

**UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E NEGÓCIOS
MESTRADO PROFISSIONAL**

PIETRO REINHEIMER DE FRANCESCHI

MODELAGENS PREDITIVAS DE *CHURN*: O CASO DO BANCO DO BRASIL

PORTO ALEGRE

2019

PIETRO REINHEIMER DE FRANCESCHI

MODELAGENS PREDITIVAS DE *CHURN*: O CASO DO BANCO DO BRASIL

Dissertação de Mestrado apresentado(a) como requisito parcial para obtenção do título de Mestre em Administração, pelo Programa de Pós-Graduação em Gestão e Negócios da Universidade do Vale do Rio dos Sinos – UNISINOS.

Orientador: Prof. Dr. Guilherme Trez

Porto Alegre

2019

D278m De Franceschi, Pietro Reinheimer.
Modelagens preditivas de *churn*: o caso do banco do brasil
/ Pietro Reinheimer De Franceschi. – Porto Alegre, 2019.
119 f. : il. ; 30 cm

Dissertação (mestrado) – Universidade do Vale do Rio dos
Sinos, Programa de Pós-Graduação em Gestão de Negócios,
Porto Alegre, 2019.

“Orientador: Prof. Dr. Guilherme Trez.”

1. Marketing de relacionamento. 2. Administração.
3. Bancos. I. Título.

CDU 658.89

Catálogo na publicação:
Bibliotecária: Lenise Di Domenico - CRB 10/1757

PIETRO REINHEIMER DE FRANCESCHI

MODELAGENS PREDITIVAS DE *CHURN*: O CASO DO BANCO DO BRASIL

Dissertação de Mestrado apresentado(a) como requisito parcial para obtenção do título de Mestre em Administração, pelo Programa de Pós-Graduação em Gestão e Negócios da Universidade do Vale do Rio dos Sinos – UNISINOS.

Aprovado em 31/07/2019

BANCA EXAMINADORA

Orientador: Prof. Dr. Guilherme Trez - UNISINOS

Membro: Prof. Dr. João Zani - UNISINOS

Membro: Profa. Dra. Vanessa Martins Pires - FURG

AGRADECIMENTOS

Enquanto eterno aprendiz, a disposição e interesse em dar mais um passo nesta caminhada em busca do conhecimento e do desenvolvimento, pessoal e profissional, através do Mestrado, não eram recentes. Buscar o ingresso em um programa reconhecido com a nota máxima da CAPES, de dupla titulação, foi uma decisão que sempre me trará orgulho. Sinto-me abençoado por ter tido a oportunidade de viver esta experiência transformadora, e sou profundamente grato a muitos.

Conciliar um curso de mestrado com uma exigente atividade profissional gera uma série de privações. Além das aulas, a carga de leituras e trabalhos nos exigiam, certamente, mais ainda. O cansaço e esgotamento, por vezes, nos torna menos pacientes e tolerantes. À Isadora, registro, portanto, toda minha gratidão por ser essa grande parceira que tenho em minha vida, pelo apoio incondicional e alento irrestrito.

Sem o incentivo familiar, durante esta e todas as etapas da minha formação, certamente essa jornada não teria acontecido. Meus amados pais e irmãos, Beto, Kika, Thiago e Laura, e também em memória de meus estimados avós, Vô Norberto, Vô Lena, Vô Arlindo e Vô Silma, essa é por todos vocês. Estendo ainda minha gratidão aos meus cunhados, amado sobrinho e afilhados, primos, tios e avós, de sangue e coração, bem como a todos amigos e irmãos que a vida me proporcionou.

Fazer parte de uma empresa como o Banco do Brasil é motivo de muito orgulho para mim. Registro meus profundos agradecimentos aos colegas da Agência 3866, que permitiram me dedicar a esta dupla jornada sempre com toda a confiança por suas conduções durante minhas ausências. Também registro meu agradecimento aos líderes desenvolvedores do Banco do Brasil que me encorajaram a esta etapa: Me. Gustavo Henrique da Rosa, Me. Edson Bündchen, Me. Júlio César Rodrigues da Silva, Me. Fabiano Mattana e Me. Sandro Genari.

