of customers to the company. The value of the customer is related to the economic value of the customer's relationship with the company. In this relationship, companies can obtain information regarding customer behavior and build valuable databases that help to understand each other's profile and measure their value. Based on a database of customers of a large retail company in the supermarket sector, this work aims to propose a model for classifying high-value customers and analyze the response of these and other customers to a cross-selling relationship action, understanding the impact of this action on increasing customer value. The work contributes to the understanding of the customer profile with greater and lesser propensity to respond to a CRM action and the consequent generation of value to the customer base, being useful to practitioners and academics. Using a secondary data model extracted from databases in a Brazilian supermarket, machine learning classification models were applied to a base of more than 1,500 customers to find models that explain high-value customers.

Keywords

Customer value; customer relationship action; retail; customer models; machine learning

Sumário

1. INTRODUÇÃO	12
1.1 Objetivo	15
1.2 Relevância do estudo	16
1.3 Delimitação do estudo	17
2. REVISÃO DE LITERATURA	18
2.1 Valor do Cliente	18
2.1.1 CLV como métrica de valor do cliente	19
2.2 Gestão de valor do cliente em um contexto multicanal	25
2.3 Aumento de valor do cliente	29
2.3.1 <i>Cross-selling</i> como ferramenta de aumento de valor do cliente	32
3. MÉTODO	37
3.1 Tipo de Pesquisa	38
3.2 Dados coletados	39
3.3 Procedimentos de tratamento e variáveis utilizadas	43
3.4 População e amostra	45
3.5 Aplicação de algoritmos de <i>Machine Learning</i>	49
3.6 Limitações do estudo	54
4. ANÁLISE E RESULTADOS	55

4.1 Análise e resultados gerais dos modelos	55
4.1.1 Modelos gerados.....	55
4.1.2 Acurácia dos modelos	55
4.1.2.1 Acurácia dos modelos - Etapa 1.....	56
4.1.2.2 Acurácia dos modelos - Etapa 2.....	57
4.2 Análise dos modelos com maior acurácia	59
4.2.1 Análise e resultados do modelo de regressão logística - Etapa 1 ...	59
4.2.2 Análise e resultados do modelo de regressão logística - Etapa 2 ...	61
4.2.3 Análise e resultados dos modelos de árvore de decisão: Etapa 1 ..	62
4.2.4 Análise e resultados dos modelos de árvore de decisão: Etapa 2 ..	65
5. CONCLUSÃO	70
5.1 Implicações teóricas	72
5.2 Implicações gerenciais	74
5.3 Limitações do estudo e recomendações de pesquisas futuras	75
6. REFERÊNCIAS BIBLIOGRÁFICAS	76

Lista de figuras

Figura 3.1: processo de aplicação do <i>machine learning</i> supervisionado de KOTSIANTIS (2007)	50
Figura 4.1. – Exemplo de estrutura de árvore de decisão	53
Figura 4.2. – Estrutura de árvore de decisão do algoritmo J48 – Etapa 1	63
Figura 4.3. – Estrutura de árvore de decisão do algoritmo REPTree – Etapa 1	63
Figura 4.4. – Estrutura de árvore de decisão do algoritmo J48 – Etapa 2	66
Figura 4.5. – Estrutura de árvore de decisão do algoritmo REPTree – Etapa 2	66

Lista de Tabelas

Tabela 2.1. – <i>Drivers</i> de CLV em contextos B2C	22
Tabela 3.1. – <i>Tabela de variáveis coletadas</i>	40
Tabela 3.2. – Variáveis base final das duas etapas do estudo	44
Tabela 3.3. – Total de clientes em cada amostra e premissas	48
Tabela 3.4. – Exemplo de matriz de confusão	51
Tabela 4.1. – Nomes dos algoritmos para referência	55
Tabela 4.2. – Matrizes de confusão dos modelos calculados na etapa 1	56
Tabela 4.3. – Resumo dos resultados de desempenho dos modelos na etapa 1	56
Tabela 4.4. – Matrizes de confusão dos modelos calculados na etapa 2	57
Tabela 4.5. – Resumo dos resultados de desempenho dos modelos etapa 2	58
Tabela 4.6. – Atributos e <i>odds ratio</i> estimados pelo modelo de regressão logística da etapa 1 para classe sim	59
Tabela 4.7. – Atributos e <i>odds ratio</i> estimados pelo modelo de regressão logística da etapa 2 para classe sim	61
Tabela 4.8. – Resultados sumarizados	68

1

INTRODUÇÃO

Nas últimas décadas do século XX ocorreram grandes mudanças mercadológicas. Neste contexto de transformações, construir uma orientação para o mercado, ou seja, focar em clientes e concorrentes, e não mais em produtos, passa a ser uma vantagem competitiva (HUNT; ROBERT M. MORGAN, 1995). Com um entendimento cada vez maior da importância da lealdade de cada cliente, empresas formulam estratégias com abordagens mais centradas nestes (DYER; SINGH, 1998; JAIN; SINGH, 2002), tratando-os como ativos e focando em adquiri-los e retê-los. Uma vez que estão orientadas para o mercado, empresas alinham suas organizações em torno dos clientes (KUMAR; SHAH, 2009) e precisam considerar como suas atividades de marketing impactam em seus relacionamentos com cada um deles (REINARTZ; KUMAR, 2003).

Desta forma, torna-se fundamental que as empresas construam relacionamentos compreendendo os clientes e sendo capazes de entregar valor a eles. Adotar uma orientação de entrega de valor ao cliente exige que as empresas aprendam sobre seus mercados e clientes alvo e traduzam estes aprendizados em performance superior com clientes (ROBERT B. WOODRUFF, 1997). O aumento do valor percebido pelo cliente em cada interação com a empresa aumenta o nível de satisfação, levando a taxas maiores de retenção e, possivelmente, a um aumento de valor do cliente para a empresa. A compreensão dos clientes pela empresa é facilitada pelo avanço tecnológico. Por meio de aplicativos *mobile*, plataformas de *e-commerce* e redes sociais, dentre outras tecnologias, clientes e empresas se conectam e estabelecem, assim, uma aproximação (VENKATESAN, 2017).

Dentro desse contexto, de avanço tecnológico, orientação para o mercado e entrega de valor ao cliente, inicia-se a prática de gestão de relacionamento com os clientes com uso de base de dados. A prática de analisar e utilizar base de dados de marketing para determinar práticas e métodos corporativos que maximizarão o valor da vida útil de cada cliente individual é chamada CRM, ou *Customer*

Marketing Relationship (KUMAR; REINARTZ, 2006). Como se vê pela definição, maximizar o valor do cliente é um objetivo de suma importância dentro do CRM. Estratégias de CRM tem se mostrado essenciais para ajudar empresas a terem melhor desempenho, permitindo que elas identifiquem os melhores clientes e satisfaçam suas necessidades para que os mesmos permaneçam leais (THOMAS; SULLIVAN, 2005). Estudos já verificaram uma relação de efeito positivo entre os resultados da empresa e a implementação de estratégias de CRM (REINARTZ et al., 2004).

Uma parte dos estudos empíricos dentro do contexto de CRM, com o uso de grandes bases de dados de clientes, examinam como as empresas podem aumentar as medidas de lealdade (KUMAR; VENKATESAN; REINARTZ, 2008; VERHOEF, 2003; VERHOEF; FRANCES; HOEKSTRA, 2001) e como preveem essas métricas (FADER; HARDIE; LEE, 2005; NESLIN et al., 2006a), ou como empresas podem influenciar e otimizar o valor do cliente (RUST; VERHOEF, 2005; SCHMITT; SKIERA; BULTE, 2011; VENKATESAN; KUMAR, 2004a; VENKATESAN; KUMAR; BOHLING, 2007). O valor dos clientes é um dos fatores aos quais os gestores mais deveriam dar atenção (SCHMITT; SKIERA; BULTE, 2011).

Valor do cliente se refere ao valor econômico do relacionamento do cliente com a empresa (KUMAR, 2018). Os clientes podem interagir com as empresas e criar valor de maneiras diferentes (KUMAR et al., 2010). A proposição total de valor entre uma empresa e um cliente pode ser capturada pelo valor que o cliente gera à empresa e através do valor gerado pela empresa ao cliente (PEPPERS; ROGERS, 2011). Quando se considera a contribuição do valor do cliente para a empresa, uma parte crucial é a contribuição dele em períodos futuros (KUMAR, 2018).

O cliente tem o valor atual, que é a receita menos o custo de servir, e também o valor de longo prazo, que pode aumentar ou diminuir a partir de experiências com a marca, com a empresa, influências externas e mudanças em suas necessidades (PEPPERS; ROGERS, 2011). O valor dos clientes deve ser gerenciado e medido (GUPTA; LEHMANN; STUART, 2004). Taxas de retenção e contribuições na

margem são os dois principais componentes de valor do cliente (GUPTA; ZEITHAML, 2006; WIESEL; SKIERA; VILLANUEVA, 2008).

Aumentar o valor da base de clientes deve ser o primeiro objetivo de uma empresa com estratégia direcionada para clientes, já que estes são a fonte de todas as receitas de curto prazo e de criação de valor de longo prazo (PEPPERS; ROGERS, 2011). Uma das ferramentas utilizadas por empresas para aumento de valor do cliente é o *cross-selling*, ou venda cruzada. *Cross-selling* é a prática de vender novos produtos ou serviços para um consumidor existente, e o objetivo desta prática é oferecer o produto certo, para o consumidor certo, na hora certa, usando o canal de comunicação certo para obter sucesso de longo prazo (LI; SUN; MONTGOMERY, 2011). Desta forma, é importante que se conheça o cliente e seu comportamento.

Para conhecer cada cliente, a obtenção de dados é de suma importância. As empresas obtêm valor a partir da base de clientes (KUMAR; REINARTZ, 2016). As empresas podem obter dados através dos diferentes canais de vendas e comunicação e, à medida que as empresas mantêm e expandem seus relacionamentos com clientes em vários canais, as atividades de gestão de relacionamento com o cliente ganham cada vez mais relevância e se tornam mais complexas (THOMAS; SULLIVAN, 2005).

Os varejistas, em particular, durante o processo de transação, capturam muitos dados que podem ser agregados ao nível cliente, como tipos de produtos comprados, transações prévias e valor histórico do cliente (VERHOEF et al., 2002). Esses dados também podem ser agrupados ao nível loja, permitindo saber qual a venda total, quantidade de visitas, venda total por categorias, dentre outras informações (BUCKLIN; GUPTA, 1999). Enquanto os modelos de varejos tradicionais dependem de um programa de lealdade para relacionar o cliente individual com as respectivas transações, a natureza do modelo de negócio de um varejo *online* permite que as transações sejam associadas aos clientes diretamente nas bases de dados (VERHOEF et al., 2010).

No entanto, muitos varejistas não coletam os dados corretamente ou não os analisam adequadamente, fazendo com que muitas implementações de CRM sejam

mal sucedidas (VERHOEF et al., 2010). Pesquisas mostraram que a qualidade dos dados está positivamente relacionada com a performance da empresa (JAYACHANDRAN et al., 2004; ZAHAY; GRIFFIN, 2002). O volume cada vez maior de dados em um ritmo intenso gerou novos problemas na gestão e aproveitamento de dados e, o principal, é que abordagens estatísticas comuns para avaliar a relevância dos resultados se tornaram pouco efetivas (ZERBINO et al., 2017). Já ferramentas de *data mining* constroem uma representação da realidade na constituição de um modelo que ajuda na tomada de decisão (SAMIZADEH; MEHREGAN, 2015).

Data Mining, ou mineração de dados, é a atividade de analisar uma quantidade relevante de dados com o intuito de encontrar padrões e relações significativas (BERRY; GS LINOFF, 2000). A mineração de dados é uma das mais relevantes aplicações para *Machine Learning*, ou Aprendizado de Máquina (KOTSIANTIS, 2007). Entender os clientes através de interações pessoais não é algo possível para as empresas, então, as empresas estão cada vez mais recorrendo a dados transacionais para entender o comportamento de clientes (KUMAR; PANSARI, 2016). Conhecendo quais clientes são mais e menos valiosos, as empresas podem priorizar seus esforços competitivos, dedicando maior esforço, tempo e recursos aos clientes mais propensos a produzir maiores retornos (PEPPERS; ROGERS, 2011), e a Mineração de Dados é uma técnica que pode ser usada para conhecer melhor os clientes.

1.1 Objetivo

A partir de uma base de dados de clientes de uma grande empresa de varejo de supermercados, este trabalho tem como objetivo propor um modelo de classificação de clientes de alto valor e de clientes que respondem a uma ação de relacionamento de *cross-selling* através de algoritmos *machine learning* para implementação de *data mining*. Este estudo divide-se, então, em duas etapas: na primeira verifica-se quais atributos são relevantes na classificação de clientes de alto valor e, na segunda, quais atributos são relevantes na classificação de clientes que respondem a uma ação de relacionamento de *cross-selling*.

Nas duas etapas, mais de dez atributos relacionados diretamente ao valor do cliente, dentro do contexto do varejo supermercadista, são considerados e, dentre estes, o comportamento de compra multicanal. Com isso, o estudo se propõe a contribuir com a literatura que aborda valor do cliente dentro do contexto de CRM e também em um contexto multicanal.

1.2

Relevância do estudo

Muitas empresas coletam e mantêm valiosos dados secundários do comportamento de compra e interação de clientes que servem como uma boa fonte para descobrir aspectos do relacionamento empresa-cliente, incluindo os motivadores de compra cruzada (KUMAR; GEORGE; PANCRAS, 2008). Porém, conseguir bases de dados adequadas para aplicação de modelos de resposta é um desafio no contexto de pesquisas acadêmicas, uma vez que tais bases devem possuir grandes quantidades de registros para serem analisadas, com muitas variáveis referentes a cada um destes registros. Este estudo, então, pode contribuir com esta lacuna ao utilizar uma base de dados de um varejo supermercadistas, com mais de 200 mil clientes identificados, para desenvolver um modelo preditivo de clientes que respondem a uma ação de relacionamento de *cross-selling* e de clientes de alto valor.

Muitos estudos já se propuseram a investigar e medir o valor do cliente (BOLTON; LEMON; VERHOEF, 2004; REINARTZ; KUMAR, 2000; RUST; LEMON; ZEITHAML, 2004; SHAH et al., 2016; VENKATESAN; KUMAR, 2004b), sendo alguns deles com análises ao nível agregado (BOLTON; LEMON; VERHOEF, 2004; REINARTZ; KUMAR, 2000, 2003; RUST; LEMON; ZEITHAML, 2004; SHAH et al., 2016; SUNDER; KUMAR; ZHAO, 2016; VENKATESAN; KUMAR, 2004a). No entanto, dentre as pesquisas que este estudo abrangeu, não foram localizados estudos com aplicação de métodos *machine learning*, o que indica que a aplicação deste método dentro de pesquisas acadêmicas na temática de valor do cliente ainda não é muito disseminada.

Quando se tenta estabelecer relações entre múltiplos fatores e durante processos de análises, pode-se cometer erros, e o *machine learning*, geralmente, é aplicado com sucesso para solucionar esse tipo de problema (KOTSIANTIS, 2007).

Somado a isso, o uso de técnicas de mineração de dados para análises de padrões de consumo não é facilmente encontrado na literatura acadêmica. Assim, através de algoritmos *machine learning* para implementação de *data mining*, este estudo contribui com pesquisas acadêmicas relacionadas a valor do cliente utilizando um método ainda não muito disseminado neste campo de estudos. Em especial, esta pesquisa contribui com a temática de valor do cliente dentro do contexto de relacionamento com o cliente, e também no contexto de varejo multicanal, uma vez que a base de dados utilizada é de um supermercado com operação *online* e em lojas físicas.

Adicionalmente, no campo acadêmico, não é tão vasta a quantidade de pesquisas desenvolvidas para identificar os motivadores de *cross-selling* no contexto de valor do cliente (HEINER EVANSCHITZKY, NEERU MALHOTRA, 2017; KUMAR; GEORGE; PANCRA, 2008) e, além disso, não é possível para as empresas alcançar todos os clientes com este tipo de ação. Assim, é fundamental identificar os clientes com maior propensão a responder a ações *cross-selling* para maximizar o retorno sobre os investimentos (KUMAR; GEORGE; PANCRA, 2008). Este trabalho contribui, então, com praticantes e acadêmicos na compreensão de modelos de clientes com maior propensão de resposta a uma ação de CRM.

1.3

Delimitação do estudo

A proposta deste estudo é modelar o comportamento de um grupo de consumidores de uma grande rede de supermercados considerando o valor do cliente e a resposta de clientes quando estimulados por uma ação de CRM que tem como objetivo aumentar o valor do cliente através do incremento de produtos comprados por eles. Portanto, este estudo contempla apenas a temática de valor do cliente dentro dos estudos no campo de CRM, e considera perfil de clientes apenas no contexto de varejo supermercadista.

2 REVISÃO DE LITERATURA

2.1 Valor do Cliente

As discussões acerca de valor do cliente se moldam a partir de estudos que defendem a orientação de mercado como uma fonte potencial de vantagem competitiva (HUNT; ROBERT M. MORGAN, 1995; ROBERT B. WOODRUFF, 1997; SLATER, 1997a). A orientação para o mercado é a coleta de informações de clientes e concorrentes, e o uso do conhecimento obtido a partir disto, para guiar estratégias que entreguem valor ao cliente (HUNT; ROBERT M. MORGAN, 1995; SLATER, 1997b).

Hunt, Robert e Morgan (1995) diferenciam o significado de valor do cliente na perspectiva da organização e na do cliente. Na primeira, é a quantificação monetária de um cliente individual para uma empresa. Já na segunda perspectiva, valor do cliente significa considerar o que os clientes querem e acreditam que podem obter ao comprar e usar um produto de uma empresa. Qualquer negócio sustentável cria valor aos seus clientes através de ofertas da empresa e, neste processo, obtém valor de seus clientes em forma de lucros (KUMAR, 2018). A proposta de um negócio sustentável é criar valor aos clientes, primeiramente, e extrair parte deste valor fornecido ao cliente como lucratividade, criando, assim, valor para a empresa (KUMAR; REINARTZ, 2016).

Gupta et al. (2004, p. 7) definem valor do cliente como “a soma esperada de ganhos futuros descontados com base em premissas-chave sobre taxa de retenção e margem de lucro”. O valor de todos os clientes é determinado pela taxa de aquisição e pelo custo de adquirir novos clientes. Kumar e Reinartz (2012, p. 4) definem valor do cliente como “o valor econômico da relação do cliente com a firma, expresso com base na margem de contribuição ou lucro líquido”. O valor do cliente está relacionado ao lucro que pode ser obtido pela empresa no relacionamento com o

cliente. Clientes com muito ou pouco tempo de vida na empresa podem ser altamente rentáveis (REINARTZ; KUMAR, 2000).

Segundo Kumar e Reinartz (2000, p. 18) “lucros de tempo de vida de um cliente individual são modelados como uma função da duração do tempo de vida de um cliente, fluxo de receitas sobre o curso do tempo de vida de um cliente e o custo da empresa associado com o intercâmbio de marketing.” Nesse mesmo estudo, os autores afirmam que a busca de variedade entre as empresas, o tédio dos clientes de interagir com a mesma empresa, ações de empresas concorrentes e a ausência de contratos são alguns fatores que podem levar a custos que excedam a receita, e esta lacuna pode aumentar com o tempo, até um ponto no qual um relacionamento negativo prevalece. Clientes que tem maior tempo de vida e geram altas receitas representam os clientes mais valiosos para a empresa.

Comportamentos de compra vigentes dependem de comportamentos de compra passados, e essa característica explica o uso comum do RFM (Recência, Frequência e Valor Monetário) nos métodos de segmentação dentro da base de marketing (BOLTON; LEMON; VERHOEF, 2004). Recência é o momento da compra mais recente, frequência é o número de compras passadas e o valor monetário é o gasto médio na compra por transação (FADER; HARDIE; LEE, 2005). Há estudos que evidenciam que mudanças em características comportamentais, como compras passadas, são bons indicadores de comportamentos futuros do cliente (DWYER, 1997; ROSSI; MCCULLOCH; ALLENBY, 1996). Um estudo apresentou um modelo formal que demanda atributos da RFM para fazer projeções específicas do valor do tempo de vida para um grupo de clientes, relacionando o modelo RFM com o CLV (FADER; HARDIE; LEE, 2005).