Ao meu orientador Prof. Dr. Guilherme Trez, minha gratidão por toda generosidade, confiança e apoio em todas as etapas desta dissertação e, por sua condição de coordenador deste Programa, estendo meu reconhecimento a todos os professores do MPGN. Agradeço aos professores Dr. Luiz Felipe Vallandro, Dra. Vanessa Martins Pires e Dr. João Zani, pelas valiosas contribuições a esta dissertação. Minhas considerações também pelo direcionamento técnico prestado pelos professores Me. Davi Simon e Dr. Daniel Vancin.

Aos meus queridos colegas do mestrado, mais do que amigos, toda gratidão e orgulho de ter estado junto de vocês. Aprendi muito com cada um. Vocês são os melhores.

Enfim, obrigado a todos e a vida por terem me proporcionado esta oportunidade, e por ter conseguido chegar até esse momento. Vamos em frente.

RESUMO

Este estudo teve por objetivo comparar modelos preditivos para identificação de clientes de uma instituição financeira que tendem ao fenômeno do *churn*. A perda ou abandono de clientes (*churn*), ainda que em uma perspectiva parcial (desidratação), representa impactos significativos aos resultados das organizações. O contexto de estudo é uma das 100 maiores empresas financeiras do mundo. A metodologia aplicada correspondeu a uma fase qualitativa, com o objetivo de mapear variáveis consideradas preditoras da evasão de clientes, aderentes a Subramanya e Somani (2017). Posteriormente foram desenvolvidas análises quantitativas em uma amostra de 2343 clientes fornecida pelo Banco do Brasil S.A., contemplando 118 variáveis explicativas, com posição histórica mensal de até 17 meses anteriores a fevereiro de 2019. O tratamento estatístico dos dados e o desenvolvimento dos modelos preditivos da desidratação foram realizados por meio do *software* RSTUDIO. Dentre as técnicas de mineração de dados que podem ser utilizadas na execução dos modelos preditivos utilizou-se a Regressão Logística, Regressão Logística com seleção *Stepwise* de variáveis independentes, Florestas Aleatórias (*Random Forests*) e Redes Neurais, tendo em vista a sugestão verificada na literatura de que tais metodologias são promissoras na previsão de *customer churn*. Os resultados mais expressivos foram alcançados por meio da Regressão Logística, tendo sido observada acurácia (índice de acertos) em patamar de 81%. O trabalho evidencia que o desenvolvimento de ações de retenção pode contribuir significativamente para os resultados da empresa.

Palavras chave: Evasão de clientes. Modelagens preditivas. Mineração de dados. Regressão Logística. Stepwise. Random Forests. Redes Neurais.

ABSTRACT

This study aimed to compare predictive models for customer identification of a financial institution that tend to churn phenomenon. The loss or abandonment of customers, even in a partial perspective (dehydration), represents significant impacts on the results of organizations. The context of study is one of the 100 largest financial companies in the world. The applied methodology corresponded to a qualitative phase, aiming to map variables considered as predictors of customer dropout, adhering to Subramanya and Somani (2017). Subsequently, I performed quantitative analyzes on a sample of 2343 clients provided by Banco do Brasil SA, comprising 118 independent variables with monthly historical position up to 17 months prior to February 2019. Then I performed data statistical treatment and developed predictive models of dehydration through the RSTUDIO software. Among the data mining techniques that can be used in the execution of predictive models, Logistic Regression, Logistic Regression with Stepwise Selection of Independent Variables, Random Forests and Neural Networks were used, considering the suggestion found in literature that such methodologies are promising in predicting customer churn. The most expressive results were achieved through logistic regression, with accuracy (hit ratio) of 81%. The work shows that the development of retention actions can contribute significantly to the company's results.

Keywords: Customer churn. Prediction churn. Data mining. Logistic regression. Stepwise. Random forests. Neural networks.

LISTA DE QUADROS

Quadro 1 - Evolução da Margem de Contribuição na amostra de trabalho.....	22
Quadro 2 - Simulador Customer Equity.....	23
Quadro 3 - Principais constructos do Marketing de Relacionamento com impacto em fidelidade.....	28
Quadro 4 - Síntese teórica de estudos/modelos preditivos do churn	51
Quadro 5 - Caracterização dos entrevistados na etapa qualitativa	56
Quadro 6 - Proposta de variáveis para análises preditivas	58
Quadro 7 - Output variável proxy	67
Quadro 8 - Matriz de confusão para um classificador binário.....	73
Quadro 9 - Resumo da Matriz de Confusão e métricas para avaliação de modelos.	74
Quadro 10 - Matriz de Confusão dos resultados da Regressão Logística	81
Quadro 11 - Consolidação dos resultados analíticos da Regressão Logística.....	81
Quadro 12 - Matriz de Confusão dos resultados da Regressão Stepwise	83
Quadro 13 - Consolidação dos resultados analíticos da Regressão Stepwise.....	83
Quadro 14 - Matriz de Confusão dos resultados da Random Forests.....	84
Quadro 15 - Consolidação dos resultados analíticos da Random Forests.....	84
Quadro 16 - Matriz de Confusão dos resultados das Redes Neurais Artificiais	85
Quadro 17 - Consolidação dos resultados analíticos das Redes Neurais Artificiais..	86
Quadro 18 - Consolidação dos resultados analíticos das modelagens preditivas.....	88

LISTA DE FIGURAS

Figura 1 - Estágios da relação cliente-empresa	39
Figura 2 - Processos de negócios	40
Figura 3 - Taxa de <i>churn</i> em clientes com ocorrência canais de Ouvidoria	44
Figura 4 - Ciclo de aprendizagem em Data Mining	49
Figura 5 - Modelo de framework preditivo de <i>churn</i>	50
Figura 6 - Verificação de correlação.....	63
Figura 7 - Histograma da quantidade agregada de produtos	64
Figura 8 - Gráfico de densidade da variação da quantidade de produtos	65
Figura 9 - Variação da quantidade de produtos (contagem)	65
Figura 10 – Gráfico de densidade da oscilação negativa de produtos.....	66
Figura 11 - Representação gráfica - Random Forests	71
Figura 12 - Representação figurativa de Redes Neurais.....	72
Figura 13 - Curva ROC e AUC	75
Figura 14 - Representação gráfica do resultado A3 das Redes Neurais.....	87

LISTA DE TABELAS

Tabela 1 - Impacto da retenção sobre a base de clientes.....	43
Tabela 2 - Analise descritiva das variáveis explicativas	77
Tabela 3 - Teste Kolmogorov-Smirnov (KS).....	79

SUMÁRIO

1 INTRODUÇÃO	13
1.1 DEFINIÇÃO DO PROBLEMA DE PESQUISA	18
1.2 OBJETIVOS	19
1.2.1 Objetivo Geral	19
1.2.2 Objetivos Específicos	19
1.3 DELIMITAÇÃO DA PESQUISA.....	20
1.4 JUSTIFICATIVA E RELEVÂNCIA	20
1.5 ESTRUTURA DE TRABALHO	24
2 REFERENCIAL TEÓRICO	25
2.1 MARKETING DE RELACIONAMENTO.....	25
2.2 CUSTOMER LIFETIME VALUE	33
2.3 CHURN	41
2.3.1 Abordagens preditivas aplicáveis ao fenômeno do <i>churn</i>	48
3 MÉTODO DE PESQUISA	54
3.1 DELINEAMENTO DA PESQUISA.....	54
3.2 ETAPA QUALITATIVA	55
3.3 ETAPA QUANTITATIVA.....	60
3.3.1 Universo e amostra	60
3.3.2 Tratamento dos dados	61
3.3.3 Construção da variável dependente	64
3.3.4 Modelos e métodos preditivos para o fenômeno do churn	68
3.3.4.1 Regressão Logística.....	69
3.3.4.2 Regressão Logística Stepwise	70
3.3.4.3 Random Forests.....	71
3.3.4.4 Redes Neurais.....	71
3.3.5 Métricas de avaliação de modelos de classificação	72
4 ANÁLISE E DISCUSSÃO DOS RESULTADOS	76
4.1 ANÁLISE DESCRITIVA DOS DADOS	76
4.1.1 Teste K-S	78
4.2 CONSTRUÇÃO DOS MODELOS DE CLASSIFICAÇÃO.....	80
4.3 REGRESSÃO LOGÍSTICA.....	81
4.4 REGRESSÃO LOGÍSTICA STEPWISE	82