2.1.1 CLV como métrica de valor do cliente

Para compreender o valor de vida útil do cliente, uma métrica com grande aceitação é o CLV (*Customer Lifetime Value*), que surgiu como uma métrica efetiva para o CRM e concede à empresa uma visão de longo prazo, avançada e com orientação aos lucros da base de clientes (SUNDER; KUMAR; ZHAO, 2016). O

CLV tem sido usado por empresas para formular e implementar estratégias específicas para clientes (KUMAR; PANSARI, 2016).

O CLV representa o valor presente do fluxo de caixa de um cliente (VENKATESAN; KUMAR; BOHLING, 2007) acumulado ao longo do seu tempo de vida (KUMAR; PANSARI, 2016). Essa métrica pode ser usada para identificar os clientes mais e menos rentáveis, classificando-os em diferentes segmentos de lucratividade (baixo, médio e alto). A computação do CLV implica em prever o fluxo de caixa futuro de cada cliente, incorporando em uma única equação os elementos de receita, despesas e comportamentos de clientes que orientam sua lucratividade (KUMAR; SHAH, 2009). Dentre os componentes do *Customer Lifetime Value* estão a frequência de compra, a margem de contribuição e os custos de relacionamento (KUMAR; PANSARI, 2016; VENKATESAN; KUMAR, 2004b).

A frequência de compra é uma variável contável que mede o número de compras que um cliente fez. Calcula-se a margem de contribuição bruta deduzindo-se da receita recebida o custo do bem vendido em cada compra. Os custos de relacionamento são os custos incorridos para cada vez mais aumentar o valor do cliente no relacionamento com a empresa. A margem de contribuição é afetada pelas atividades de compra direcionadas ao cliente (KUMAR; PANSARI, 2016). Venkatesan e Kumar (2004b) definem o CLV em uma função que prediz a margem de contribuição, a propensão do cliente a continuar no relacionamento com a empresa e os recursos de marketing alocados ao cliente, conforme segue:

$$CLV_i = \sum_{t=1}^n \frac{(Contribuição\ de\ margem\ futura_{it} - Custo\ Futuro_{it})}{(1 + r)^t}$$

i=índice do cliente

t=índice de tempo

n=horizonte de previsão

r=taxa de desconto

Um outro estudo (RUST; LEMON; ZEITHAML, 2004) defende que o valor do tempo de vida de cada cliente resulta da frequência de compra da categoria, quantidade média de compras, e os padrões de trocas de marcas combinados com a margem de contribuição da empresa. O CLV ganhou rápida aceitação como métrica de aquisição, crescimento e retenção dos clientes corretos na gestão de relacionamento com o cliente (VENKATESAN; KUMAR, 2004a).

Os modelos de CLV ajudam a quantificar o relacionamento da empresa com os seus clientes e, conseqüentemente, permitem que a empresa tome decisões a partir de informações estruturadas, além de ajudar na identificação dos clientes mais lucrativos para alocação de recursos de marketing (JAIN; SINGH, 2002). Um estudo propôs estimar o valor do cliente, a partir de um software com variáveis e estimativas para esse fim, com base no histórico transacional, financeiro e interações do cliente com um varejo brasileiro de supermercados (VICHROSKI et al., 2010). Um outro estudo investigou a influência do valor do cliente na intenção de recomendação, calculando o CLV a partir de dados de clientes de uma instituição financeira (DA COSTA; CLARO; BORTOLUZZO, 2018).

Alguns estudos fornecem orientações para cálculo do CLV e retorno sobre investimento ao nível agregado (BOLTON; LEMON; VERHOEF, 2004; REINARTZ; KUMAR, 2000, 2003; SHAH et al., 2016). Dentro dos estudos ao nível agregado, há alguns com evidências empíricas de uma relação entre ações de marketing e CLV (REINARTZ; KUMAR, 2003; RUST; LEMON; ZEITHAML, 2004; SUNDER; KUMAR; ZHAO, 2016; VENKATESAN; KUMAR, 2004a).

Dentre as pesquisas empíricas que trazem uma relação entre CLV e ações de marketing, há estudos que recomendam o CLV como uma métrica para seleção de clientes e definição de programas de relacionamento (REINARTZ; KUMAR, 2003; RUST; LEMON; ZEITHAML, 2004). Há achados que concluem que interações por vários canais influenciam o valor do tempo de vida do cliente de maneira não linear (VENKATESAN; KUMAR, 2004b). Um dos estudos considerou o impacto de ofertas concorrentes e a troca de marcas no modelo de CLV, permitindo a criação de um modelo que contém atração e retenção de cliente em um contexto de troca de marcas (RUST; LEMON; ZEITHAML, 2004).

Há ainda um grupo de pesquisas que investigaram como características relacionadas aos clientes influenciam os componentes do CLV (SUNDER; KUMAR; ZHAO, 2016; VENKATESAN; KUMAR, 2004a). Alguns estudos mostram que o CLV era aplicado considerando-se a retenção como a única fonte relevante de ganho (GUPTA; LEHMANN; STUART, 2004), ignorando outros comportamentos do cliente. Outros estudos consideraram a contribuição de outros comportamentos do cliente na aplicação do CLV através das receitas que geram para a empresa, como uso de serviço e compras adicionais (BOLTON; LEMON; VERHOEF, 2004).

Em um estudo de Kumar e Reinartz (2016), são investigadas e apresentadas pesquisas que examinaram o aspecto de aplicabilidade do valor do cliente em tempo real e que forneceram informações para implementação. Os atributos que direcionam o valor do cliente determinam a natureza do relacionamento entre firma e cliente, e ajudam a estimar o nível de lucratividade e do CLV para cada cliente. Os autores classificam estes atributos em dois tipos: características de troca e heterogeneidade do cliente, ambos apresentados em estudo anterior (REINARTZ; KUMAR, 2003). As características de troca contemplam o conjunto de variáveis que definem e descrevem atividades de relacionamento em sentido mais amplo, e a heterogeneidade do cliente faz referência a indicadores demográficos e psicográficos que ajudam a empresa na segmentação de clientes e na gestão do relacionamento empresa-cliente. A heterogeneidade do cliente e as características de troca são determinadas pela natureza do negócio, ou seja, se é B2C (*Business to Consumer*) ou B2B (*Business to Business*).

A tabela 2.1 contém a lista apresentada na pesquisa (KUMAR; REINARTZ, 2016) dos atributos que direcionam o valor do cliente em empresas B2C, que é o caso de empresas no setor de varejo de supermercados (com formato de entrega de produtos e serviços a consumidores finais):

Tabela 2.1: *Drivers* de CLV em contextos B2C

Características de Troca	<ul style="list-style-type: none"> - Níveis de gastos passados do cliente - Comportamento de compra cruzada - Comportamento de compra focado - Média do tempo entre compras - Participação em programas de lealdade - Retorno do cliente - Contatos iniciados pelo cliente - Frequência dos contatos de marketing - Tipo dos contatos de marketing - Comportamento multicanal
--------------------------	---

	<ul style="list-style-type: none"> - Intensidade de uso do negócio - Intensidade do uso de ofertas
Heterogeneidade do Cliente	<ul style="list-style-type: none"> - Idade - Gênero - Renda - Localização física dos clientes

(Kumar e Reinartz, 2016)

Neste mesmo estudo, diferentes abordagens identificadas na literatura de marketing são apresentadas para entender e medir o valor dos clientes. Estas abordagens são classificadas em duas grandes categorias: abordagem agregada e abordagem individual. Na primeira, a média do valor de tempo de vida de um cliente deriva de um valor de tempo de vida de um segmento. Já na segunda abordagem, é computado o CLV de um cliente ao longo do seu tempo de vida com a empresa. Esta estratégia contribui para que empresas personalizem estratégias de acordo com as necessidades de clientes e com a potencial lucratividade futura de cada cliente.

Kumar (2018) amplia a visão de valoração do cliente propondo a CVT, ou *Customer Valuation Theory*. Esta é uma teoria para metrificar o valor futuro do cliente, que pode ser visto com um ativo. Na definição de Kumar, um ativo pode ser amplamente definido como qualquer atributo físico, organizacional ou humano que permita à empresa gerar e implementar estratégias que melhorem sua eficiência e efetividade no mercado. No entanto, para o autor, valorar ações de bolsa de valores, por exemplo, não é o mesmo que valorar clientes, já que quando se trata de valorar clientes a relação entre investimentos e ganhos não é linear. O valor e o volume de investimentos em clientes não são escaláveis e um novo balanço de portfólio de clientes é relativamente mais difícil, por exemplo.

A partir dos determinantes de cliente como ativo, Kumar (2018) apresenta uma abordagem para valorar clientes de uma maneira geral. Nesta abordagem, a *CFP (Customer Future Porfitability)* ou Rentabilidade Futura do Cliente, depende de comportamentos transacionais passados e vigentes, dos esforços de marketing da empresa, das variáveis demográficas e do ambiente no qual esses clientes estão inseridos. O autor representa a rentabilidade futura do cliente, conforme segue:

$$CFP_i = f(\text{Comportamento Transacional}_i, \text{Custo de Marketing}_i, \text{Variáveis Demográficas} / \text{Variáveis da Empresa}_i, \text{Fatores Ambientais e Econômicos})$$

O estudo indica como as empresas podem usar a CVT para valorar ativos de clientes, gerir portfólio de clientes e nutrir clientes rentáveis. A CVT foca em dois aspectos das contribuições financeiras do cliente: sua natureza (direta e indireta) e seu escopo (amplo e profundo). Assim, é um mecanismo para medir o valor futuro de cada cliente com base em: (1) contribuição direta do valor econômico do cliente, (2) intensidade da contribuição direta do valor econômico e (3) amplitude da contribuição indireta do valor econômico contabilizando a volatilidade e a vulnerabilidade do fluxo de caixa do cliente.

A contribuição do valor econômico direto é o valor econômico do relacionamento do cliente com a empresa, expresso como lucro líquido (ou margem de contribuição). A intensidade da contribuição do valor econômico direto é a intensidade e a exclusividade do valor direto de contribuição do cliente para a empresa através de compras pessoais que tenham produzido resultado financeiro significativo. E a amplitude da contribuição do valor econômico indireto é o valor de contribuição indireto do cliente para a empresa através de indicações, influência *online* em potenciais clientes e outras compras de clientes.

Kumar (2018) cria uma relação entre a CVT e o valor da firma, relacionando conceitos, métricas e estratégias. Ele define três estratégias relacionadas a cada uma das métricas e conceitos para nutrir clientes rentáveis: relacionamento, interação e engajamento. Selecionar clientes considerando o seu valor futuro pode ser uma estratégia efetiva para as empresas. A lucratividade futura do cliente depende de seu comportamento passado, dos esforços de marketing, da identidade e perfil dos clientes e do ambiente (KUMAR, 2018).

Estudos abordam como valorizar clientes é um caminho viável para valorizar empresas, incluindo empresas com alto crescimento e ganhos negativos (GUPTA; LEHMANN; STUART, 2004). Clientes selecionados com base no valor do tempo de vida geram lucros muito maiores em períodos futuros do que os selecionados a partir de outras métricas. Alguns fatores sob o controle da gestão, como comunicações de marketing, explicam a variação da lucratividade na duração do tempo de vida (REINARTZ; KUMAR, 2003). Quando os gestores definem regras de alocação e recursos que maximizam o tempo de vida do cliente, há um maior potencial para melhorar a lucratividade (VENKATESAN; KUMAR, 2004a).

2.2

Gestão de valor do cliente em um contexto multicanal

O contexto de Gestão de Valor do Cliente, ou *Customer Value Management* (CVM), envolve a otimização do valor da base de clientes de uma empresa (VERHOEF; DOORN; DOROTIC, 2007). O valor da base de clientes de uma empresa é a soma do valor de vida útil de seus clientes atuais e futuros (GUPTA; LEHMANN; STUART, 2004). O CVM foca na análise de dados individuais de clientes potenciais e vigentes, com tópicos específicos de investigação como previsão do valor da vida útil do cliente e desenvolvimento de modelos para ofertas específicas ao cliente (VERHOEF; DOORN; DOROTIC, 2007).

Dentro do campo de pesquisa de CVM, seis temas de pesquisa podem ser identificados: (1) Pesquisa em Análises de Cliente: Métodos e Questões Técnicas; (2) Pesquisa em Métodos de Aquisição de Clientes; (3) Pesquisa em Valor de Vida Útil do Cliente e Relações entre o valor da Firma; (4) Pesquisa em Implementação de Gestão de Valor ao Cliente; (5) Pesquisa em Determinantes de Retenção de Cliente e Expansão de Cliente; e (6) Pesquisa em Canais em Gestão de Valor ao Cliente (VERHOEF; DOORN; DOROTIC, 2007). Complementando a literatura de pesquisas sobre canais no contexto de gestão de valor do cliente, o varejo multicanal tem ganhado importância.

Canal é o ponto de contato, ou meio através do qual empresa e cliente interagem (NESLIN et al., 2006b). No contexto do varejo, canais desempenham a atividade de transferir produtos e serviços das empresas para os consumidores e para outras empresas (KUMAR; REINARTZ, 2006). De acordo com a tecnologia correspondente e as respectivas funções, os canais podem ser categorizados em quatro grupos: (1) canais *offline*, que incluem lojas físicas e catálogos; (2) canais *online*, que incluem *e-mails* e *websites*; (3) canais *mobile*, que incluem *websites mobile* e aplicativos; e (4) outros pontos de contato, como mídia social, boca a boca, anúncio e promoções (LIU; LOBSCHAT; VERHOEF, 2018).

O conjunto de atividades no qual varejistas vendem produtos ou serviços através de mais de um canal é chamado varejo multicanal (BECK; RYGL, 2015; LEVY; WEITZ, 2009). O gerenciamento multicanal no contexto do varejo é um elemento-chave do CRM (PAYNE; FROW, 2005; VERHOEF; DOORN;

DOROTIC, 2007). Varejos multicanais oferecem maior acesso a marcas, produtos e serviços, quando comparados a varejistas com um único tipo de canal, criando maior variedade de escolhas de compra, pagamentos e entregas aos clientes (LIU; LOBSCHAT; VERHOEF, 2018). Adotar estratégias multicanais aumentam a conveniência dos clientes no processo de compras e concede às empresas uma oportunidade de explorar sua ampla base de clientes (KUMAR; BEZAWADA; TRIVEDI, 2018).

Em um contexto multicanal, os dados deveriam estar integrados entre os canais para fornecer uma visão completa em relação às atividades dos clientes, bem como facilitar o marketing personalizado (NESLIN et al., 2006c). A identificação do cliente em cada canal permite individualizar a oferta de marketing e abre a oportunidade de vender produtos complementares que empresas concorrentes talvez não estejam aptas a ofertar (KUMAR; REINARTZ, 2006). O maior benefício dos dados ao nível individual da jornada de compra é entender como os pontos de contato no decorrer da jornada impactam outras variáveis de longo prazo relacionadas ao cliente, como satisfação, lealdade, valor do tempo de vida, patrimônio líquido de clientes e de marcas (KANNAN; REINARTZ; VERHOEF, 2016).

Para entregar um modelo de negócio satisfatório em um contexto no qual uma grande quantidade de dados está disponível para análise, é fundamental que o varejista desenvolva estratégias multicanais que se encaixem nas necessidades dos clientes e otimizem a vida útil dos mesmos (VERHOEF; DOORN; DOROTIC, 2007). Estratégias de vários canais com alto grau de integração podem fornecer promoções, produtos e preços consistentes em todos os canais (YRJÖLÄ; SAARIJÄRVI; NUMMELA, 2018). As atividades de marketing são influentes na migração de clientes em todos os canais (ANSARI; MELA; NESLIN, 2008). A migração para o canal *online* em particular pode ser útil para o varejista. Como o comportamento de navegação *online* dos clientes pode ser facilmente mapeado, este tipo de canal concede ao varejista a oportunidade de capturar dados potencialmente mais ricos do que os do canal tradicional (VERHOEF et al., 2010).

Na literatura sobre varejo multicanal, alguns estudos se concentraram em abordar como as estratégias de marketing podem ser aplicadas neste contexto. São

discutidas questões como, por exemplo, criação de estratégias multicanais bem-sucedidas no ambiente de varejo (ZHANG et al., 2010), se preços e promoções deveriam ser diferentes entre os canais (GREWAL et al., 2010), estratégias de sortimento por canal (BERRY et al., 2010) e como os varejistas aproveitam as estratégias de vários canais em relação às propostas de valor do cliente (YRJÖLÄ; SAARIJÄRVI; NUMMELA, 2018). Varejos multicanais devem implementar estratégias de marketing para atrair e reter clientes e aumentar as vendas (LIU; LOBSCHAT; VERHOEF, 2018). Neslin et al. (2006b) e Neslin e Shankar (2009) dão uma visão geral da pesquisa nesse campo.

Dentro da literatura acadêmica de varejo multicanal, é apresentado o conceito de “gestão do cliente multicanal”, definido como o desenho, a coordenação, o desenvolvimento e avaliação dos canais para melhorar o valor do cliente através de aquisição, retenção e desenvolvimento efetivos de cliente (NESLIN et al., 2006b). Dentro dos estudos de gestão de valor do cliente no contexto de varejo multicanal, foram desenvolvidos estudos relacionados a percepção do comportamento do consumidor (DHOLAKIA et al., 2010), segmentação e perfil do consumidor (KEEN et al., 2004; KONUS; VERHOEF; NESLIN, 2008) e seleção e migração de canais (ANSARI; MELA; NESLIN, 2008; VENKATESAN; KUMAR; RAVISHANKER, 2007) em um ambiente de varejo multicanal.

Importantes contribuições foram feitas neste campo, como identificação da incidência e do valor de vários segmentos de compras (THOMAS; SULLIVAN, 2005) e a abordagem analítica para prever mudanças nas escolhas de canais ao longo do tempo (ANSARI; MELA; NESLIN, 2008; SULLIVAN; THOMAS, 2004). Os estudos revelam uma diversidade considerável em termos de como os consumidores percebem e utilizam os diferentes canais. Quando interagem com um canal em particular, os consumidores podem apresentar diferentes comportamentos que podem ser influenciados pela natureza particular do canal e vice-versa (DHOLAKIA et al., 2010). Clientes que interagem através de canais que criam vínculos sociais ou econômicos mais fortes têm níveis maiores de satisfação e comprometimento, e também de compra cruzada (BOLTON; LEMON; VERHOEF, 2004).

Outros estudos investigaram gastos e valor do cliente em contextos multicanais (BOLTON; LEMON; VERHOEF, 2004; KUMAR; BEZAWADA; TRIVEDI, 2018; KUMAR; REINARTZ, 2016; KUSHWAHA; VENKATESH SHANKAR, 2007; MELIS et al., 2016; VENKATESAN; KUMAR; RAVISHANKER, 2007). Há estudos que evidenciam que o valor de um cliente individual depende do canal de aquisição devido às diferentes taxas de retenção em cada canal (KEANE; WANG, 1995). No contexto do varejo, ser multicanal concede ao varejista a oportunidade de melhorar a lucratividade do cliente oferecendo a ele uma variedade de opções transacionais (VERHOEF et al., 2010).

Um estudo empírico investigou os efeitos moderadores que podem estimular ou reter o aumento da participação nos gastos de um varejo de supermercados e confirmou que compradores multicanais aumentam a participação dos gastos nas visitas *online* na rede, e o tamanho desse aumento depende das características da rede e dos consumidores (MELIS et al., 2016). Alguns estudos sugerem que compras multicanais podem levar a uma maior lucratividade do cliente (KUSHWAHA; VENKATESH SHANKAR, 2007; VENKATESAN; KUMAR; RAVISHANKER, 2007). Estudos dentro deste campo comprovaram que clientes multicanais geram maiores receitas para as empresas, compram mais itens, compram em mais categorias e compram mais frequentemente do que compradores de um único canal (KUMAR; BEZAWADA; TRIVEDI, 2018; THOMAS; SULLIVAN, 2005; VENKATESAN; KUMAR; RAVISHANKER, 2007).

Um estudo investiga antecedentes de compradores multicanal usando dados comportamentais passados e vigentes e quantifica as consequências da compra multicanal ao longo de três dimensões: gasto, frequência de visita e lucratividade do cliente. O resultado do estudo é que compras multicanais tem efeitos positivos nos gastos dos clientes, frequência de visita e lucratividade do cliente (KUMAR; BEZAWADA; TRIVEDI, 2018)

Ao investigar os efeitos das compras multicanal e atividades de compra na lucratividade do cliente através de informações longitudinais de transações de clientes, Venkatesan, Kumar e Ravishanker (2007) ratificaram que compras multicanais estão relacionadas a maiores lucratividades do cliente. Os achados sugerem que, para aumentar a lucratividade, as empresas devem incentivar clientes

a comprarem em múltiplos canais. Gerar uma experiência multicanal ao cliente tem o potencial de melhorar dois aspectos cruciais do CRM: retenção e crescimento de cliente.

2.3

Aumento de valor do cliente

Valor do cliente é uma das métricas essenciais no gerenciamento de clientes e que ajuda as empresas a direcionarem seus esforços de marketing com maior confiança (KUMAR; REINARTZ, 2006). Quando a empresa entende o valor individual do cliente em relação a outros clientes, ela pode alocar seus recursos com maior efetividade (PEPPERS; ROGERS, 2011). As empresas podem aumentar o valor da sua base de clientes atraindo novos clientes, aumentando a retenção dos clientes e criando crescimento de valor dos clientes (VERHOEF; DOORN; DOROTIC, 2007).

O aumento de valor dos clientes pode ser feito com a venda de mais produtos ou serviços a um cliente, com a atualização dos clientes para níveis mais altos de serviço, com o aumento do uso de serviços pelos clientes e com a adoção pelos clientes de produtos ou serviços recém-desenvolvidos (BOLTON; LEMON; VERHOEF, 2008). É muito difícil listar todos os determinantes do crescimento de valor de clientes, uma vez que esse conceito é multidimensional e o número de estudos em cada uma das dimensões é muito limitado (VERHOEF; DOORN; DOROTIC, 2007). Verhoef et al. (2007) consideram cinco métricas para crescimento de clientes: (1) venda-cruzada, (2) uso de serviço, (3) valorização, (4) adoção de novos produtos/serviços e (5) participação do cliente/participação da carteira.

Participação do cliente x em um fornecedor y é o percentual de produtos ou serviços comprados pelo cliente x em um mercado específico ou em uma categoria definida no fornecedor y . Participação da carteira do cliente x em um fornecedor y é o percentual de dinheiro gasto pelo cliente x em um mercado ou em uma categoria definida no fornecedor y (VERHOEF; DOORN; DOROTIC, 2007). Para os varejistas, aumentar sua participação nos gastos totais de seus principais clientes é uma importante meta (MÄGI, 2003). A participação dos clientes é considerada uma

das melhores métricas para lealdade no comportamento do cliente (FOURNIER; YAO, 1997; VERHOEF, 2003).

No varejo, a participação nos gastos e a participação das visitas às lojas são os dois aspectos de participação do cliente e, embora ambos estejam correlacionados, os fatores que determinam onde os clientes fazem a maior parte de suas compras em volume nem sempre são os mesmos fatores que determinam quais lojas visitam com mais frequência (MÄGI, 2003). O objetivo de um varejo é garantir atividades repetidas de compras futuras para gerar receitas futuras (KUMAR; PANSARI, 2016).

Garantir as atividades de compras futuras é um dos objetivos das ações de relacionamento (VENKATESAN; KUMAR, 2004b). Muitos estudos mostram uma relação positiva entre satisfação e medidas de intenção de recompra (ANDERSON; FORNELL; LEHMANN, 1994; BOULDING et al., 1993; FORNELL, 1992; JONES; MOTHERSBAUGH; BEATTY, 2000), enquanto outros sugerem que os clientes satisfeitos não exibem necessariamente um alto nível de comportamento de recompra (MITTAL; KAMAKURA, 2001; REICHHELD, 1993).

O comportamento de clientes pode ser influenciado por atividades de marketing e, conseqüentemente, pode influenciar no valor da empresa. Os instrumentos de marketing estão relacionados ao CLV através dos custos da empresa para produzir os comportamentos necessários. As ações podem ser de marketing indireto ou direto. Marketing direto engloba comunicações direcionadas entre vendedor e cliente (BOLTON; LEMON; VERHOEF, 2004).

O efeito do marketing direto promocional na extensão do relacionamento funciona indiretamente, mediado pela amplitude do relacionamento, e pode ter um efeito negativo no uso de longo prazo do serviço, mediado pelo comprometimento (BOLTON; LEMON; VERHOEF, 2004). Estruturando suas atividades de marketing em torno de atividades que ajudam a estabilizar o comportamento de gasto dos clientes, as empresas podem reduzir a volatilidade do fluxo de caixa da empresa (SHAH et al., 2016).

Estudos empíricos analisaram como o marketing influencia o valor da empresa através do nível de fluxos de caixa futuros do cliente (BOLTON; LEMON;

VERHOEF, 2004; GUPTA; LEHMANN; STUART, 2004; KUMAR; SHAH, 2009; RUST; LEMON; ZEITHAML, 2004). Rust et al. (2004) defendem que as empresas devem considerar os efeitos dos gastos das estratégias de marketing no seu patrimônio líquido de clientes, relacionando as melhorias no patrimônio líquido de clientes aos gastos demandados para isso. O patrimônio líquido de clientes é definido como o valor total descontado do tempo de vida somado de todos os clientes vigentes e potenciais.

Um outro estudo desenvolve um modelo que relaciona o patrimônio líquido de clientes à capitalização de mercado. Os resultados indicam que a relação entre ambos é moderada pelos fatores de risco em forma de volatilidade e vulnerabilidade de fluxo de caixa dos clientes. Os achados mostram que uma estrutura baseada em patrimônio líquido de cliente pode, seguramente, prever a capitalização de mercado de uma empresa e que estratégias de marketing direcionadas a aumentar o patrimônio líquido de clientes aumentam o preço das ações da empresa e movimentam as expectativas de mercado (KUMAR; SHAH, 2009).

Outros estudos mostraram como características comportamentais de clientes, em alguns casos influenciadas por ações de marketing, podem afetar o valor do cliente como ativo (BOLTON; LEMON; VERHOEF, 2004; SHAH et al., 2016). No estudo de Shah et al. (2016) é feita uma análise empírica com dados de uma das maiores empresas de varejo dos Estados Unidos, com cinco tipos de dados: (1) dados transacionais de clientes, (2) dados de marketing, (3) dados de margem de produtos, (4) dados de características de clientes e (5) dados macroeconômicos. Os achados mostram que iniciativas de marketing ao nível do cliente tem efeitos positivos e significativos nas compras.

O estudo de Bolton et al. (2004) apresenta uma estrutura integrada chamada CUSAMS (*Customer Asset Management of Service*) proposta para ser o ponto de partida de um conjunto de proposições em relação a como seis categorias de instrumentos de marketing – (1) preço, (2) programas de qualidade em serviço, (3) promoções de marketing direto, (4) instrumentos de relacionamento de marketing, (5) anúncios e (6) canais de distribuição - influenciam o comportamento do consumidor em um relacionamento com a empresa, influenciando assim o valor do cliente. Os pesquisadores afirmam que os diferentes comportamentos de compra

são reflexo da extensão, da profundidade e da amplitude do relacionamento da empresa com o cliente. A extensão corresponde à retenção, a profundidade é refletida na frequência do uso do serviço ao longo do tempo e a amplitude se reflete em acréscimo de compra ou *cross-selling*.

2.3.1

Cross-selling como ferramenta de aumento de valor do cliente

O *cross-selling* é um importante motivador de comportamento de compra multicanal (KUMAR; VENKATESAN, 2005) e de valor do tempo de vida do cliente (VENKATESAN; KUMAR, 2004a). *Cross-selling*, ou venda cruzada, é a prática de vender novos produtos ou serviços para um consumidor existente, e o objetivo dessa prática é oferecer o produto certo, para o consumidor certo, na hora certa, usando o canal de comunicação certo para obter sucesso de longo prazo (LI; SUN; MONTGOMERY, 2011).

Entender os padrões de compra de cada consumidor entre várias categorias de produtos ou serviços é necessário para a implementação de uma estratégia eficiente de *cross-selling*, ou seja, o conhecimento sobre o comportamento de compra cruzada deve influenciar as estratégias de *cross-selling* (VERHOEF et al., 2010). Compra cruzada, ou *cross-buying*, é comprar produtos ou serviços adicionais de um fornecedor de serviços em adição aos que já são comprados correntemente (NGOBO, 2004).

A principal vantagem de identificar variáveis de interação entre empresa e cliente como motivadores de comportamento de compra cruzada é que empresas podem identificar clientes mais propensos a esse tipo de compra (KUMAR; GEORGE; PANCRAS, 2008). Um desafio gerencial é melhorar as taxas de resposta de campanhas deste tipo e evitar alcançar clientes com mensagens irrelevantes (LI; SUN; MONTGOMERY, 2011). A sinergia de interesses entre as ofertas da empresa (quando comparadas com as de seus concorrentes) e os desejos do cliente mantém o interesse do cliente em seguir em um relacionamento sem contrato com uma empresa (REINARTZ; KUMAR, 2000).

Em algumas empresas de serviços, o relacionamento geralmente não é contratual, envolvendo altos custos de troca (KUMAR; GEORGE; PANCRAS,

2008). Em relações contratuais, a empresa sabe quando a relação vai acabar e usufrui de um fluxo de caixa relativamente constante por um período de tempo (KUMAR; REINARTZ, 2016). Já em contextos não contratuais, o foco é prever atividades futuras dos clientes (VENKATESAN; KUMAR, 2004a). Otimizar o valor para o cliente, bem como fazer um bom uso dos dados de clientes para influenciá-los e prever a sua fidelidade, são ações relevantes no novo contexto de orientação para o mercado. Muitos segmentos, incluindo serviços financeiros, seguros, telecomunicações e varejo, colocam o *cross-selling* no topo das prioridades estratégicas (LI; SUN; MONTGOMERY, 2011).

Não é por acaso que empresas de serviço, incluindo o varejo, muitas vezes utilizam ferramentas e ações que estimulem o *cross-selling*. Este tipo de estímulo permite que empresas aumentem a receita de contribuição de clientes já existentes (KUMAR; GEORGE; PANCRAS, 2008). O *cross-selling* pode criar relacionamentos duradouros entre o cliente e a empresa, servindo como propaganda para a marca, sinal de qualidade e educando consumidores sobre a oferta de produtos (LI; SUN; MONTGOMERY, 2011).

Taxa de *cross-selling*, participação dos clientes e até influência de outros clientes podem ser considerados orientadores econômicos do modelo de tempo de vida de um cliente, com diferentes graus de importância, de acordo com o modelo do negócio (PEPPERS; ROGERS, 2011). Kumar e Reinartz (2003) mediram empiricamente, em uma relação não contratual de cliente e empresa, a duração do tempo de vida do cliente. Em relações não contratuais, o cliente tem a liberdade de comprar produtos e serviços simultaneamente de diferentes empresas (KUMAR; REINARTZ, 2016). Os autores concluem em seu estudo que a rentabilidade da duração do tempo de vida dos clientes está positivamente relacionada com os níveis de gastos, posse pelo cliente de algum instrumento de lealdade da empresa, quantidade de esforços de contato, renda do cliente, e também com o grau de comportamento de *cross-selling* do cliente (REINARTZ; KUMAR, 2003).

Estudos identificaram importantes motivadores de *cross-selling*, como, por exemplo, o valor total do pagamento, a satisfação, os esforços de marketing e os canais de aquisição (KUMAR; GEORGE; PANCRAS, 2008; VERHOEF; DONKERS, 2005; VERHOEF; FRANCES; HOEKSTRA, 2001). Pesquisas

indicam que promoções efetivas de marketing direto influenciam positivamente o *cross-selling* (BOLTON; LEMON; VERHOEF, 2004; VERHOEF; FRANCES; HOEKSTRA, 2001). Verhoef et al. (2001) defendem que a utilidade subjetiva de cada cliente depende primeiramente de sua satisfação corrente e percepção de preço. Assim, na decisão de compra de serviços adicionais, os clientes consideram a qualidade dos serviços fornecidos e o preço do serviço. As percepções de preço refletem na construção do patrimônio líquido de pagamento e as percepções da qualidade do serviço refletem em satisfação, e ambos afetarão o *cross-selling*.

A percepção de qualidade do consumidor influencia na sua atitude em relação a determinado produto e serviço. Quanto mais positiva a atitude e o comportamento, e quanto maior o nível de conectividade, seja entre clientes, entre clientes e funcionários e de ambos com a empresa, maior o nível de engajamento. Há um impacto positivo de altos níveis de engajamento com a performance da firma, e o nível de engajamento pode ser aprimorado pela identificação de estratégias relevantes implementadas (KUMAR; PANSARI, 2015).

Um outro estudo contribuiu com as discussões ao investigar empiricamente, usando avaliações de acionistas, as consequências relacionadas ao valor de comportamentos de engajamento de clientes estimulados pela empresa. Descobriu-se que iniciativas de empresas de engajar clientes, em média, diminuem o valor de mercado, o que é provável porque os acionistas são sensíveis ao risco de que essas iniciativas não funcionem (BECKERS; VAN DOORN; VERHOEF, 2018).

A venda de produtos adicionais pode gerar impactos negativos no relacionamento empresa-cliente. Tentativas de vendas direcionadas com frequência e erradamente podem despertar a indignação do cliente e, no pior caso, fazer com que o cliente rompa o relacionamento com a empresa. Adicionalmente, tentativas mal sucedidas de aumentar a gama de produtos com o cliente são sinônimo de má alocação de recurso (VERHOEF et al., 2010). Alcançar clientes com os produtos certos e no momento certo pode levar a uma melhoria considerável na qualidade do relacionamento empresa-cliente (KUMAR; VENKATESAN; REINARTZ, 2008).

Há ainda dentro da literatura estudos que investigaram respostas a campanhas de *cross-selling* (LI; SUN; MONTGOMERY, 2011; RAMSEY;

BAPNA, 2016). Um estudo mostrou que mineração de texto de anotações livres de serviços prestados ao cliente, como uma forma de segmentá-los, pode resultar em modelos preditivos que são tão bons quanto, ou melhores, do que modelos desenvolvidos usando apenas dados estruturais (RAMSEY; BAPNA, 2016). Outro estudo propôs um modelo de resposta que reconhece a evolução da demanda dos clientes por vários produtos, a possibilidade dos papéis multifacetados das solicitações de *cross-selling* para promoção, anúncio e educação, e a preferência heterogênea do cliente por canais de comunicação (LI; SUN; MONTGOMERY, 2011).

Ainda no contexto das pesquisas acadêmicas, há um estudo (NETZER; LATTIN, 2005) que propõe um modelo que facilita o uso de dados transacionais para avaliar a efetividade das ações de marketing de relacionamento, bem como encontros de clientes com marcas nas dinâmicas do relacionamento com o cliente e o subsequente comportamento de compra. O modelo proposto é o HMM (Hidden Markov Model). O HMM relaciona o estado latente do relacionamento com o comportamento de compra observado. O modelo permite examinar métodos através dos quais a empresa pode alterar o nível de relacionamento do cliente e, conseqüentemente, afetar o comportamento de compra de longo prazo.

Analisar a estimativa do HMM pode ajudar as empresas na avaliação da efetividade das diferentes ações de marketing para levar os clientes a um maior nível de relacionamento com a empresa (NETZER; LATTIN, 2005). Posteriormente, um estudo usou a abordagem de Netzer (2005) para avaliar a eficácia dos encontros dos clientes com as marcas e o comportamento de compra subsequente no contexto do varejo (LI; LI; JIA, 2006).

Outros estudos buscaram identificar clientes com maior probabilidade de *cross-buy* (NGOBO, 2004; VERHOEF; FRANCES; HOEKSTRA, 2001). Estudos identificaram o canal de aquisição do cliente (VERHOEF; DONKERS, 2005), o tipo de serviço (NGOBO, 2004), o número total de serviços obtidos em períodos anteriores (VERHOEF; FRANCES; HOEKSTRA, 2001), os níveis de gasto dos lares, características demográficas como gênero, educação, renda (LI; SUN; WILCOX, 2005) e idade (VERHOEF; FRANCES; HOEKSTRA, 2001) como fatores de impacto nas compras cruzadas. Programas de lealdade e o número de e-

mails diretos em períodos anteriores também são determinantes importantes das compras cruzadas (VERHOEF; FRANCES; HOEKSTRA, 2001).

Um estudo investigou o modelo preditivo do melhor jeito de vender o produto certo, ao cliente certo, no tempo certo (KUMAR; VENKATESAN; REINARTZ, 2008). Os achados indicam que adotar uma campanha de venda focada no cliente pode aumentar significativamente os lucros da empresa e os retornos sobre os investimentos. Adicionalmente, os resultados mostram que campanhas focadas melhoram o relacionamento entre empresa e cliente.

Um dos estudos, que tinha como objetivo identificar os motivadores de compra cruzada em um contexto não contratual, escolheu o serviço de varejo para tal investigação (REINARTZ et al., 2008). Nesse contexto, os pesquisadores propuseram uma estrutura para análise dos motivadores que tinham como influenciadores do comportamento de compra cruzada os seguintes fatores: características de troca, esforços de marketing, características dos clientes e características dos produtos. Os resultados ao nível cliente poderiam ser observados pelos aumentos nas receitas e aumento em margem de contribuição por pedido, bem como pelo aumento no número de pedidos em um dado período de tempo, o que gera como resultado final aumento em CLV (KUMAR; GEORGE; PANCRAS, 2008).

Reinartz et al. (2008) utilizaram dados transacionais das interações entre cliente e uma grande empresa de varejo para avaliar os motivadores de compra cruzada. O estudo evidencia que, dentro das características dos clientes, renda dos lares e idade são preditores relevantes. Já no contexto de características dos produtos, os resultados mostram que clientes que nas suas primeiras compras adquiriram categorias para casa, crianças, homens e mulheres demonstraram maior propensão à compra cruzada àqueles que compraram produtos para uso externo ou categorias de bagagem. Adicionalmente, a pesquisa mostrou que métricas baseadas em clientes são significativamente diferentes para os diferentes níveis de compra cruzada, como as médias de receita e média de pedidos.

3 MÉTODO

A partir de uma abordagem quantitativa, utilizou-se uma base de dados de clientes de uma grande empresa de varejo no setor supermercadista, criada e disponibilizada com exclusividade para este estudo pela equipe do supermercado em questão. Bases de dados internas de clientes de empresas podem fornecer informações relevantes para análises de comportamento do consumidor.

Alguns estudos empíricos que usam grandes bases de dados de clientes analisam como as empresas podem aumentar suas taxas de retenção, compra cruzada e participação do cliente (KUMAR; VENKATESAN; REINARTZ, 2008; VERHOEF, 2003) e como as firmas podem prever essas métricas (NESLIN et al., 2006a). Há outros estudos que focaram, especificamente, em como as empresas podem otimizar e influenciar o valor do cliente (RUST; VERHOEF, 2005; VENKATESAN; KUMAR; BOHLING, 2007).

Este estudo propõe um modelo de classificação de clientes de alto valor e de clientes que respondem a uma ação de *cross-selling*. Para isso, foi necessário tratar cada cliente de maneira individualizada no processo de análise. Quando segmentos individuais de clientes que exibem um ganho são identificados, as empresas podem implementar iniciativas de CRM e ações de marketing customizadas para cada cliente individualmente ou para cada segmento de clientes (KUMAR; REINARTZ, 2016).

A partir de critérios pré-estabelecidos, foram feitas análises dos dados coletados e compilados através da tabulação dos indicadores. Ao final do estudo, são apresentadas as conclusões do cruzamento das informações levantadas e discutidas na pesquisa, de modo a oferecer resultados empíricos e teóricos. A seguir, são apresentados o tipo de pesquisa, os dados coletados, os tratamentos aplicados aos dados, critérios para definição das variáveis utilizadas nos modelos, o processo de amostragem - incluindo as características da população estudada, o

método de coleta de dados e os procedimentos utilizados para tratar e analisar os dados das amostras - e a aplicação dos algoritmos de *machine learning*.