4.5 RANDOM FORESTS	83
4.6 REDES NEURAIIS	85
4.7 SÍNTESE DOS RESULTADOS	88
5 CONCLUSÕES	90
5.1 IMPLICAÇÕES TEÓRICAS.....	90
5.2 IMPLICAÇÕES PRÁTICAS	91
5.3 LIMITAÇÕES DO MÉTODO.....	92
5.4 RECOMENDAÇÕES PARA TRABALHOS FUTUROS	94
REFERÊNCIAS.....	95
APENDICE A – VARIÁVEIS EXPLICATIVAS FORNECIDAS PELO BANCO DO BRASIL.....	103
APENDICE B – PADRÃO DOS DADOS RECEBIDOS.....	106
APENDICE C – REGRESSÃO LOGÍSTICA.....	107
APENDICE D – REGRESSÃO LOGÍSTICA STEPWISE	109
APENDICE E – COEFICIENTES DA REGRESSÃO LOGÍSTICA E REGRESSÃO STEPWISE.....	111
APENDICE F – RANDOM FORESTS.....	115
APENDICE G – REDES NEURAIIS.....	118

1 INTRODUÇÃO

No atual ambiente de negócios, cada vez mais competitivo, os clientes podem facilmente alternar entre os concorrentes. Por isso, a previsão da perda ou abandono de clientes – *churn* – é uma questão primordial no gerenciamento do relacionamento com o cliente, conforme Zhu et al. (2017). Outro ponto destacado por estes autores é que os clientes de longo prazo tendem a ser menos sensíveis aos concorrentes, geram ticket médio mais elevado, produzindo margens de contribuição e lucros maiores. Neste sentido, salientam que as empresas estão alterando o foco destinado a aquisição de novos clientes para a retenção de seus clientes, especialmente os principais. Desta forma, por ajudar as empresas a direcionar ações de retenção para os clientes certos, a previsão do *churn* é, portanto, reconhecidamente uma prioridade para os gestores de clientes e suas organizações.

A pesquisa de Reichheld e Sasser Jr. (1990) apresentou pontos relevantes sobre a necessidade de uma efetiva gestão da retenção de clientes. Foram elencados exemplos em que o crescimento exponencial dos resultados de uma organização passa por minimizar a ocorrência do *churn*. Estes autores evidenciaram que no caso de agências bancárias, o estudo empírico mostrou que reduzir em 5% o índice de *churn* é capaz de elevar os lucros em 85%. Com este propósito, cada vez mais, as organizações têm atuado neste sentido. Portanto, a compreensão do fenômeno do *churn* permite que a organização implemente estratégias e ações sobre as variáveis-chave que o influenciam, na tentativa de minimizá-lo (BOTELHO; TOSTES, 2010).

Neslin et al. (2006) afirmam que a compreensão do fenômeno *churn* e a definição de estratégias claras para blindar esse processo é relevante para a perenidade das organizações, destacando a importância das soluções de CRM – *Customer Relationship Management* preponderantemente pela capacidade de realizar análises preditivas. Isto é, por conhecer profundamente o cliente e suas necessidades, e pelo mapeamento de seu momento de vida, de forma antecipada, consegue-se identificar e prever as próximas etapas do relacionamento com os clientes. Pelo exposto acima, torna-se possível prever o fenômeno do *churn*.