3.1

Tipo de Pesquisa

Este estudo é dividido em duas etapas. Em ambas, utilizou-se uma abordagem quantitativa de análise aplicando-se modelos de *machine learning* de classificação a uma base de dados secundários extraídos de bancos de dados de um supermercado brasileiro. Os dados refletem uma amostra representativa do universo de compradores de um supermercado com operação em lojas físicas e *online*. Na primeira etapa, buscou-se encontrar modelos que expliquem clientes de alto valor. Na segunda etapa do estudo, analisou-se a resposta de clientes a uma ação de relacionamento, a fim de se compreender modelos de clientes que respondem a uma ação de *cross-selling*.

Para a primeira etapa do estudo, a amostra se divide em dois grupos: (i) clientes de alto valor e (ii) outros clientes. Já para a segunda etapa, dois novos grupos foram criados: (iii) respondem à ação e (iv) não respondem à ação. Os dados foram dispostos como uma matriz de informação em que cada linha da base analisada, chamada de instância, representa um único cliente, identificado a partir do “código do cliente”, cada um associado a um único CPF. Para rodar modelos de classificação, é necessário que se defina a classe. Classe é a variável-alvo sobre a qual pretende-se aprender para que futuros dados possam ser classificados pela máquina como pertencentes ou não à classe em questão. As demais variáveis disponíveis para cada instância referem-se a características dos respectivos clientes, avaliadas pelos algoritmos e chamadas atributos.

Para a classe definiu-se a variável binária “sim” / “não”. Na primeira etapa do estudo, essa variável assume o valor “sim” para o grupo de “clientes de alto valor” (i) e não para “outros clientes” (ii). Seguindo a mesma lógica, na segunda etapa definiu-se como classe a variável binária “sim” para o grupo “responde à ação de relacionamento” (iii) e “não” para o grupo “não responde à ação de relacionamento” (iv).

Para as duas etapas do estudo, utilizou-se um *software* específico para desenvolver modelos de aprendizado de máquinas: o WEKA. Aplicou-se seis tipos de algoritmos de classificação, que serão indicados posteriormente. De maneira geral, esses algoritmos avaliam os atributos de cada instância para desenvolver um modelo que explique a sua respectiva classe.

3.2

Dados coletados

Todos os dados foram coletados de uma base de clientes de um varejo brasileiro supermercadista com operação em lojas físicas e *online*, referente ao período de setembro de 2017 a setembro de 2019. A base original é composta de inúmeras variáveis. Parte das variáveis coletadas para esta pesquisa basearam-se em um estudo empírico elaborado com dados de uma das maiores empresas de varejo dos Estados Unidos (SHAH et al., 2016). Foram utilizados nesse estudo, que tinha com o objetivo relacionar marketing ao comportamento do cliente e comportamento do cliente ao nível futuro do fluxo de caixa e volatilidade do cliente, cinco tipos de dados - (1) dados transacionais de clientes, (2) dados de marketing, (3) dados de margem de produtos, (4) dados de características de clientes e (5) dados macroeconômicos.

Três dos cinco tipos de dados utilizados no estudo de Shah et al (2016) foram coletados para esta pesquisa: dados transacionais (que compuseram a maior parte da base de dados) - frequência de visita, receita, uso de serviços na loja, margem de contribuição do cliente e perfil de produtos comprados; dados de marketing - se o cliente foi ou não impactado pela ação de marketing de relacionamento disparada e os gastos e volumes adquiridos através desta ação específica; e dados de características de clientes - variáveis demográficas como idade, bairro de moradia e gênero do cliente. Adicionalmente, também foram coletados dados de interação, como elogios e reclamações do cliente no SAC da empresa.

Os dados capturados existiam tanto para lojas físicas quanto para a operação *online*, sendo possível, portanto, mapear quais clientes compram em ambos os canais ou exclusivamente em um ou em outro. Na tabela 3.1 estão dispostas as variáveis de interesse coletadas em fontes secundárias do varejista, especificando o

tipo de dado e se o mesmo consta ou não nos dados apresentados no estudo de Shah et al. (2016).

Tabela 3.1 – Tabela de Variáveis Coletadas

Dado	Comentário	Tipo de dado	Apresentado no estudo de Shah (2016)?
Código do cliente	Todas as informações do banco de dados do varejista estão associadas ao código do cliente. Por isso, este foi utilizado como código identificador de cada cliente neste estudo.	Dados de características de clientes	Sim
Gênero	Masculino ou Feminino.		
Bairro moradia	Bairro no qual o cliente reside.		
Data de Nascimento	Dia, mês e ano em que o cliente nasceu. Este dado foi utilizado para calcular a idade do cliente. Bairro da loja na qual o cliente mais compra.		
Localização da loja	Bairro da loja na qual o cliente mais compra.	Dados transacionais	Sim
Margem % de contribuição do cliente para a rede varejista	O valor percentual da margem de contribuição do cliente, calculada pela rede varejista.		
Gasto total com consumo local nas lojas físicas	Valor gasto pelo cliente com alguns serviços e produtos oferecidos dentro da loja para consumo local.		
Em quantas lojas físicas comprou	Quantidade de lojas físicas em que o cliente		

	comprou no respectivo mês.		
Quantas vezes usou serviço <i>delivery</i> na loja física	Quantidade de vezes em que o cliente utilizou o serviço de <i>delivery</i> dentro da loja física.		
Data da última compra	Dado que informa qual foi a última data em que o cliente comprou em alguma loja da rede.		
Frequência na loja (nas lojas físicas e <i>online</i>)	Informa qual a frequência média do cliente nas lojas da rede varejista para realizar compras.		
Quantidade de transações (nas lojas física e <i>online</i>)	Informa a quantidade de transações realizadas pelo cliente em suas visitas à loja.		
Gasto nas compras (nas lojas física e <i>online</i>)	Indica qual o gasto mínimo, médio e máximo do cliente com as compras no respectivo mês.		
Quantidade de itens na cesta (nas lojas física e <i>online</i>)	Indica qual a quantidade mínima, média e máxima de itens na cesta do cliente no respectivo mês.		
Perfil dos itens na cesta (nas lojas física e <i>online</i>)	Avaliar se o cliente compra mais itens exclusivos (definidos assim pela própria rede de supermercado, como itens que são exclusivos ao sortimento deste varejista) ou itens <i>commodities</i> (definidos		

	pela rede varejista como itens de maior competitividade por preço).		
Reclamou no SAC	Quantas vezes o cliente reclamou no SAC sobre qualquer aspecto durante sua interação com os serviços e suas experiências junto à rede varejista.	Dados de interação	Não
Elogiou no SAC	Quantas vezes o cliente elogiou no SAC sobre qualquer aspecto durante sua interação com os serviços e suas experiências junto à rede varejista.		
Última atualização de cadastro	Data em que o cliente fez a última atualização dos dados cadastrais no banco de dados da rede varejista.		
Uso da ação de marketing	Meses em que o cliente fez uso da ação de marketing de relacionamento avaliada neste estudo.	Dados de marketing	Sim
Valor gasto na ação de relacionamento	Valor total do gasto médio do cliente cada vez que utilizou o programa de relacionamento avaliado.		
Quantidade de itens comprados na ação de relacionamento	Quantidade média de itens comprados pelo cliente cada vez que utilizou o programa de relacionamento avaliado.		

Todos os dados foram combinados em uma identificação única do cliente. Os dados fundidos representam valiosos detalhes em termos de características dos clientes, dados de compra e iniciativas de marketing no período de setembro de 2017 a setembro de 2019.

3.3

Procedimentos de tratamento e variáveis utilizadas

Toda a base, contendo mais de 250 mil clientes identificados pela rede varejista, foi disponibilizada, mês a mês, do período de setembro de 2017 a setembro de 2019, totalizando 25 bases mensais. Considerando que a unidade de análise é o cliente, e os mesmos códigos precisam estar presentes nas bases de cada um dos meses para a consolidação da base final utilizada na análise, optou-se nesse estudo por não utilizar as 25 bases, a fim de se obter uma base final nas duas etapas com maior quantidade de clientes analisados para obtenção de melhores resultados.

Na primeira etapa do estudo, foram considerados somente clientes que estavam presentes nas bases de todos os meses de 2019 (janeiro a setembro). Desta forma, a base totalizou 161.489 clientes. Já na segunda etapa do estudo, foram considerados meses do ano de início da ação (2017) e do ano mais recente das bases disponibilizadas (2019). Assim, o período de outubro de 2017 a dezembro de 2017 foi considerado como o período sem uso da ação, e o período dos três meses mais recentes da base, julho de 2019 a setembro de 2019, como o período de uso da ação de relacionamento.

Algumas informações extraídas dos bancos de dados do varejista foram utilizadas diretamente na base de dados final, sobre a qual os modelos foram gerados. Outras variáveis foram criadas a partir de dados existentes. Para definir as variáveis finais a serem utilizadas nas duas etapas do estudo levou-se dois critérios em consideração: ter os tipos de variáveis apresentadas na seção 3.2 deste capítulo – variáveis transacionais, de interação, demográficas e de marketing – bem como alguns dos atributos que direcionam o valor do cliente em empresas B2C (*Business to Consumer*), definidos em um estudo de Kumar e Reinartz (2016).

A tabela 3.2 contém todas as variáveis presentes nas bases finais de cada uma das etapas do estudo e suas respectivas descrições, bem como a identificação do tipo de dado, do tipo da variável (se categórica ou numérica) e do respectivo atributo direcionador do valor do cliente do estudo de Kumar e Reinartz (2016).

Tabela 3.2 – Variáveis base final das duas etapas do estudo

Tipo de dado	Tipo de variável	Nome da variável	Descrição	Atributo estudo Kumar e Reinartz (2016)
demográfico	categórica	GÊNERO	M/F	Gênero
demográfico	numérica	IDADE	Idade no último mês do período analisado (set/19).	Idade
transacional	numérica	MÉDIA FREQ. MENSAL LOJA DE JAN A SEP 2019	Média de janeiro a setembro de 2019 da frequência mensal nas lojas físicas.	Intensidade de uso do negócio
transacional	numérica	MÉDIA GASTO CONSUMO LOCAL NA LOJA DE JAN A SEP 2019	Média de janeiro a setembro de 2019 do gasto com alimentos, bebidas e outros produtos para consumo local disponíveis para venda nas lojas físicas.	Níveis de gastos passados do cliente
interação	categórica	RECLAMA COM FREQUÊNCIA	SIM (reclamou em mais de 65% dos meses, ou seja, 6 ou mais meses dos 9 analisados) / NÃO (não reclamou em nenhum dos meses, ou reclamou em 5 ou menos meses dos 9 analisados).	Não há atributo associado a este no estudo de Kumar e Reinartz (2016)
interação	categórica	ELOGIA COM FREQUÊNCIA	SIM (elogiou em mais de 65% dos meses, ou seja, 6 ou mais meses dos 9 analisados) / NÃO (não elogiou em nenhum dos meses, ou elogiou em 5 ou menos meses dos 9 analisados).	Não há atributo associado a este no estudo de Kumar e Reinartz (2016)
transacional	categórica	USA <i>DELIVERY</i> COM FREQUÊNCIA	SIM (usou serviço de <i>delivery</i> na loja física em mais de 65% dos meses, ou seja, 6 ou mais meses dos 9 analisados) / NÃO (não usou serviço de <i>delivery</i> na loja física em nenhum dos meses, ou elogiou em 5 ou menos meses dos 9 analisados).	Intensidade de uso do negócio
transacional	categórica	CONSOME LOCALMENTE COM FREQUÊNCIA	SIM (usou serviço de consumo local na loja física em mais de 65% dos meses, ou seja, 6 ou mais meses dos 9 analisados) / NÃO (não usou serviço de consumo local na loja física em nenhum dos meses, ou usou em 5 ou menos meses dos 9 analisados).	Intensidade de uso do negócio
transacional	categórica	MAIOR TENDÊNCIA A COMPRAR PRODUTOS EXCLUSIVOS	SIM (em mais de 50% dos meses comprou apenas produtos que são exclusivos ao sortimento daquela rede, ou seja, em 5 ou mais meses dos 9 analisados) / NÃO (nunca comprou apenas produtos exclusivos ao sortimento daquela loja, ou fez isso em	Comportamento de compra focado

			menos de 5 dos 9 meses analisados).	
transacional	categórica	MAIOR TENDÊNCIA A COMPRAR PRODUTOS <i>COMMODITIES</i>	SIM (em mais de 50% dos meses comprou apenas <i>commodities</i> – itens de categorias com maior competitividade de preço - ou seja, 5 ou mais meses dos analisados) / NÃO (nunca comprou apenas produtos <i>commodities</i> - itens de categorias com maior competitividade de preço - ou fez isso em menos de 5 dos 9 analisados).	Comportamento de compra focado
transacional	categórica	CLIENTE MULTICANAL	SIM (em mais de 50% dos meses comprou no <i>online</i> e na loja física, ou seja, 5 ou mais meses dos 9 analisados) / NÃO (nunca ou em 4 ou menos meses dos 9 analisados comprou no <i>online</i> e na loja física).	Comportamento multicanal
Classe Etapa 1	Categórica	CLIENTE ALTO VALOR	Essa é a variável que define o status do cliente: cliente de alto valor ou outros clientes	-
Classe Etapa 2	Categórica	RESPONDE A AÇÃO	Essa é a variável que define o status do cliente: cliente responde ou não à ação	-

3.4 População e amostra

O varejista em questão tem uma média de 650 mil clientes frequentando suas lojas físicas e *online* mensalmente. Deste total, em torno de 250 mil clientes são identificados e constituem uma rica base de dados que foi compartilhada neste estudo. Na primeira etapa, a proposta é modelar clientes de alto valor. Valor do cliente é a quantificação monetária de um cliente individual para uma empresa (HUNT; ROBERT M. MORGAN, 1995).

A escolha do modelo para entender e mensurar o valor dos clientes é decidido pela disponibilidade de volume e variedade dos dados, tempo e recursos técnicos, bem como o uso intencional da CLV e, independentemente do modelo escolhido, os resultados devem ser úteis para a tomada de decisões gerenciais que levem a mudanças em tempo real (KUMAR; REINARTZ, 2016). A rede varejista já identifica os clientes de maior e menor valor na base utilizando a regra de RFM (Recência, Frequência e Valor Monetário).

A RFM é usada como uma ferramenta de segmentação para classificar segmentos de clientes como baixo RFM ou alto RFM (KUMAR; REINARTZ,

2016). Comportamentos de compra vigentes dependem de comportamentos de compra passados. Essa característica explica o uso comum do RFM (Recência, Frequência e Valor Monetário) nos métodos de segmentação dentro de bases de marketing (BOLTON; LEMON; VERHOEF, 2004).

Adicionalmente a essa classificação, constava também na base de dados compartilhada a margem de contribuição de cada cliente. Kumar e Reinartz (2012, p. 4) definem valor do cliente como “o valor econômico da relação do cliente com a firma, expresso com base na margem de contribuição ou lucro líquido”. Desta forma, considerou-se a margem de contribuição e o gasto médio mensal por compra de cada cliente para criar as duas amostras da primeira etapa do estudo: (i) “clientes de alto valor” e “outros clientes” (ii). Como as duas variáveis selecionadas para a construção das amostras poderiam variar no mês a mês, para o cálculo das variáveis finais, considerou-se a média das margens de contribuição e a média dos gastos médios mensais de cada cliente nos 9 meses contemplados na análise.

Clientes com a média da margem de contribuição acima do valor mínimo estabelecido pela regra de negócio da empresa - que por questões de confidencialidade não serão apresentadas, mas pode-se afirmar que está acima da média trabalhada pelo setor supermercadista - e com a média do gasto médio mensal por compra superior ou igual a R\$300,00, foram considerados na amostra do grupo (i), que com essas premissas totalizou 1.882 clientes. O segundo tipo de amostra consiste em outros clientes, que não são de alto valor. Nesta base do grupo (ii) encontram-se os outros 159.607 clientes que, junto aos 1.882 clientes de alto valor, totalizam os 161.489 clientes da base completa.

Em algumas linhas de clientes havia um ou outro caso de campos demográficos como gênero e idade sem preenchimento, ou com dados de idade identificados como não estando corretos. Todas essas linhas foram retiradas da base, chegando-se a uma base final de 154.864 clientes no grupo (ii) e 1.792 clientes no grupo (i). A performance dos algoritmos de *machine learning* é avaliada, normalmente, usando-se acurácia preditiva. No entanto, não é apropriada quando a base de dados está desequilibrada, ou seja, se as categorias de classificação não estão representadas, aproximadamente, de maneira igual (CHAWLA et al., 2002)

Com 154.864 clientes no grupo (ii) e 1.792 clientes no grupo (i), as categorias de classificação do estudo em questão ficariam desequilibradas, ou seja,

as categorias de classificação (clientes de alto valor e outros clientes) não estariam representadas de maneira proporcional, uma vez que na base de clientes do grupo (ii) havia quase 87 vezes mais clientes do que na base de clientes do grupo (i), ocorrendo *oversampling*. Isto quer dizer que a amostra está acima da frequência mínima de amostragem, determinada pelo critério de Nyquist-Shannon, no qual a frequência da amostragem deve ser o dobro da amostra considerada (FARROW; KIM; BILLINGE, 2011). A fim de equilibrar a base de “outros clientes” com a amostra de “clientes de alto valor” para a análise final, foi aplicado o método de amostra aleatória simples na base de outros clientes, chegando-se a uma amostra com 3.584 no grupo (ii), totalizando 5.376 clientes, juntando-se as duas amostras.

O terceiro e quarto tipos de amostras, utilizadas na segunda etapa do estudo, consistem em clientes impactados pela ação de relacionamento e que fizeram uso da oferta de *cross-selling* proposta pela ação. O cliente foi impactado através de um *e-mail* marketing e, por meio de cadastro do CPF, poderia escolher um determinado número de produtos para colocar na sua cesta com um percentual de desconto previamente definido pela rede de supermercados.

A ação teve início em abril de 2017 e, nos primeiros três meses, 800 clientes foram impactados. Após esse período, a rede de supermercados expandiu pouco a pouco a base de clientes impactados pela ação, chegando a 9% de conversão após o período de um ano. Assim, para gerar a base que originou as amostras dos grupos (iii) e (iv), agrupou-se o período equivalente a dois anos de ação, de outubro de 2017 a setembro de 2019, totalizando 96.973 clientes.

Posteriormente, mapeou-se os clientes que fizeram uso da ação no período analisado e, nesta base, que totalizava 5.675 clientes, selecionou-se àqueles que ficaram os 3 últimos meses do primeiro ano de informações compartilhadas (de outubro de 2017 a dezembro de 2017) sem fazer uso da ação e que, após serem impactados, fizeram uso da ação, ao menos, nos 3 meses mais recentes das bases compartilhadas, ou seja, de julho 2019 a setembro de 2019. Com estas premissas e mais a limpeza de campos de clientes sem informação demográfica, de gênero ou dados de idade incorretos, chegou-se a um total de 2.135 clientes.

Para o terceiro tipo de amostra, grupo (iii), selecionou-se dentro da base de 2.135 clientes os que responderam à ação de relacionamento dentro do período

analisado. Já para o quarto tipo de amostra, considerou-se os clientes que não responderam à ação. A ação de relacionamento construída e disparada pela rede de supermercados tinha como principal objetivo aumentar o valor do cliente através do aumento dos gastos e incremento do volume de itens por compra. Portanto, para definir clientes que responderam ou não à ação, usou-se como premissa a regra de negócio da rede varejista, que seria ter incremento de 3 unidades ou mais na média de itens comprados mensalmente no período de uso da ação comparado ao período sem uso da ação, e ter variação positiva na média dos gastos médios mensais no período de uso da ação comparado ao período sem uso da ação.

A regra de negócio do varejista está de acordo com a definição de compra cruzada feita em estudos anteriores, definida como o número total de diferentes categorias de produtos que um cliente comprou desde o período inicial (KUMAR; GEORGE; PANCRA, 2008; REINARTZ; KUMAR, 2003; VENKATESAN; KUMAR, 2004a). No entanto, no caso da regra do varejista o período inicial considerado não foi a primeira compra do cliente na rede, mas compras que antecederam o uso da ação de *cross-selling* implementada.

Dentro das premissas aqui apresentadas, chegou-se a um total de 610 clientes para o grupo (iii) e, a fim de se obter um equilíbrio na base que foi utilizada para rodar o modelo final, foi feita uma seleção por amostragem aleatória simples na base de clientes que foram impactados pela ação, mas não responderam, para que se chegasse aos 1.220 clientes do grupo (iv). O conjunto final de dados consiste em 1.830 clientes, que representam um elevado percentual da receita total gerada pelos clientes que fizeram uso da ação de relacionamento no período observado. A tabela 3.3 sumariza o total de clientes em cada grupo de amostra com as respectivas premissas.

Tabela 3.3 – Total de clientes em cada amostra e premissas

		Total Clientes	Premissas
ETAPA 1	Base utilizada para as amostras	161.489	Clientes que compraram em cada um dos 9 meses mais recentes das informações compartilhadas (janeiro de 2019 a setembro de 2019).
	Amostra grupo (i): Clientes de Alto Valor	1.792	Clientes com margem de contribuição > do que o mínimo estabelecido pela rede varejista e gasto médio mensal > ou = R\$300,00.
	Amostra grupo (ii): Outros Clientes	3.584	Clientes com margem de contribuição < do que o mínimo estabelecido pela rede

			varejista e gasto médio mensal < R\$300,00.
ETAPA 2	Base utilizada para as amostras	2.135	Clientes que não fizeram uso da ação de outubro de 2017 a dezembro de 2017 e fizeram uso da ação, ao menos, de julho 2019 a setembro de 2019.
	Amostra (iii): Clientes que responderam à ação de <i>cross-selling</i>	610	Clientes com incremento de 3 unidades ou mais na média de volume de itens na cesta de compras e com incremento na média dos gastos, comparando os meses mais recentes em que fez uso da ação (julho, agosto e setembro de 2019) com meses em que não fez uso da ação (outubro, novembro e dezembro de 2017).
	Amostra (iv): Clientes que não responderam à ação de <i>cross-selling</i>	1.220	Clientes sem incremento ou com queda de unidades na média de volume de itens na cesta de compras e sem incremento ou queda na média dos gastos, comparando os meses mais recentes em que fez uso da ação (julho, agosto e setembro de 2019) com meses em que não fez uso da ação (outubro, novembro e dezembro de 2017).

3.5

Aplicação de algoritmos de *Machine Learning*

Uma vez coletados os dados e sendo devidamente tratados, de acordo com as premissas apresentadas previamente, estes foram consolidados em uma única base processada por um *software* estatístico denominado WEKA (AHER; L.M.R.J., 2011; HALL, 2009; SHARMA; JAIN, 2013). Este *software* é próprio para desenvolvimento de modelos de *machine learning*, permitindo criar, testar e comparar a qualidade dos modelos, além de possuir os algoritmos mais utilizados em processos de aprendizado de máquinas (HALL, 2009). Os algoritmos de classificação aplicados avaliam os atributos de cada instância e suas respectivas classes para desenvolver um modelo que as explique.

Considerando o tipo de problema que este estudo trabalha, ou seja, o de classificação, que implica em reconhecer um cliente como pertencente a um dos grupos pré-definidos (“clientes de alto valor” ou “outros clientes” e “respondem à ação ou “não respondem à ação”), a abordagem utilizada é conhecida como aprendizado supervisionado. Este tipo de abordagem se opõe a metodologias de aprendizado não supervisionados.

No aprendizado supervisionado, um modelo é criado a partir de dados já classificados, em que o algoritmo apenas aprende a relacionar os atributos àquela classe previamente definida. A meta do aprendizado supervisionado é construir um modelo conciso de distribuição de rótulos de classes dentro de características preditivas (KOTSIANTIS, 2007). Kotsiantis (2007) apresentou uma figura que descreve o processo de aplicação do *machine learning* supervisionado a um problema do mundo real, conforme figura 1:

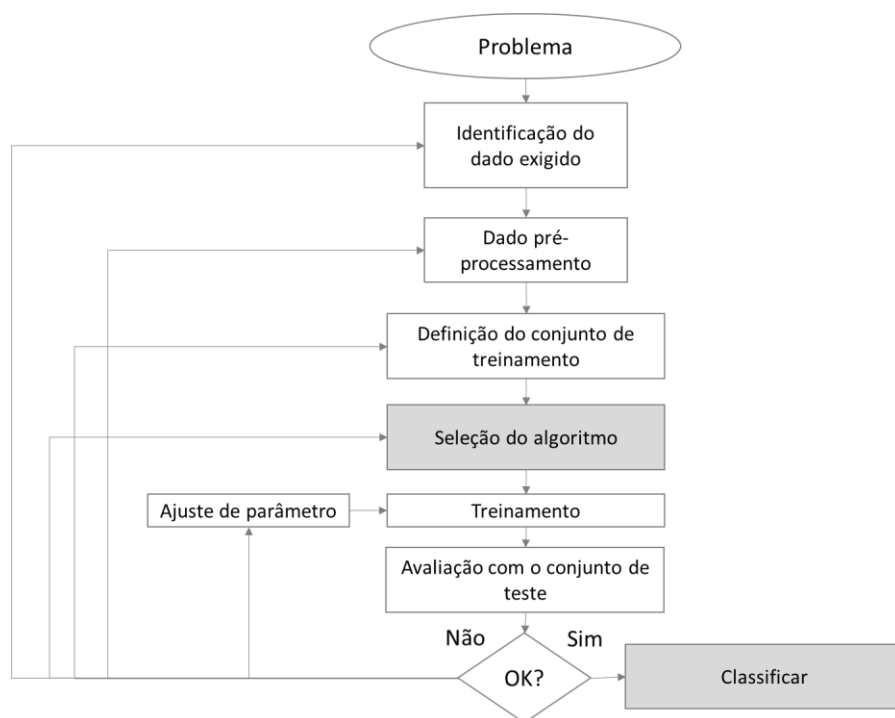


Figura 3.1: processo de aplicação do *machine learning* supervisionado (KOTSIANTIS, 2007)

Criar e treinar modelos é diferente de fazer previsões. Para criar e treinar modelos - caso de técnicas de aprendizado supervisionado, como neste estudo - utiliza-se uma base de instâncias com classe conhecida para que os algoritmos calculem e modelem as relações entre os atributos e a classe. Já no processo de previsão, também conhecido como *scoring*, o objetivo é aplicar um modelo previamente calculado sobre uma base de instâncias com classe desconhecida, mas que possua observações com os mesmos atributos utilizados na base que originou o modelo. O *output* do modelo é um *score* para cada instância. Este estudo se limitará, nas duas etapas, ao modelo supervisionado, ou seja, utilizará uma base de instâncias com classe conhecida para que os algoritmos calculem e modelem as relações entre os atributos e a classe.

As acurácias *out-of-sample* dos modelos e suas capacidades de generalização foram testadas com o método de validação cruzada, conhecido como *k-fold cross-validation*, com $k = 10$. Essa metodologia consiste em dividir a amostra em k grupos de igual tamanho, onde $k-1$ grupos são utilizados para estimar os parâmetros, e um grupo é usado para teste de estimativa de classe. Em seguida, troca-se o grupo de teste, usando os demais $k-1$ grupos para estimação. Esse procedimento é realizado k vezes e, ao fim, tem-se uma estatística da quantidade de previsões sobre a classe corretamente realizadas sobre o total de previsões. Essa estatística é utilizada como medida da acurácia de um modelo de classificação e, naturalmente, quanto maior seu valor, melhor é o poder preditivo do modelo em questão, assim como sua capacidade de generalização.

Para compreender a acurácia dos modelos, é importante analisar as matrizes de confusão geradas em cada um deles. Matrizes de confusão comparam as classes reais com as previsões realizadas sobre as instâncias da base classificada. Neste estudo, a classe usada é binária: sim/não. Na primeira etapa, se o cliente é de alto valor (sim) ou se ele é outro tipo de cliente (não); e na segunda etapa, se ele responde à ação de relacionamento (sim) ou se ele não responde à ação de relacionamento (não). Portanto, obtém-se uma matriz 2×2 , a exemplo da matriz representada na tabela 3.4, retirada de um dos algoritmos dos resultados deste trabalho.

Tabela 3.4. – Exemplo de matriz de confusão

Realidade	Previsão		Total
	A - SIM	B - NÃO	
A - SIM	1.110	682	1.792
B - NÃO	187	3.397	3.584
Total	1.297	4.097	n=5.376

Cada número na tabela representa contagens da amostra. Neste exemplo, a amostra é igual a 5.376, sendo 1.792 instâncias pertencentes a classe “sim” e 3.584 instâncias pertencentes a classe “não”. Adicionalmente, também é possível enxergar na matriz os resultados das previsões.

Desta forma, relacionando as informações, é possível observar que das 1.792 instâncias positivas, 1.110 foram classificadas corretamente, ou seja, são “verdadeiro positivo”. Já 682 instâncias foram classificadas como negativo, ou seja,

“falso negativo”, ou erro tipo I. Analogamente, das 3.584 instâncias pertencentes a classe “não”, 3.397 foram classificadas corretamente, sendo consideradas, assim, como “verdadeiro negativo”. Já 187 foram classificadas incorretamente, sendo “falso positivo”, ou erro tipo II.

Já a acurácia total de um modelo é dada pela razão entre as previsões verdadeiras (verdadeiro positivo + verdadeiro negativo) e o tamanho total da amostra. Nesse caso, $(1.110 + 3.397) / 5.376 = 83,84\%$. Além da acurácia total do modelo, outras estatísticas podem ser inferidas a partir da análise da matriz de confusão. Abaixo segue um breve descritivo das medidas utilizadas neste estudo para fins comparativos:

Sensibilidade, *Recall* ou Taxa de Verdadeiro Positivo: Mede o percentual de previsões de positivos, quando a classe real é “positivo”. No exemplo acima, $1.110/1.792 = 61,94\%$.

Especificidade ou Taxa de Verdadeiro Negativo: Mede o percentual de previsões de negativo quando a classe real é “negativo”. No exemplo acima, $3.397/3.584 = 94,78\%$.

Precisão: É a taxa de acertos quando a previsão é “positivo”. Considerando o exemplo acima, calcula-se conforme segue: $1.110/1.297 = 85,58\%$

Nas duas etapas deste estudo foram considerados seis algoritmos de classificação: regressão logística (PENG; LEE; INGERSOLL, 2002; WILSON; LORENZ, 2015), árvore de decisão, árvore de decisão J48 - também conhecido como C4.5 (KOTSIANTIS, 2007; QUINLAN, 1996), *K-Nearest Neighbors* (ALTMAN, 1992), *Support Vector Machines* (CORTES; VAPNIK, 1995; FRADKIN; MUCHNIK, 2006) e *Naive Bayes* (RISH, 2001). Todos foram rodados e seus resultados comparados, com maior aprofundamento em algoritmos de melhor acurácia: regressão logística (PENG; LEE; INGERSOLL, 2002; WILSON; LORENZ, 2015), árvore de decisão e árvore de decisão J48.

O modelo de regressão logística tem a capacidade de indicar como as chances de ocorrência de um evento binário, chamadas de *odds ratio*, são afetadas por um determinado atributo. Se a cada 10 clientes de alto valor 6 são mulheres, então a probabilidade de um cliente de alto valor ser mulher é 0,6, e a probabilidade complementar, é de 0,4 e, neste caso, a *odds ratio* seria a razão entre ambos, ou

seja, 0,6/0,4. Isso significa que mulheres tem 1,5 vezes mais chances de serem *clientes de alto valor* do que homens.

Sendo a *odds ratio* uma razão entre probabilidades, que oscilam entre 0 e 1, os valores variam de 0 a infinito positivo. Quando o resultado da *odds ratio* é um número entre 0 e 1, significa que o atributo ao qual o resultado se refere está associado a uma menor chance de ocorrência da variável dependente. Em oposição, resultados maiores do que 1 refletem que o atributo aumenta as chances de ocorrência da classe, sendo maior esta chance quanto mais alto for o valor do resultado. Já quando o resultado é igual a 1 ou muito próximo disso, significa que o atributo não tem influência sobre a classe.

Em algoritmos de árvore de decisão, os dados são divididos em ramos, também conhecidos como *splits*, e são classificados a partir de regras (nós). Na ponta de cada ramo, novas regras podem ser aplicadas, criando novos nós que levam a novos *splits*. Quando um nó não é mais subdividido é chamado de folha. O primeiro nó é conhecido como nó raiz e os nós intermediários, nós de decisão. A figura 3.2 ilustra uma árvore de decisão genérica para fins de exemplo.

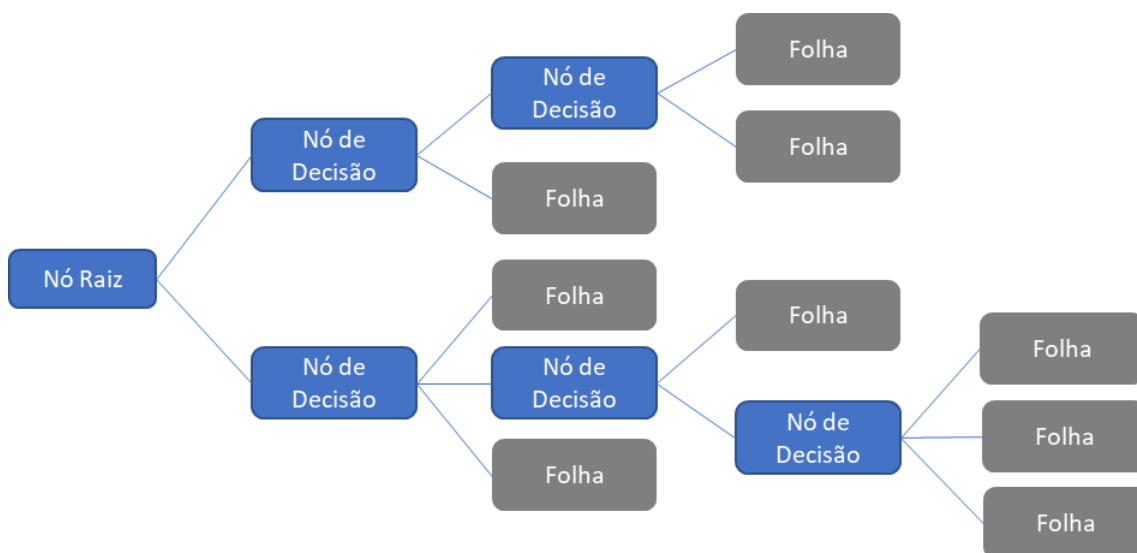


Figura 3.2. – Exemplo de estrutura de árvore de decisão.

Uma vez criada a estrutura da árvore, novos dados, não classificados, realizam o caminho dos ramos de acordo com as regras impostas em cada nó. Os ramos podem ter tamanhos diferentes e os nós podem ser divididos em dois ou mais

grupos. Quando chega na folha, o dado é identificado como pertencente a uma das classes.

3.6 Limitações do estudo

Uma das limitações do estudo é que os custos com as ações de marketing não foram compartilhados pela rede varejista, impossibilitando que métricas de CLV fossem aplicadas ou calculadas. Desta forma, para definir a amostra de clientes de alto valor, o estudo considerou a regra de negócio do varejista, que é baseada em métricas e modelos já apresentados em estudos acadêmicos, como RFM (BOLTON; LEMON; VERHOEF, 2004; FADER; HARDIE; LEE, 2005) e margem de contribuição (KUMAR; REINARTZ, 2012).

Além disso, algumas variáveis consideradas para os atributos dos modelos possuem nomenclaturas que são particulares à rede varejista em questão. Fazem parte deste grupo as seguintes variáveis: tendência a comprar produtos *commodities* e tendência a comprar produtos exclusivos. O termo *commodities* abrange produtos e categorias com maior competitividade de preço dentro do setor. Já o termo “produtos exclusivos” refere-se a produtos que são exclusivos ao sortimento da rede de supermercados considerada neste estudo.

Outra limitação do estudo é que não foi possível ter acesso a outras ações de relacionamento com o cliente além da ação aqui considerada. Isso limita compreender possíveis impactos no comportamento do cliente causados por outras ações de relacionamento praticadas pela rede varejista, além da ação avaliada neste estudo. Adicionalmente, alguns motivadores já identificados como sendo relevantes na influência de comportamento de compra cruzada em estudos acadêmicos anteriores, como a satisfação, os esforços de marketing e os canais de aquisição (KUMAR; GEORGE; PANCRAS, 2008; VERHOEF; DONKERS, 2005; VERHOEF; FRANCES; HOEKSTRA, 2001) não constavam na base de dados, o que pode ter gerado impacto na precisão dos modelos gerados.

4 ANÁLISE E RESULTADOS

Este capítulo será dividido em duas seções: (1) análise e resultados gerais dos seis modelos e (2) análise dos modelos com maior acurácia. Na seção referente à análise e resultados gerais dos modelos, são apresentados os algoritmos gerados para as análises e os resultados gerais, incluindo a acurácia de cada modelo das duas etapas do estudo. Já na seção 2 deste capítulo, são explicados em mais detalhes cada algoritmo com modelos de melhor acurácia e os resultados obtidos com cada um nas duas etapas do estudo.

4.1 Análise e resultados gerais dos modelos

4.1.1 Modelos gerados

Para a análise dos dados nas duas etapas do estudo, considerou-se alguns modelos específicos de *machine learning*. A tabela 4.1 apresenta os algoritmos utilizados e seus respectivos nomes no *software* WEKA.

Tabela 4.1. – Nomes dos algoritmos para referência

Algoritmo	Nome Weka
Regressão Logística	Logistic
Árvore de Decisão	REPTree
Árvore de Decisão J48	J48
K-Nearest Neighbours	IBk
<i>Support Vector Machine</i>	SMO
Naïve Bayes	N.Bayes

4.1.2 Acurácia dos modelos

A partir de matrizes de confusão, as acurácias dos modelos são calculadas. Matrizes de confusão comparam as classes reais com as previsões realizadas sobre as instâncias da base classificada. Neste estudo, a classe usada é binária: sim/não.

4.1.2.1

Acurácia dos modelos - Etapa 1

Na tabela 4.2 estão dispostas as matrizes de confusão de cada um dos modelos, obtidos pelo método de validação cruzada 10 vezes. Neste caso apresentado, “sim” representa a classe positivo, daqueles que são “clientes de alto valor”, e “não” representa a classe negativo, dos clientes que não estão no grupo de “clientes de alto valor”.

Tabela 4.2. – Matrizes de confusão dos modelos calculados na etapa 1

LOGISTIC	A - SIM	B - NÃO	Total	SMO	A - SIM	B - NÃO	Total
A - SIM	1.110	682	1.792	A - SIM	1.132	660	1.792
B - NÃO	187	3.397	3.584	B - NÃO	258	3.326	3.584
Total	1.297	4.079	5.376	Total	1.390	3.986	5.376

J48	A - SIM	B - NÃO	Total	IBk	A - SIM	B - NÃO	Total
A - SIM	1.102	690	1.792	A - SIM	1.203	589	1.792
B - NÃO	186	3.398	3.584	B - NÃO	678	2.906	3.584
Total	1.288	4.088	5.376	Total	1.881	3.495	5.376

REPTree	A - SIM	B - NÃO	Total	N. Bayes	A - SIM	B - NÃO	Total
A - SIM	1.149	643	1.792	A - SIM	1.544	248	1.792
B - NÃO	256	3.328	3.584	B - NÃO	1.355	2.229	3.584
Total	1.405	3.971	5.376	Total	2.899	2.477	5.376

Abaixo, na tabela 4.3, o resumo da acurácia, da sensibilidade, da especificidade e da previsão dos modelos calculados na etapa 1 deste estudo.

Tabela 4.3. – Resumo dos resultados de desempenho dos modelos na etapa 1

	<i>Logistic</i>	<i>J48</i>	<i>REPTree</i>	<i>SMO</i>	<i>IBk</i>	<i>N. Bayes</i>	Média
Acurácia do modelo	83,84%	83,71%	83,28%	82,92%	76,43%	70,18%	80,06%
Sensibilidade	61,94%	61,50%	64,12%	63,17%	67,13%	86,16%	67,34%
Especificidade	94,78%	94,81%	92,86%	92,80%	81,08%	62,19%	86,42%
Precisão	85,58%	85,56%	81,78%	81,44%	63,96%	53,26%	75,26%
cliente alto valor/total	24,13%	23,96%	26,13%	25,86%	34,99%	53,92%	31,50%

De maneira geral, os seis modelos apresentaram bons resultados de acurácia, sendo todos superiores a 70%. O modelo *Logistic* e os dois modelos de árvore de decisão (*J48* e *REPTree*) obtiveram os mais altos índices de acurácias dentre os algoritmos testados. Notadamente, estes três algoritmos demonstraram também o melhor desempenho na capacidade de prever corretamente a classe de “clientes de alto valor”, com os seguintes resultados de precisão: 85,58% - *Logistic*, 85,56% - *J48* e 81,78% - *REPTree*. Analisando o mesmo critério para os demais modelos, observa-se que o modelo SMO apresenta um bom desempenho também, pois prevê corretamente mais de 81% da classe de “cliente de alto valor” e possui acurácia de

82,92%. Já o IBk e o *N. Bayes* foram os que apresentaram menor acurácia, com 76,43% e 70,18% respectivamente. As taxas de precisão de ambos também foram as menores, com 63,96% para o IBk e 53,26% para o *N. Bayes*.