Lejeune (2011) afirma que as perdas decorrentes de *churn* podem ser substancialmente reduzidas através do desenvolvimento de uma estrutura adequada

de CRM, que inclui funcionalidades de *data warehousing*¹ e *data mining*². No contexto bancário, as evidências empíricas denotam que, previamente a ocorrência do *churn*, algum ou vários sinais, de forma isolada ou conjunta, são marcadores sintomáticos de um processo de *churn*, assumindo uma característica recorrentemente preditora (VAN DEN POEL; LARIVIÈRE, 2004). Estes sinais, em forma de dados, apresentam a evolução da história do cliente com a organização e revelam a tendência ao *churn*, abrangendo desde o cancelamento de um produto, a redução na quantidade de transações, a transferência de domicílio dos proventos ou a migração do portfólio de investimentos para outra instituição, além de reclamações em canais de ouvidoria, dentre muitas outras, por exemplo.

Amparada na teoria e abordagens do Marketing de Relacionamento, o CRM produz insights importantes a partir da identificação de necessidades dos clientes, o que contribui para melhorar a percepção destes, eleva as taxas de performance nas ofertas e abordagens, além de contemplar toda a gestão de relacionamento com cliente e a definição de prioridades. Para Payne (2006), as estratégias de CRM, portanto, têm em seu propósito maior sistematizar práticas de relacionamento com vistas a elevar a satisfação dos clientes e aumentar sua rentabilidade através de novos negócios/transações, mas também para blindar a ocorrência do *churn*, ponto que ganha relevância como fator essencial para elevação sustentável do patamar de resultados das organizações, conforme evidencia o mesmo autor.

Para Payne (2006), o CRM é um processo que contribui de forma consistente com a implementação e execução da estratégia corporativa, uma vez que suas informações apresentam-se mais tempestivas do que indicadores financeiros tradicionais. As razões para adoção de sistemas de CRM são muitas: melhorar o relacionamento com os clientes, diminuir custos, elevar a eficiência e produtividade, aumentar as receitas e a lucratividade. Além disso, os sistemas de CRM convergem as organizações para o marketing centrado no cliente distanciando-as do marketing centrado no produto. Esse posicionamento reconhece o protagonismo e poder do cliente.

¹ Data Wharehousing é uma estrutura que concentra dados e informações relevantes para a gestão corporativa.

² Data mining tem convencionada sua tradução para “mineração de dados”, que consiste em ferramentas e técnicas com capacidade de descoberta de conhecimento relevante dentro do Data Warehousing, através do encontro de correlações e padrões dentro dos dados armazenados.

A evolução do comportamento do cliente, cada vez mais acostumado ao avanço tecnológico, configura maior complexidade a este momento. Marroquín (2014) definiu o perfil do consumidor na dita Era/Sociedade da Informação: com mais acesso à informação e com expectativas cada vez maiores em relação a qualidade de produtos, serviços e atendimento, os consumidores estão mais conscientes e racionais, não toleram a falta de transparência e exigem marcas comprometidas com o bem-estar das pessoas, ambientalmente responsáveis e que compartilhem valor com a sociedade. A realidade posta é que os clientes estão cada vez mais conectados, informados, conscientes e exigentes.

No mercado financeiro, contexto da presente pesquisa, essa perspectiva é ainda mais emblemática em virtude da transformação digital. Para Rogers (2016), a forma como você se conecta com seus clientes e cria valor para eles foi atropelada pelas tecnologias digitais. Liébana-Cabanillas et al. (2016) apresentaram resultados empíricos, no contexto de um banco espanhol, sobre a relação entre experiência do usuário e a satisfação. Conceitos como User Experience (UX) e Omni-channel são essenciais, uma vez que é preciso estar presente em todas as plataformas e proporcionar a melhor experiência.