As taxas altas de especificidade de cinco dos seis modelos apresentados mostram baixo risco de se cometer o erro tipo I, ou seja, concluir erradamente que um cliente está no grupo de *outros clientes* quando, na realidade, ele é *cliente de alto valor*, demonstrando que um elevado percentual classificado como *outros clientes*, de fato, não é cliente de alto valor. Os resultados indicam que menos de 8% dos que estão no grupo de *clientes de alto valor* foram classificados incorretamente como *outros clientes* nos modelos *Logistic*, *J48*, *REPTree* e *SMO*.

Já o risco de se cometer um erro tipo II, ou seja, classificar um cliente do grupo de *outros clientes* como sendo um *cliente de alto valor* - o falso positivo, é inferior a 40% em todos os modelos, uma vez que a taxa de sensibilidade ou verdadeiro positivo é superior a 60% em todos os casos. Neste aspecto, o modelo que apresenta o menor risco é o *N. Bayes*. Observa-se, portanto, uma facilidade de identificar o cliente de alto valor baseando-se nos atributos utilizados neste estudo.

4.1.2.2

Acurácia dos modelos - Etapa 2

Na tabela 4.4 são dispostas as matrizes de confusão de cada um dos modelos, obtidos pelo método de validação cruzada 10 vezes. Aqui, “sim” representa a classe positivo, daqueles que “responderam à ação de relacionamento”.

Tabela 4.4 – Matrizes de confusão dos modelos calculados na etapa 2.

LOGISTIC	A - SIM	B - NÃO	Total
A - SIM	175	435	610
B - NÃO	136	1.084	1.220
Total	311	1.519	1.830

REPTree	A - SIM	B - NÃO	Total
A - SIM	170	440	610
B - NÃO	144	1.076	1.220
Total	314	1.516	1.830

J48	A - SIM	B - NÃO	Total
A - SIM	171	439	610
B - NÃO	160	1.060	1.220
Total	331	1.499	1.830

SMO	A - SIM	B - NÃO	Total
A - SIM	229	381	610
B - NÃO	214	1.006	1.220
Total	443	1.387	1.830

N. Bayes	A - SIM	B - NÃO	Total
A - SIM	293	317	610
B - NÃO	292	928	1.220
Total	585	1.245	1.830

IBk	A - SIM	B - NÃO	Total
A - SIM	259	351	610
B - NÃO	367	853	1.220
Total	626	1204	1.830

Abaixo, na tabela 4.5, o resumo da acurácia, da sensibilidade, da especificidade e da previsão dos modelos calculados.

Tabela 4.5. – Resumo dos resultados de desempenho dos modelos etapa 2

	<i>Logistic</i>	<i>REPTree</i>	<i>SMO</i>	<i>J48</i>	<i>N. Bayes</i>	<i>IBk</i>	<i>Média</i>
Acurácia do modelo	68,80%	68,09%	67,49%	67,27%	66,72%	60,77%	66,52%
Sensibilidade	28,69%	27,87%	37,54%	28,03%	48,03%	42,46%	35,44%
Especificidade	88,85%	88,20%	82,46%	86,89%	76,07%	69,92%	82,06%
Precisão	56,27%	54,14%	51,69%	51,66%	50,09%	41,37%	50,87%
Respondem à ação/valor total	16,99%	17,16%	24,21%	18,09%	31,97%	34,21%	23,77%

De maneira geral, os seis modelos não apresentaram altos resultados de acurácia, sendo todos inferiores a 70%. Uma possibilidade para tal cenário é que alguns motivadores já identificados em estudos anteriores como sendo relevantes na influência de comportamento de compra cruzada, como por exemplo e-mails previamente recebidos e canal de aquisição do cliente (VERHOEF; DONKERS, 2005; VERHOEF; FRANCES; HOEKSTRA, 2001), não constavam na base de dados. Tal situação pode ter impactado na acurácia dos modelos gerados.

No contexto geral, o modelo *Logistic* e o modelo de árvore de decisão *REPTree* obtiveram os melhores índices de acurácia dentre os algoritmos testados. Notadamente, estes dois algoritmos demonstraram também o melhor desempenho na capacidade de prever corretamente a classe de clientes que “respondem à ação”, com precisão de 56,27% e 54,14%, respectivamente. Analisando o mesmo critério para os demais modelos, observa-se que nenhum apresenta um bom desempenho, já que o *SMO* prevê corretamente apenas 51,69% e o *J48* apenas 51,66% dos clientes que respondem à ação. Enquanto o *N. Bayes* e o *IBk* não preveem corretamente nem 51% dos clientes que respondem à ação, apresentando as menores taxas.

As taxas de especificidade maiores do que 85% em três dos seis modelos apresentados mostram baixo risco de se cometer o erro tipo I, ou seja, concluir erradamente que um cliente está no grupo de *clientes que não respondem à ação* quando, na realidade, ele é um *cliente que responde à ação*, demonstrando que um elevado percentual que o modelo classifica como *clientes que não respondem à ação*, de fato, é um cliente que não respondeu à ação implementada. Os resultados indicam que menos de 15% dos que estão no grupo de *clientes que respondem à ação* foram classificados incorretamente como *clientes que não respondem à ação*.

Já o risco de se cometer um erro tipo II, o falso positivo, ou seja, classificar um cliente do grupo de *clientes que não respondem à ação* como sendo um *cliente que responde à ação*, é igual ou maior do que 52%, uma vez que a taxa de sensibilidade ou verdadeiro positivo é inferior a 49% em todos os casos, chegando a 27,87% no modelo *REPTree*, que apresentou a menor sensibilidade quando comparado aos demais. Neste aspecto, o modelo que apresenta o menor risco do erro tipo II é o *N.Bayes*. Observa-se, portanto, que não se pode identificar com facilidade o cliente que *responde à ação* baseando-se nos atributos utilizados neste estudo.

4.2

Análise dos modelos com maior acurácia

Os modelos de regressão logística e árvore de decisão foram os que apresentaram melhor acurácia nas duas etapas. Considerando isso, nesta seção, os resultados destes modelos serão apresentados com maior aprofundamento e detalhamento para ambas etapas deste estudo. Apesar de na segunda etapa do estudo o modelo J48 não ter ficado entre os três modelos com melhor acurácia, sendo superado pelo SMO, ele será considerado para aprofundamento a fim de se obter melhor comparação com os modelos da etapa 1 e com o algoritmo *REPTree* da etapa 2.

4.2.1

Análise e resultados do modelo de regressão logística - Etapa 1

A tabela 4.6 revela os atributos e suas respectivas *odds ratio* listadas na etapa 1 deste estudo.

Tabela 4.6. – Atributos e *odds ratio* estimados pelo modelo de regressão logística da etapa 1 para classe sim

Atributo	Odds Ratio
Gênero=M	0.8463
Idade	1.0084
Média da frequência mensal em loja de janeiro a setembro de 2019	0.7052
Média do gasto com consumo local de janeiro a setembro de 2019	1.0053
Reclama com frequência=NÃO	0.9608
Usa <i>delivery</i> com frequência=NÃO	0.0725
Consome localmente com frequência=NÃO	4.2771
Maior tendência a comprar produtos exclusivos=NÃO	10.2937
Maior tendência a comprar produtos <i>commodities</i> =NÃO	22.7788
Cliente é multicanal=NÃO	0.0311

A partir dos resultados apresentados, é possível inferir, ainda que com algum nível de incerteza, algumas informações. Na tabela observam-se 10 atributos dos 11 considerados para gerar o modelo e, nessa análise, a variável dependente é ser um cliente de alto valor. Os atributos “não reclamar com frequência”, “média dos gastos mensais com consumo local” e “idade” parecem não ter influência sobre a classe, diante dos resultados apresentados. Já “não ter uma maior tendência a comprar produtos *commodities*”, “não ter uma maior tendência a comprar produtos exclusivos” e “não consumir localmente com frequência” são atributos que aumentam a chance de ocorrência da classe.

Os maiores *odds ratio* foram para “não ter uma maior tendência a comprar produtos *commodities*” e “não ter uma maior tendência a comprar produtos exclusivos”. Ou seja, os dados indicam que os clientes de mais alto valor não têm tendência a comprar um ou outro tipo de perfil de produtos em particular, colocando variados perfis de categorias e produtos em sua cesta de compras. No estudo de Reinartz et al. (2008), os autores afirmam que compradores que se engajam em comprar mais de categorias variadas são mais engajados com o negócio, o que reflete maior retenção, maiores receitas, maior participação na carteira e maior lucratividade. Incrementar essas variáveis gera impacto direto no valor do cliente (REINARTZ; KUMAR, 2003; VERHOEF; DOORN; DOROTIC, 2007).

Em contrapartida, os atributos “não ser multicanal” e “não usar o *delivery* com frequência” pressupõem uma menor chance de ocorrência da variável dependente “cliente ser de alto valor”, apresentando o resultado de menor *odds ratio* o atributo “não ser multicanal”. Conclui-se, então que, se o cliente não é multicanal, ele tem menor probabilidade de ser um cliente de alto valor. Esse resultado corrobora com achados da literatura que concluíram que clientes multicanal tem maior probabilidade de serem de alto valor, pois há um efeito positivo desse perfil de cliente nos gastos e na lucratividade (KUMAR; BEZAWADA; TRIVEDI, 2018; KUMAR; SHAH; VENKATESAN, 2006; VENKATESAN; KUMAR; RAVISHANKER, 2007).

4.2.2

Análise e resultados do modelo de regressão logística - Etapa 2

A tabela 4.7 revela os atributos e suas respectivas *odds ratio* listadas na etapa 2 deste estudo.

Tabela 4.7. – Atributos e *odds ratio* estimados pelo modelo de regressão logística da etapa 2 para classe sim.

Atributo	Odds Ratio
Gênero=M	1.2926
Idade	0.9941
Média da frequência mensal em loja de janeiro a setembro de 2019	0.8998
Média do gasto com consumo local de janeiro a setembro de 2019	1.0011
Reclama com frequência=NÃO	1.9713
Usa delivery com frequência=NÃO	0.4512
Consome localmente com frequência=NÃO	1.4617
Maior tendência a comprar produtos commodities=NÃO	0.7537
Cliente é multicanal=NÃO	0.5145
CLIENTE DE ALTO VALOR=NÃO	0.5368

Ainda que com algum nível de incerteza, dado que a acurácia dos modelos estimados para esta etapa não foi alta, é possível inferir a partir dos resultados apresentados algumas informações. Observando-se a tabela 4.7, pode-se inferir pelos resultados que a “média dos gastos mensais com consumo local” e “idade” parecem não ter influência sobre a classe. A idade não ter influência sobre a classe diverge do achado do estudo de Kumar et al. (2008) que conclui que pessoas com idades intermediárias tem relação positiva com maiores níveis de compra cruzada.

Já “não reclamar com frequência” e “não consumir localmente com frequência” são atributos que pressupõem uma maior chance de ocorrência da variável dependente “cliente responder à ação de relacionamento”. O atributo com maior *odds ratios* é “não reclamar com frequência”. Pode-se inferir que clientes que não reclamam estão mais satisfeitos com o negócio. O resultado desse modelo corrobora com achados da literatura que sugeriram que maiores níveis de satisfação levam a maiores intenções de compra (LI; SUN; WILCOX, 2005) e diverge de achados que concluíram que compras cruzadas não se associam fortemente com satisfações relacionadas a experiências com serviços (NGOBO, 2004).

Em contrapartida, “não ser um cliente de alto valor”, “não ser multicanal” e “não usar serviços de *delivery*” da loja física com frequência são atributos que diminuem a chance de ocorrência da classe. Os dados indicam que os clientes com

menor propensão de resposta à ação de relacionamento implementada possuem essas características comportamentais. Não usar o serviço de *delivery* com frequência e não ser multicanal têm as menores *odds ratio*, indicando que clientes com esse perfil têm menos chances de resposta à ação de *cross-selling*.

Estudos anteriores já investigaram a influência de atributos similares no valor do cliente. Observou-se relações de efeitos direto no valor do cliente, tanto em relação ao uso de serviços (REINARTZ; KUMAR, 2003) como em relação ao comportamento de compra multicanal (KUMAR; BEZAWADA; TRIVEDI, 2018; KUMAR; SHAH; VENKATESAN, 2006), encontrando relação de efeito direto no valor do cliente.

Assim sendo, comparando-se os resultados de regressão logística das duas etapas desse estudo, observa-se que os dois atributos com as menores *odds ratio* na etapa 1, ou seja, os que indicam menor chance de o cliente ser de alto valor, são os mesmos atributos com as menores *odds ratio* na etapa 2 e, consequentemente, indicando menor chance de o cliente responder à ação de relacionamento. Ou seja, clientes que não são multicanal e não usam o serviço de *delivery* com frequência na loja física tem menor chance de serem clientes de alto valor e também menor chance de responder à ação de *cross-selling*.

Portanto, os atributos principais a indicar menor chance de o cliente ser de alto valor são também os principais, por este modelo, a indicar menor chance de o cliente responder à ação de relacionamento. Desta forma, pode-se inferir que clientes de alto valor tem maior chance de responder à ação de relacionamento de *cross-selling* aqui considerada. Não é por acaso que, na etapa 2 do estudo, ao considerar o atributo “cliente de alto valor” para gerar o modelo, este atributo aparece como sendo o terceiro com o menor resultado de *odds ratio*, indicando uma menor chance de o cliente que não tem alto valor responder à ação de relacionamento.

4.2.3

Análise e resultados dos modelos de árvore de decisão: Etapa 1

As figuras 4.1 e 4.2 retratam as árvores geradas pelos algoritmos J48 e REPTree respectivamente na etapa 1 deste estudo.

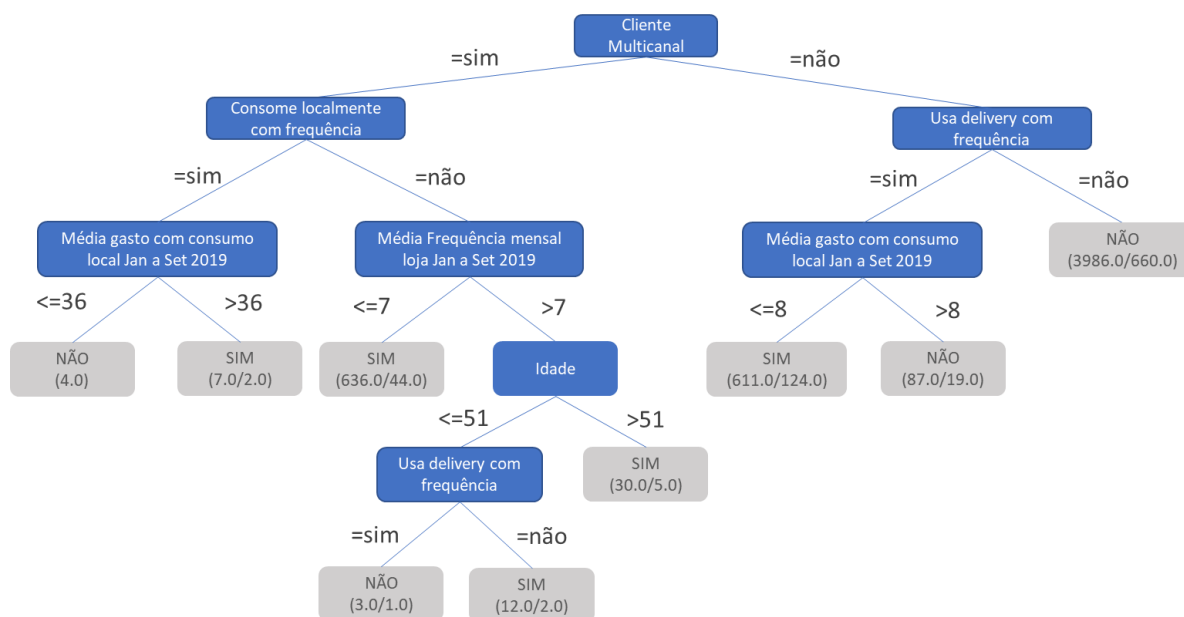


Figura 4.1. – Estrutura de árvore de decisão do algoritmo J48 - Etapa 1.

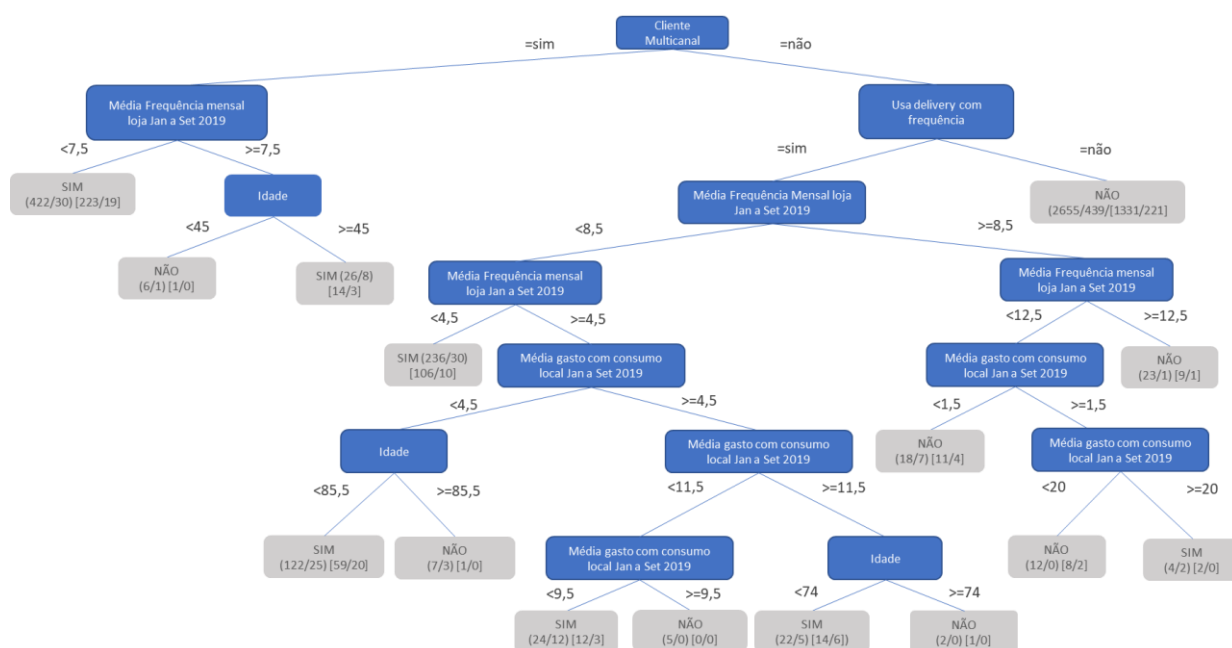


Figura 4.2. – Estrutura de árvore de decisão do algoritmo REPTree - Etapa 1.

Comparando-se os modelos, observa-se que o modelo REPTree gerou oito níveis enquanto o J48 gerou seis níveis e que, em ambos, dos onze atributos considerados, cinco foram vistos como relevantes para a classe. Os atributos observados em ambos são: cliente multicanal (compra na loja física e na loja

online), usa serviço de *delivery* na loja física (uso de serviços), média da frequência mensal na loja física (quantas vezes, em média, frequenta as lojas físicas dentro de um mês), média dos gastos com consumo local (gastos com serviços e produtos consumidos localmente) e idade.

Estudos anteriores dentro da literatura que aborda atributos relacionados ao valor do cliente já concluíram relações de causa e efeito de variáveis consideradas relevantes dentro dos modelos gerados pelos algoritmos das árvores na etapa 1 deste estudo. Investigou-se nesse contexto uso de serviços e aumento da lealdade do cliente (REINARTZ; KUMAR, 2003), efeitos da compra multicanal nos gastos e lucratividade dos clientes (KUMAR; BEZAWADA; TRIVEDI, 2018; VENKATESAN; KUMAR; RAVISHANKER, 2007) e frequência de visita e aumento da lealdade do cliente (KUMAR; BEZAWADA; TRIVEDI, 2018; REINARTZ et al., 2008).

Analisando os primeiros nós, pode-se tirar alguns *insights* para a rede varejista em questão. Tanto no algoritmo REPTree como no J48, o nó raiz é “cliente multicanal”, e divide-se em dois ramos: ser ou não ser multicanal. O *split* do nó raiz indica o atributo que o modelo entendeu como sendo o mais relevante para a definição da classe. Assim, considerando que a maior parte dos clientes de alto valor ficaram no ramo “sim” em ambos os modelos gerados pelos algoritmos das árvores, conclui-se que ser multicanal é um atributo relevante para ser um cliente de alto valor. Esse resultado corrobora com outros achados na literatura.

Estudos concluíram que compras multicanais podem levar a uma maior lucratividade do cliente (KUMAR; BEZAWADA; TRIVEDI, 2018; KUSHWAHA; VENKATESH SHANKAR, 2007; VENKATESAN; KUMAR; RAVISHANKER, 2007) e há estudos que evidenciaram que clientes multicanais geram maiores receitas para as empresas, compram mais itens, compram em mais categorias e compram mais frequentemente do que compradores de um único canal (THOMAS; SULLIVAN, 2005). Venkatesan, Kumar e Ravishanker (2007) apresentam como uma possível razão para clientes serem mais lucrativos quando compram em mais de um canal o fato de as empresas poderem fornecer serviços adicionais aos clientes através dos múltiplos canais, como pedir *online* e retirar na loja física. De acordo com os autores, os clientes são expostos a mais serviços

fornecidos pela empresa e, portanto, espera-se que fiquem mais satisfeitos e desenvolvam uma relação mais profunda com a empresa.

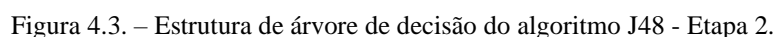
A perspectiva apresentada pelos autores corrobora com o que pode ser verificado ao analisar os dois primeiros níveis das duas árvores apresentadas neste estudo. Em ambos os modelos de árvore, o nó raiz é ser multicanal e o nó de decisão do nível seguinte ao ramo “não” é “uso do serviço de *delivery* com frequência na loja física”, indicando que, àqueles que não são multicanal e não usam o serviço de *delivery* na loja física, não são de alto valor. O mesmo é observado no modelo de regressão logística gerado na etapa 1 desse estudo, que apresentou o atributo “não ser multicanal” e “não usar o serviço *delivery*” como atributos que pressupõem uma menor chance de ocorrência da variável dependente “cliente ser de alto valor”. Pode-se concluir, então, que clientes que não utilizam serviços oferecidos pela rede varejista são menos propensos a serem de alto valor.

Já os que não são multicanais, mas usam o serviço de *delivery* na loja e possuem uma frequência média mensal na loja de até 4,5 vezes são clientes de alto valor, como pode ser observado no modelo REPTree. Analisando especificamente este modelo, o nó seguinte ao ramo sim do nó raiz é a média da frequência mensal do cliente na loja, indicando que, se o cliente é multicanal e tem uma frequência de até 7,5 vezes ao mês na loja física, ele é um cliente de alto valor. Uma vez que a frequência de compra está dentre os componentes do *Customer Lifetime Value* (KUMAR; PANSARI, 2016; VENKATESAN; KUMAR, 2004b), o resultado apresentado pelo modelo é pertinente.

4.2.4

Análise e resultados dos modelos de árvore de decisão: Etapa 2

As figuras 4.3 e 4.4 retratam as árvores geradas pelos algoritmos J48 e REPTree, respectivamente, na etapa 2 deste estudo.



Comparando-se os modelos, observa-se que o modelo J48 gerou seis níveis, enquanto o REPTree gerou treze níveis. Em ambos, os atributos “média de frequência mensal na loja física” - quantas vezes em média o cliente frequenta as lojas físicas dentro de um mês - e “cliente multicanal” - cliente com comportamento

de compra em loja física e em loja *online* - foram considerados relevantes para a classificação. Assim como nestes modelos, os atributos “cliente multicanal” e a “média de frequência mensal na loja física” também foram identificados como relevantes nos mesmos algoritmos para obtenção de modelo preditivo de clientes de alto valor, gerados na etapa 1 deste estudo.

Tanto no algoritmo REPTree como no J48, gerados na etapa 2, a média de frequência mensal na loja física é o *split* do nó raiz sendo, portanto, o atributo mais relevante para definir se o cliente responde ou não à ação. O estudo de Reinartz et al. (2008) mostra que comprar com mais frequência resulta em um aumento das categorias compradas, ou seja, altos níveis de lealdade levam a maior *cross-buy*. Neste estudo, os autores afirmam que compradores que se engajam em comprar mais de categorias variadas são mais engajados com o negócio, o que reflete maior retenção, maiores receitas, maior participação na carteira e maior lucratividade.

O *split* seguinte ao nó raiz, para se chegar ao maior número de clientes que respondem à ação de *cross-selling*, em ambos os modelos gerados pelos algoritmos das árvores na etapa 2, é ser cliente multicanal. Esse resultado confirma outros achados na literatura. Estudos indicam que clientes multicanais compram mais itens, compram em mais categorias e compram mais frequentemente do que compradores de um único canal (THOMAS; SULLIVAN, 2005).

Pessoas com idades intermediárias tem relação positiva com maiores níveis de compra cruzada (KUMAR; GEORGE; PANCRAS, 2008). Seguindo a sequência até a folha com o maior número de clientes na classe sim, no modelo REPTree, aparece o atributo idade, indicando que pessoas com 32 anos ou mais, com uma frequência média mensal em loja de até 6,5 e com gastos de até R\$52,00 com itens e serviços para consumo local tem maior chance de responder à ação de relacionamento implementada pela rede varejista.

Interessante observar também que, nos dois modelos de árvore, clientes que tem uma frequência de visita média mensal à loja física superior a 8 vezes não respondem à ação de compra cruzada. Tal resultado pode ser atribuído à hipótese de o cliente que vai mais vezes à loja comprar mais categorias e ter um maior ticket médio por compra, não tendo o incremento esperado na cesta de compras após o uso da ação. Clientes que fazem compras ocasionalmente têm níveis de interação

muito baixos com a empresa para desenvolver uma relação de confiança com esta (KUMAR; GEORGE; PANCRAAS, 2008), já clientes que tem maior nível de interação - e o atributo de frequência de visita pode ser um indicativo do nível de interação - tem maior chance de construir uma relação de maior confiança, levando a maior lealdade e, conseqüentemente, entregando maior valor à empresa.

Desta forma, a média de frequência mensal de visitas à loja física e ser multicanal são atributos relevantes para o cliente responder à ação de *cross-selling*, enquanto ser multicanal e fazer uso de serviço são atributos relevantes para modelar clientes de alto valor. Os resultados da etapa 1 do estudo, que investigou modelos preditivos de clientes de alto valor, corroboram com pesquisas anteriores que investigaram relações de causa e efeito dessas variáveis e valor do cliente, como lealdade e aumento dos gastos (REINARTZ et al., 2008; REINARTZ; KUMAR, 2003). Assim também os resultados da etapa 2 do estudo, que investigou modelos preditivos de clientes que respondem à ação de relacionamento de *cross-selling*, convergem com estudos de motivadores de comportamentos de compra cruzada (KUMAR; GEORGE; PANCRAAS, 2008). A tabela 4.8 sumariza os resultados desta pesquisa.

Tabela 4.8. – Resultados sumarizados

Algoritmo	Atributos mais relevantes			
	Cliente é de Alto Valor	Cliente não é de Alto Valor	Cliente Responde à Ação	Cliente não Responde à Ação
Regressão Logística	- Não ter tendência a comprar apenas produtos <i>commodities</i> - Não ter tendência a comprar apenas produtos exclusivos	- Não ser multicanal - Não usar serviço <i>delivery</i> com frequência	- Não reclamar com frequência - Não consumir localmente com frequência	- Não usar serviço <i>delivery</i> com frequência - Não ser multicanal
J48	- Ser multicanal - Não consumir produtos e	- Não ser multicanal	- Ter frequência média de visita	- Ter frequência média de visita mensal em loja

	serviços localmente com frequência - Ter frequência média de visita mensal em loja de até 7 vezes	- Não usar serviço <i>delivery</i> com frequência	mensal em loja de até 8 vezes - Ser multicanal	superior a 8 vezes
REPTree	- Ser multicanal - Ter frequência média de visita mensal em loja de até 7,5 vezes	- Não ser multicanal - Não usar serviço <i>delivery</i> com frequência	- Ter frequência média de visita mensal em loja de até 8,5 vezes - Ser multicanal	- Ter frequência média de visita mensal em loja superior a 8,5 vezes

5 CONCLUSÃO

Este trabalho teve como objetivo propor um modelo de classificação de clientes de alto valor e de clientes que respondem a uma ação de relacionamento de *cross-selling*. A partir de uma base de dados de clientes de uma grande empresa varejista no setor de supermercados, na primeira etapa do estudo verificou-se quais atributos - relacionados diretamente ao valor do cliente - são relevantes na classificação de modelos de clientes de alto valor e, na segunda etapa, na classificação de clientes que respondem a uma ação de relacionamento de compra cruzada.

Os modelos gerados na etapa 1 apresentaram, de maneira geral, uma boa acurácia. Observa-se, portanto, uma facilidade de identificar clientes de alto valor baseando-se nos atributos utilizados neste estudo. Já na segunda etapa, os modelos gerados não apresentaram uma acurácia elevada. Mesmo assim, é possível inferir algumas informações, ainda que com um certo nível de incerteza.

Ao analisar os modelos com melhor acurácia nas duas etapas do estudo, observou-se que o atributo multicanal é relevante tanto para classificar modelos de clientes de alto valor como clientes que respondem a uma ação de relacionamento de compra cruzada. Adicionalmente, os atributos “uso de serviço *delivery* na loja física” e “frequência de visita a lojas” também foram indicados por alguns dos modelos gerados como sendo relevantes para as classificações.

Além destes atributos, na primeira etapa do estudo, verificou-se pelo algoritmo de regressão logística que “não ter tendência a comprar produtos *commodities*” é o atributo mais relevante, seguido de “não ter tendência a comprar produtos exclusivos”. Os dados indicam que clientes com essas características têm maior chance de serem clientes de alto valor. Isso quer dizer que, na maior parte das compras de clientes de alto valor, não são colocados apenas produtos de

categorias com maior competitividade de preço na cesta de compras ou produtos que sejam exclusivos ao sortimento da rede varejista.

Compradores que se engajam em comprar mais de categorias variadas são mais engajados com o negócio, o que reflete maior lucratividade (REINARTZ et al., 2008). Comprar categorias variadas reflete em variáveis que geram impacto direto no valor do cliente: maior retenção, maiores receitas, maior participação na carteira, maior lucratividade (REINARTZ; KUMAR, 2003; VERHOEF; DOORN; DOROTIC, 2007).

Tanto na primeira como na segunda etapa do estudo, foi observado no algoritmo de regressão logística e nos algoritmos das árvores (REPTree e J48) que o comportamento de compra multicanal é um atributo relevante nas classificações consideradas neste estudo. Nas duas etapas, observou-se convergência entre os resultados dos três algoritmos que tiveram seus resultados aprofundados.

Através do modelo gerado pelo algoritmo de regressão logística na etapa 1, foi verificado que clientes que não são multicanais teriam menores chances de serem de alto valor. Convergentemente, nos dois algoritmos das árvores o modelo entendeu que não ser multicanal é o atributo mais relevante para o cliente não ser de alto valor. Já na etapa 2, ser multicanal foi considerado um atributo relevante para a classificação de clientes que respondem à ação de *cross-selling* nos modelos das árvores J48 e REPTree. Assim também, pelo algoritmo de regressão logística, observou-se que clientes que não são multicanais tem menor chance de responder à ação de compra cruzada.

No geral, essa pesquisa mostrou que o atributo de comportamento de compra multicanal é relevante tanto para classificação de clientes de alto valor como para classificação de clientes que respondem a ações de compra cruzada. Adicionalmente, os resultados mostraram que a frequência e o uso ou não de serviços na loja física são preditores também relevantes para classificação das classes consideradas neste estudo. Os achados contribuem com discussões acadêmicas relacionadas a valor do cliente e com praticantes que atuam, em particular, em empresas do segmento varejista.

5.1

Implicações teóricas

Nas duas etapas do estudo, mais de dez atributos relacionados ao valor do cliente foram considerados e algoritmos de *machine learning* foram aplicados. Apesar de alguns estudos já terem investigado e medido o valor do cliente (BOLTON; LEMON; VERHOEF, 2004; REINARTZ; KUMAR, 2000; RUST; LEMON; ZEITHAML, 2004; SHAH et al., 2016; VENKATESAN; KUMAR, 2004b), dentre as pesquisas que este estudo abrangeu, não foram localizados estudos com aplicação de métodos *machine learning*, o que indica que a aplicação deste método dentro de pesquisas acadêmicas na temática de valor do cliente ainda não é muito disseminada.

Através de algoritmos *machine learning* para implementação de *data mining*, este estudo contribui, então, com as discussões acadêmicas relacionadas a valor do cliente através do uso de um método pouco utilizado nas pesquisas dentro dessa temática. Uma vez que o *machine learning* é aplicado com sucesso para solucionar o problema de erros cometidos durante análises ou quando se tenta estabelecer relações entre múltiplos fatores (KOTSIANTIS, 2007), este estudo contribui com as discussões acadêmicas relacionadas a valor do cliente evidenciando atributos relevantes para modelar clientes de alto valor e clientes que responderam a uma ação de *cross-selling*.

Dentre os atributos considerados, encontra-se o comportamento de compra multicanal. Tanto na primeira como na segunda etapa, verificou-se que o atributo multicanal é relevante para modelar as classes consideradas. Os resultados apresentados neste estudo levam, então, à conclusão que o atributo multicanal é relevante para modelos preditivos de clientes de alto valor e de clientes que respondem a ações de *cross-selling*.

Os achados deste estudo são consistentes com estudos que comprovaram que clientes que compram em mais de um canal geram maiores receitas para as empresas do que compradores de um único canal (THOMAS; SULLIVAN, 2005). Os resultados corroboram também com estudos que já comprovaram que compras multicanais podem levar a uma maior lucratividade do cliente (KUMAR;

BEZAWADA; TRIVEDI, 2018; KUSHWAHA; VENKATESH SHANKAR, 2007; VENKATESAN; KUMAR; RAVISHANKER, 2007).

Os resultados dos três algoritmos de maior acurácia na etapa 1 indicaram que não usar serviços *delivery* na loja física e não ser multicanal são comportamentos que preveem clientes que não são de alto valor. Ou seja, clientes que não são de alto valor usufruem e se expõem a menos serviços ofertados pela empresa. Este resultado converge com outros achados da literatura, uma vez que já se comprovou que é esperado que clientes expostos a mais serviços fornecidos pela empresa fiquem mais satisfeitos e desenvolvam uma relação mais profunda com a empresa, se tornando mais lucrativos (VENKATESAN; KUMAR; RAVISHANKER, 2007).

Adicionalmente, é necessário quantificar os benefícios da compra cruzada em termos de melhorias nas métricas baseadas em clientes (KUMAR; GEORGE; PANCRAS, 2008) e o uso de técnicas de mineração de dados para análises de padrões de consumo não é facilmente encontrado na literatura acadêmica. Este estudo, então, contribuiu com esta lacuna ao utilizar uma ampla base de dados reais de clientes de um supermercado, com operação *online* e *offline*, para identificar modelos de clientes que respondem a uma ação de relacionamento de *cross-selling* e para desenvolver um modelo preditivo de clientes de alto valor através de métodos *machine learning*.

Além disso, no campo acadêmico, não é tão vasta a quantidade de pesquisas desenvolvidas para identificar os motivadores de *cross-selling* no contexto de valor do cliente (HEINER EVANSCHITZKY, NEERU MALHOTRA, 2017; KUMAR; GEORGE; PANCRAS, 2008), e dentre os estudos que investigaram, a maioria foca em serviços financeiros, sendo poucos os que tem foco em varejo (BOLTON; LEMON; VERHOEF, 2004; KUMAR; GEORGE; PANCRAS, 2008; VERHOEF; DONKERS, 2005; VERHOEF; DOORN; DOROTIC, 2007). Assim, essa pesquisa contribui com a teoria relacionada ao tema ao identificar atributos que influenciam na resposta a ações de compra cruzada dentro de um contexto de supermercados. Desta forma, este estudo trouxe contribuições nas discussões acadêmicas sobre valor do cliente dentro do contexto de CRM e também em um contexto multicanal.

5.2

Implicações gerenciais

Valor do cliente é uma das métricas essenciais no gerenciamento de clientes e que ajuda as empresas a direcionarem seus esforços de marketing com maior confiança (KUMAR; REINARTZ, 2006). Quando a empresa entende o valor individual do cliente em relação a outros clientes, ela pode alocar seus recursos com maior efetividade (PEPPERS; ROGERS, 2011). Através dos atributos indicados como relevantes neste estudo na classificação de clientes de alto valor, as empresas, no segmento varejista em particular, podem identificar clientes com maior propensão a serem mais valiosos. Conhecendo quais clientes são mais e menos valiosos, as empresas podem priorizar seus esforços competitivos, dedicando maior esforço, tempo e recursos aos clientes mais propensos a produzir maiores retornos (PEPPERS; ROGERS, 2011).

Um dos estímulos que permite às empresas aumentar a receita de contribuição de clientes já existentes é o *cross-selling* (KUMAR; GEORGE; PANCRAS, 2008). A venda cruzada é vista por gerentes como a melhor ferramenta de desenvolvimento de clientes (NGOBO, 2004). A principal vantagem de identificar variáveis de interação entre empresa e cliente como motivadores de comportamento de compra cruzada é que empresas podem identificar clientes mais propensos a esse tipo de compra (KUMAR; GEORGE; PANCRAS, 2008).

Um desafio gerencial é melhorar as taxas de resposta de campanhas deste tipo e evitar alcançar clientes com mensagens irrelevantes (LI; SUN; MONTGOMERY, 2011). Ao compreender os atributos relevantes na classificação de modelos de clientes que respondem a ações de compra cruzada, a empresa pode identificar pontos a serem observados no perfil do cliente para aprimorar a base de clientes impactados por ações *cross-selling*. Uma das implicações da presente pesquisa é que as empresas varejistas devem reconhecer que nem todos os clientes respondem positivamente a ações de *cross-selling*.

Portanto, os resultados deste estudo são úteis a praticantes. Os resultados sugerem que clientes que não são de alto valor, de acordo com os resultados deste estudo, parecem não ter alta probabilidade de responder a ações de *cross-selling*.

Desta forma, empresas devem focar em clientes de alto valor para segmentar a oferta de ações de *cross-selling*.

5.3

Limitações do estudo e recomendações de pesquisas futuras

Como não foi possível ter acesso a outras ações de relacionamento com os clientes, além da ação considerada neste estudo, a compreensão de possíveis impactos no comportamento do cliente causados por outras ações de relacionamento praticadas pela rede varejista ficou limitada. Portanto, em pesquisas futuras, recomenda-se contemplar as outras ações de relacionamento praticadas no período estudado para identificar modelos de clientes que respondem à ação de compra cruzada. Adicionalmente, é sugerido que em pesquisas futuras sejam considerados *e-mails* recebidos previamente e canal de aquisição do cliente, por exemplo, já que esses são alguns dos motivadores já identificados em estudos acadêmicos anteriores (VERHOEF; DONKERS, 2005; VERHOEF; FRANSES; HOEKSTRA, 2001) como sendo relevantes na influência de comportamento de compra cruzada.

Outra limitação deste estudo é que os custos com as ações de marketing não foram compartilhados pela rede varejista parceira desta pesquisa, impossibilitando que métricas de CLV fossem aplicadas ou calculadas. Desta forma, em pesquisas futuras que se proponham a modelar clientes de alto valor, recomenda-se calcular o CLV de cada cliente para identificar os modelos.