Neste ambiente de transformação digital, Rogers (2016) ressalta que a gestão de dados consiste, talvez, no maior dos desafios das organizações – criar informações valiosas a partir dos dados que você possui. O autor destaca que a inovação era uma atividade considerada de alto risco, cara e imprevisível, excessivamente dependente de intuição e suposições. Atualmente, com métodos e ferramentas ágeis para criação, prototipagens e experimentações inclusive virtuais, as ideias e soluções rapidamente podem ser testadas, inclusive em comunidades de usuários, o que significa errar rápido e errar barato, corrigir desvios rapidamente e retomar o caminho da satisfação dos usuários. Todo esse suporte da tecnologia, no entanto, nos remete a ir sobre o conceito de valor para o cliente. O que os clientes valorizam pode mudar rapidamente e os concorrentes continuamente buscam novas oportunidades, que podem ser valorizadas pelos clientes. Rogers (2016) pondera que a competição já não se limita a rivais do mesmo setor e o mercado financeiro está mudando no mundo todo com a entrada de novos players.

A dinâmica do mercado financeiro brasileiro também converge para o exposto por Rogers (2016), cuja entrada de novos players já está transformando o setor. Além das promissoras estratégias por parte dos grandes players globais de tecnologia, os

bightechs – Google, Amazon, Facebook, Apple, Samsung, Alibaba, há o notável avanço das *fintechs*³, que já são mais de 600 iniciativas no Brasil (FINTECHLAB, 2019), cujas atuações abrangem praticamente todas as áreas de atuação dos bancos convencionais, entregando soluções que vão desde meios de pagamentos e seguros a conta corrente e cartões de crédito, de gerenciamento financeiro e operações de crédito a soluções em investimentos, por exemplo. Normalmente com estruturas enxutas e altamente eficientes, as fintechs destacam-se preponderantemente pelo foco na experiência do usuário/cliente. Conforme relatório Pulse of Fintech/KPMG (2017), os investimentos globais nestas organizações ultrapassaram US\$ 120 bilhões nos últimos três anos. Estima-se que em 2025 até um terço das receitas dos bancos convencionais poderá migrar para as fintechs.

A competitividade entre as instituições do setor financeiro é reconhecida em todo mundo, bem como o pragmatismo em relação aos seus resultados financeiros. Entretanto, no Brasil, evidenciam-se algumas peculiaridades. Conforme dados atuais do Banco Central do Brasil (BACEN), existem mais de 150 instituições financeiras em operação no Sistema Financeiro Nacional (SFN), além de cerca de 400 outras entidades financeiras (sociedades de crédito e de investimentos, corretoras e distribuidoras de títulos, valores mobiliários e câmbio). No entanto, cinco grandes bancos dominam o mercado concentrando mais de 80% do total de ativos do SFN (BACEN, 2018).

Com intuito de aumentar a competição entre as instituições financeiras, o BACEN, continuamente têm avançado em uma agenda no sentido de prover ao cliente bancário maior nível de autonomia e poder de escolha. Alguns exemplos são as Resoluções nº 3.402 e 3.424 do CMN, que introduziram a Livre Opção Bancária (LOB), que é a possibilidade de o cliente receber seus proventos na Instituição Financeira (IF) de sua preferência, independente de convênio do seu empregador. Em fevereiro de 2018, o CMN editou a Resolução nº 4.639 estendendo essa possibilidade e contemplando portabilidade de proventos através da LOB para contas em bancos/plataformas digitais. Também há disposições como a Resolução CMN nº 4.292, que dispõe sobre a Portabilidade de Crédito, que é a possibilidade de transferir, também, as operações de crédito à outra Instituição Financeira escolhida pelo cliente.

³ Fintech é um termo cuja origem é a junção das palavras *finance* e *technology* e refere-se a *startups* do setor financeiro que usam a tecnologia para gerar inovações profundas e tornar os sistemas financeiros mais eficientes. (MCAULEY, 2015).