6 REFERÊNCIAS BIBLIOGRÁFICAS

AHER, S. B.; L.M.R.J., L. Data Mining in Educational System using WEKA. **International Conference on Emerging Technology Trends (ICETT)**, p. 20–25, 2011.

ALTMAN, N. S. An Introduction to Kernel and Nearest Neighbor Nonparametric Regression. **The American Statistician**, v. 46, n. 3, p. 175–185, 1992.

ANDERSON, E. W.; FORNELL, C.; LEHMANN, D. R. Customer satisfaction, Market Share and Profitability: Findings from Sweden. **Journal of Marketing**, v. 58, n. 3, p. 53–66, 1994.

ANSARI, A.; MELA, C. F.; NESLIN, S. A. Customer Channel Migration. **Journal of Marketing Research**, v. XLV, n. February, p. 60–76, 2008.

BECK, N.; RYGL, D. Categorization of multiple channel retailing in Multi-, Cross-, and Omni - Channel Retailing for retailers and retailing. **Journal of Retailing and Consumer Services**, v. 27, p. 170–178, 2015.

BECKERS, S. F. M.; VAN DOORN, J.; VERHOEF, P. C. Good, better, engaged? The effect of company-initiated customer engagement behavior on shareholder value. **Journal of the Academy of Marketing Science**, v. 46, p. 366–383, 2018.

BERRY, L. L. et al. Opportunities for innovation in the delivery of interactive retail services. **Journal of Interactive Marketing**, v. 24, n. 2, p. 155–167, 2010.

BERRY, M.; GS LINOFF. Mastering data mining: The art and science of customer relationship management. **Industrial Management & Data Systems**, v. 100, p. 245–246, 2000.

BOLTON, R. N.; LEMON, K. N.; VERHOEF, P. C. The Theoretical Underpinnings of Customer Asset Management: A Framework and Propositions for Future Research. **Journal of the Academy of Marketing Science**, v. 32, n. 3, p. 271–292, 2004.

BOLTON, R. N.; LEMON, K. N.; VERHOEF, P. C. Customer Relationships: Modeling the Customer's Upgrade Decision. **Journal of Marketing**, v. 72, n. January, p. 46–64, 2008.

BOULDING, W. et al. A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions. **Journal of Marketing Research**, v. 30, n. 1, p. 7, 1993.

BUCKLIN, R. E.; GUPTA, S. Commercial Use of UPC Scanner Data : Industry and Academic Perspectives. **Marketing Science**, v. 18, n. 3, p. 247–273, 1999.

CHAWLA, N. V et al. SMOTE: Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321–357, 2002.

CORTES, C.; VAPNIK, V. Support-Vector Networks. **Machine Learning**, v. 20, p. 273–297, 1995.

DA COSTA, S. B.; CLARO, D. P.; BORTOLUZZO, A. B. A influência do valor do cliente na recomendação: Um estudo empírico no mercado Brasileiro de previdência privada. **Revista Brasileira de Gestao de Negocios**, v. 20, n. 2, p. 210–231, 2018.

DHOLAKIA, U. M. et al. Consumer Behavior in a Multichannel , Multimedia Retailing Environment. **Journal of Interactive Marketing**, v. 24, p. 86–95, 2010.

DWYER, F. R. Customer Lifetime Valuation to Support Marketing Decision Making. **Journal of Direct Marketing**, v. 11, n. 4, p. 6–13, 1997.

DYER, J. H.; SINGH, H. The Relational View: Cooperative Strategy and Sources of Inteorganizational Competitive Advantage. **The Academy of Management Review**, v. 23, n. 4, p. 660–679, 1998.

FADER, P. S.; HARDIE, B. G. S.; LEE, K. L. RFM and CLV : Using Iso-value Curves for Customer Base Analysis. **Journal of Marketing Research**, v. 42, p. 415–30, 2005.

FARROW, C. L.; KIM, H.; BILLINGE, S. J. L. The Nyquist-Shannon sampling theorem and the atomic pair distribution function. **Physical Review B**, 2011.

FORNELL, C. A National Customer Satisfaction Barometer: The Swedish Experience. **Journal of Marketing**, v. 56, n. January, p. 6–21, 1992.

FOURNIER, S.; YAO, J. L. Reviving brand loyalty: A reconceptualization within the framework of consumer-brand relationships. **International Journal of Research in Marketing**, v. 14, p. 451–472, 1997.

FRADKIN, D.; MUCHNIK, I. Support Vector Machines for Classification. **DIMACS Series in Discrete Mathematics and Theoretical Computer Science**, v. 70, p. 13–20, 2006.

GREWAL, D. et al. Strategic Online and Offline Retail Pricing: A Review and Research Agenda. **Journal of Interactive Marketing**, v. 24, n. 2, p. 138–

154, 2010.

GUPTA, S.; LEHMANN, D. R.; STUART, J. A. Valuing customers. **Journal of Marketing Research**, v. 41, n. 3, p. 7–18, 2004.

GUPTA, S.; ZEITHAML, V. Customer Metrics and Their Impact on Financial Performance. **Marketing Science**, v. 25, n. 6, p. 718–739, 2006.

HALL, M. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, v. 11, n. 1, p. 10–18, 2009.

HEINER EVANSCHITZKY, NEERU MALHOTRA, F. V. W. AND K. N. L. Antecedents of Peripheral Services Cross-Buying Behavior. **Journal of Retailing and Consumer Services**, v. 36, n. May, 2017, p. 218–224, 2017.

HUNT, S. D.; ROBERT M. MORGAN. The Comparative Advantage Theory of Competition. **Journal of Marketing**, v. 59, n. April 1995, p. 1–15, 1995.

JAIN, D.; SINGH, S. S. Customer Lifetime Value Research in Marketing: A Review and Future Directions. **Journal of Interactive Marketing**, v. 16, n. 2, p. 34–46, 2002.

JAYACHANDRAN, S. et al. The Role of Relational Information Processes and Technology Use in Customer Relationship Management. **Journal of Marketing**, v. 69, p. 177–92, 2004.

JONES, M. A.; MOTHERSBAUGH, D. L.; BEATTY, S. E. Switching Barriers and Repurchase Intentions in Services. **Journal of Retailing**, v. 76, n. 2, p. 259–274, 2000.

KANNAN, P. K.; REINARTZ, W.; VERHOEF, P. C. The path to purchase and attribution modeling: Introduction to special section. **International Journal of Research in Marketing**, v. 33, n. 3, p. 449–456, 2016.

KEANE, T. J.; WANG, P. Applications for the lifetime value model in modern newspaper publishing. **Journal of Direct Marketing**, v. 9, n. 2, p. 59–66, 1995.

KEEN, C. et al. E-tailers versus Retailers: Which Factors Determine Consumer Preferences. **Journal of Business Research**, v. 57, p. 685–95, 2004.

KONUS, U.; VERHOEF, P. C.; NESLIN, S. A. Multichannel Shopper Segments and Their Covariates. **Journal of Retailing**, v. 84, n. 2007, p. 398–413, 2008.

KOTSIANTIS, S. B. Supervised Machine Learning: A Review of Classification Techniques. **Journal of Manufacturing Science and Engineering, Transactions of the ASME**, v. 31, n. 3, p. 249–268, 2007.

KUMAR, A.; BEZAWADA, R.; TRIVEDI, M. The Effects of Multichannel Shopping on Customer Spending, Customer Visit Frequency, and

Customer Profitability. **Journal of the Association for Consumer Research**, v. 3, n. 3, p. 294–311, 2018.

KUMAR, V. et al. Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value. **Journal of Service Research**, v. 13, p. 297–310, 2010.

KUMAR, V. A Theory of Customer Valuation: Concepts, Metrics, Strategy, and Implementation. **Journal of Marketing**, v. 82, n. 1, p. 1–19, 2018.

KUMAR, V.; GEORGE, M.; PANCRAAS, J. Cross-buying in retailing: Drivers and consequences. **Journal of Retailing**, v. 84, n. 1, p. 15–27, 2008.

KUMAR, V.; PANSARI, A. Competitive Advantage through Engagement. **Journal of Marketing Research**, v. 53, n. 4, p. 497–514, 2015.

KUMAR, V.; PANSARI, A. National culture, economy, and customer lifetime value: Assessing the relative impact of the drivers of customer lifetime value for a global retailer. **Journal of International Marketing**, v. 24, n. 1, p. 1–21, 2016.

KUMAR, V.; REINARTZ, W. **Customer Relationship Management: Concept, Strategy, and Tools**. Heidelberg, Germany: Springer, 2012.

KUMAR, V.; REINARTZ, W. Creating Enduring Customer Value. **Journal of Marketing**, v. 80, n. 6, p. 36–68, 2016.

KUMAR, V.; REINARTZ, W. J. **Customer Relationship Management: A Databased Approach**. 1. ed. Hoboken, NJ: John Wiley & Sons, Inc., 2006.

KUMAR, V.; SHAH, D. Expanding the Role of Marketing: From Customer Equity to Market Capitalization. **Journal of Marketing**, v. 73, n. 6, p. 119–136, 2009.

KUMAR, V.; SHAH, D.; VENKATESAN, R. Managing retailer profitability—one customer at a time! **Journal of Retailing**, v. 82, n. 4, p. 277–294, 2006.

KUMAR, V.; VENKATESAN, R. Who are the multichannel shoppers and how do they perform? Correlates of multichannel shopping behavior. **Journal of Interactive Marketing**, v. 19, n. 2, 2005.

KUMAR, V.; VENKATESAN, R.; REINARTZ, W. Performance Implications of Adopting a Customer-Focused Sales. **Journal of Marketing**, v. 72, n. September, p. 50–68, 2008.

KUSHWAHA, T.; VENKATESH SHANKAR. Optimal Multichannel Allocation of Marketing Efforts. **MSI Report**, p. 07–207, 2007.

LEVY, M.; WEITZ, B. A. **Retailing Management**. 9th. ed. New York: McGraw-Hills Irwin, 2009.

LI, G.; LI, C.; JIA, W. A Hidden Markov Model of Customer Relationship

Dynamics in Retailing Industry. **Research and Practical Issues of Enterprise Information Systems**, v. 205, p. 633–634, 2006.

LI, S.; SUN, B.; MONTGOMERY, A. L. Cross-Selling the Right Product to the Right Customer at the Right Time. **Journal of Marketing Research**, n. November 2010, p. 683–700, 2011.

LI, S.; SUN, B.; WILCOX, R. T. Cross-Selling Sequentially Ordered Products: An Application to Consumer Banking Services. **Journal of Marketing Research**, v. XL, n. May 2005, p. 233–239, 2005.

LIU, H.; LOBSCHAT, L.; VERHOEF, P. C. Multichannel Retailing: A Review and Research Agenda. **Foundations and Trends® in Marketing**, v. 12, n. 1, p. 1–79, 2018.

MÄGI, A. W. Share of wallet in retailing: the effects of customer satisfaction, loyalty cards and shopper characteristics. **Journal of Retailing**, v. 79, p. 97–106, 2003.

MELIS, K. et al. A bigger slice of the multichannel grocery pie: When does consumers' online channel use expand retailers' share of wallet? **Journal of Retailing**, p. 1–49, 2016.

MITTAL, V.; KAMAKURA, W. A. Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics. **Journal of Marketing Research**, v. XXXVIII, n. 2001, p. 131–142, 2001.

NESLIN, S. A. et al. Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. **Journal of Marketing Research**, v. 43, n. May 1, p. 204–211, 2006a.

NESLIN, S. A. et al. Challenges and Opportunities in Multichannel Customer Management. **Journal of Service Research**, v. 9, n. 2, p. 95–112, 2006b.

NESLIN, S. A. et al. Multichannel Customer Management. **Journal of Service Research**, v. 9, n. 2, p. 95–113, 2006c.

NESLIN, S. A.; SHANKAR, V. Key Issues in Multichannel Customer Management: Current Knowledge and Future Directions. **Journal of Interactive Marketing**, v. 23, n. 1, p. 70–81, 2009.

NETZER, O.; LATTIN, J. M. A Hidden Markov Model of Customer Relationship Dynamics. **Marketing Science**, n. March 2005, 2005.

NGOBO, P. V. Drivers of customers' cross-buying intentions. **European Journal of Marketing**, v. 38, n. 9, p. 1129–1157, 2004.

PAYNE, A.; FROW, P. A Strategic Framework for Customer Relationship Management. **Journal of Marketing**, v. 69, n. 4, p. 167–176, 2005.

PENG, C. Y. J.; LEE, K. L.; INGERSOLL, G. M. An introduction to logistic regression analysis and reporting. **Journal of Educational Research**, v. 96, n. 1, p. 3–14, 2002.

PEPPERS, D.; ROGERS, M. **Managing Customer Relationships: A Strategic Framework**. 2. ed. Hoboken, NJ: Wiley, 2011.

QUINLAN, J. R. Improved Use of Continuous Attributes in C4.5. **Journal of Artificial Intelligence Research**, v. 4, p. 77–90, 1996.

RAMSEY, G.; BAPNA, S. Text Mining to Identify Customers Likely to Respond to Cross-Selling Campaigns: Reading Notes from Your Customers. **International Journal of Business Analytics**, v. 3, n. 2, p. 2–5, 2016.

REICHHELD, F. F. Loyalty-based management. **Harvard Business Review**, v. 71, n. 2, p. 64–73, 1993.

REINARTZ, W. et al. The Customer Relationship Management Process : Its Measurement and Impact on Performance. **Journal of Marketing Research**, v. 41, n. 3, p. 293–305, 2004.

REINARTZ, W. et al. Investigating cross-buying and customer loyalty. **Journal of Interactive Marketing**, v. 22, n. 1, p. 5–21, 2008.

REINARTZ, W. J.; KUMAR, V. On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing. **Journal of Marketing**, v. 64, n. October 2000, p. 17–35, 2000.

REINARTZ, W. J.; KUMAR, V. The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration. **Journal of Marketing**, v. 67, n. 12, p. 77–99, 2003.

RISH, I. An empirical study of the naive Bayes classifier. **IJCAI 2001 workshop on empirical methods in artificial intelligence**, v. 3, n. 22, p. 41–46, 2001.

ROBERT B. WOODRUFF. Customer Value: The Next Source for Competitive Advantage. **Journal of the Academy of Marketing Science**, v. 25, n. 3, p. 256–256, 1997.

ROSSI, P. E.; MCCULLOCH, R. E.; ALLENBY, G. M. The Value of Purchase History Data in Target Marketing. **Marketing Science**, v. 15, n. 4, 1996.

RUST, R. T.; LEMON, K. N.; ZEITHAML, V. A. Return on Marketing : Using Customer Equity to Focus. **Journal of Marketing**, v. 68, n. January, p. 109–127, 2004.

RUST, R. T.; VERHOEF, P. C. Optimizing the Marketing Interventions Mix in Intermediate-Term CRM. **Marketing Science**, v. 24, n. 3, p. 477–489, 2005.

SAMIZADEH, R.; MEHREGAN, S. Retaining customers using clustering and association rules in insurance industry: A case study. **International Journal of Management and Business Research**, v. 5, n. 4, p. 261–268, 2015.

SCHMITT, P.; SKIERA, B.; BULTE, C. VAN DEN. Referral Programs and Customer Values. **Journal of Marketing**, v. 75, n. January, p. 46–59, 2011.

SHAH, D. et al. Linking Customer Behaviors to Cash Flow Level and Volatility: Implications for Marketing Practices. **Journal of Marketing Research**, v. 54, n. 1, p. 27–43, 2016.

SHARMA, T. C.; JAIN, M. WEKA Approach for Comparative Study of Classification Algorithm. **(IJARCCE) International Journal of Advanced Research in Computer and Communication Engineering**, v. 2, n. 4, p. 1925–1931, 2013.

SLATER, S. F. Developing a Customer Value-Based Theory of the Firm. **Journal of the Academy of Marketing Science**, v. 25, n. 2, p. 162–167, 1997a.

SLATER, S. F. Developing a customer value-based theory of the firm. **Journal of the Academy of Marketing Science**, v. 25, n. 2, p. 162–167, 6 maio 1997b.

SULLIVAN, U.; THOMAS, J. S. **Customer Migration: An Empirical Investigation Across Multiple Channels**: Working Paper 04-0112. University of Illinois at Urbana-Champaign, College of Business.

SUNDER, S.; KUMAR, V.; ZHAO, Y. Measuring the Lifetime Value of a Customer in the Consumer Packaged Goods Industry. **Journal of Marketing Research**, v. 53, n. 6, p. 901–921, 2016.

THOMAS, J. S.; SULLIVAN, U. Y. Managing Marketing Communications with Multichannel. **Journal of Marketing**, v. 69, n. 4, p. 239–251, 2005.

VENKATESAN, R. Executing on a customer engagement strategy. **Journal of the Academy of Marketing Science**, v. 45, p. 289–293, 2017.

VENKATESAN, R.; KUMAR, V. A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. **Journal of Marketing**, v. 68, n. October, p. 106–125, 2004a.

VENKATESAN, R.; KUMAR, V. A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. **Journal of Marketing**, v. 68, n. October, p. 106–125, 2004b.

VENKATESAN, R.; KUMAR, V.; BOHLING, T. Optimal Customer Relationship Management Using Bayesian Decision Theory : An Application for Customer Selection. **Journal of Marketing Research**, v. 44, n. November 2007, p. 579–594, 2007.

VENKATESAN, R.; KUMAR, V.; RAVISHANKER, N. Multichannel Shopping: Causes and Consequences. **Journal of Marketing**, v. 71, n. April, p. 114–132, 2007.

VERHOEF, P. C. et al. The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands. **Decision Support Systems**, v. 34, p. 471–481, 2002.

VERHOEF, P. C. Understanding the Effect of Efforts on Customer Retention and Customer Share Development. **Journal of Marketing**, v. 67, n. October, p. 30–45, 2003.

VERHOEF, P. C. et al. CRM in Data-Rich Multichannel Retailing Environments: A Review and Future Research Directions. **Journal of Interactive Marketing**, v. 24, n. 2, p. 121–137, 2010.

VERHOEF, P. C.; DONKERS, B. The effect of acquisition channels on customer loyalty and cross-buying. **Journal of Interactive Marketing**, v. 19, n. 2, p. 31–43, 2005.

VERHOEF, P. C.; DOORN, J. VAN; DOROTIC, M. Customer Value Management: An Overview and Research Agenda. **Journal of Research in Management**, v. 2, p. 51–69, 2007.

VERHOEF, P. C.; FRANCES, P. H.; HOEKSTRA, J. C. The impact of satisfaction and payment equity on cross-buying: A dynamic model for a multi-service provider. **Journal of retailing**, v. 77, p. 359–378, 2001.

VERHOEF, P. C.; FRANCES, P. H.; HOEKSTRA, J. C. The impact of satisfaction and payment equity on cross-buying: A dynamic model for a multi-service provider. **Journal of Retailing**, v. 77, p. 359–378, 2001.

VICHROSKI, T. DA S. F. et al. O Valor Real do Cliente no Processo de CRM e a Contabilidade: Um estudo de caso em empresa no setor de Varejo Supermercadista. **REGE**, v. 17, n. 4, p. 471–488, 2010.

WIESEL, T.; SKIERA, B.; VILLANUEVA, J. Customer Equity: An Integral part of Financial Reporting. **Journal of Marketing**, v. 72, p. 1–14, 2008.

WILSON, J. R.; LORENZ, K. A. Short History of the Logistic Regression Model. **Springer International Publishing**, p. 17–23, 2015.

YRJÖLÄ, M.; SAARIJÄRVI, H.; NUMMELA, H. The value propositions of multi-, cross-, and omni-channel retailing. **International Journal of Retail & Distribution Management**, v. 46, n. 11/12, p. 1133–1152, 2018.

ZAHAY, D.; GRIFFIN, A. **Are Customer Information Systems Worth It? Results from B2B Services**: Working Paper 02-113. Cambridge, MA: Marketing Science Institute, 2002.

ZERBINO, P. et al. Big Data-enabled Customer Relationship Management: A holistic approach. **Information Processing and Management**, v. 54, n.

5, p. 818–846, 2017.

ZHANG, J. et al. Crafting Integrated Multichannel Retailing Strategies. **Journal of Interactive Marketing**, v. 24, n. 2, p. 168–180, 2010.