Ponderados os aspectos acima, em um mercado mais livre e flexível que se evidencia, cuja tendência do setor bancário no Brasil é de intensificação do ambiente de concorrência, com cada vez mais provedores de serviços e soluções financeiras atuando. Conforme Rafique, Evans e Nawaz (2015), grandes níveis de concorrência, às vezes, concebem um outro fenômeno, aderente às abordagens propostas por Chan Kim⁴ do “Oceano Azul”: a comoditização. Inovações, avanços tecnológicos e lançamentos de novos produtos, serviços ou soluções, rapidamente são copiadas por concorrentes. Sob o ponto de vista dos clientes, logo, percebe-se cada vez menos diferenças entre as organizações, que dispõem de portfólios, muitas vezes, praticamente idênticos. Essa percepção eleva a sensibilidade dos clientes a fatores como preço e qualidade, e torna mais tênue as barreiras relacionadas a fidelidade.

Conforme Barney (1991), um ambiente competitivo impõe a necessidade de colocar no cerne da estratégia organizacional a criação de vantagem competitiva. Pressupõe a geração e entrega de benefícios advindos da oferta competitiva da organização, consistindo nos conhecimentos especializados, tecnologias e processos de negócios construídos, aperfeiçoados e distinguidos ao longo do tempo.

Neste cenário, o presente estudo tem por contexto um dos principais players do mercado financeiro brasileiro, o Banco do Brasil S.A. (BB). Uma das 100 maiores instituições financeiras do mundo em volume de ativos, bicentenária, fundada em 1808, encerrou o ano de 2018 com R\$1,44 trilhões de ativos em carteira. Por sua composição social, o BB é um banco de capital aberto que tem como acionista majoritário a União Federal (Brasil), e o restante de seu capital estando em livre circulação (*free float*) na BOVESPA. Portanto, ao tratar-se de um banco público, o que o sujeita a regulamentações que não se aplicam aos pares privados, seus desafios frente a este ambiente de grandes transformações podem ser considerados ainda maiores.

Em 2018, o Banco do Brasil registrou Lucro Líquido Recorrente, indicador que desconsidera receitas extraordinárias, no montante de R\$12,9 Bilhões, representando um incremento de 22% em relação ao ano de 2017 (Banco do Brasil, 2019). Se por um lado a perspectiva sobre o patamar de resultado alcançado é auspiciosa, os analistas de mercado têm apresentado dúvidas em relação a capacidade do Banco

⁴ W. Chan Kim é um teórico de negócios nascido na Coreia do Sul e autor do livro *Blue Ocean Strategy*, que pondera questões sobre a diferenciação em mercados competitivos.

do Brasil em elevar o patamar das receitas, uma vez que grande parte da evolução é representada pela efetividade de ações para conter as despesas operacionais.

Alguns grandes números relativos à base de clientes negociais do Banco do Brasil evidenciam uma aderência ao Princípio 80/20 (Pareto, 2003). Conforme dados da instituição em 2018, a base de 65 milhões de clientes contemplava 23 milhões de correntistas. Entretanto, mais de 80% do resultado organizacional é oriundo de apenas 5 milhões de clientes, evidenciando destacada concentração. Se de um lado toda a extensão da base de clientes representa grande oportunidade e potencial de rentabilização, por outro lado a perda de clientes, alvo natural do assédio de concorrentes, neste cenário que evolui também com novos entrantes, compõe grande ameaça.

Entender o cliente, entregar valor e proporcionar a melhor experiência de relacionamento são imprescindíveis para sustentabilidade da organização (KUMAR; PANSARI, 2016). Para estes objetivos o BB dispõe de um modelo de relacionamento desenvolvido para cada perfil de clientes, com propostas de valor aderentes às necessidades de cada segmento. A atuação contempla o uso de ferramentas e sistemas de CRM, com dados capturados em bases internas e externas de informações cadastrais e comportamentais, maturados com soluções de analytics, mineração de dados e big data. Busca-se promover e entregar melhor experiência aos clientes, transformar dados em efetivas informações, possibilitando conhecer e tratar cada um dos mais de 65 milhões de clientes como único.

1.1 DEFINIÇÃO DO PROBLEMA DE PESQUISA

Compreender profundamente as dinâmicas de uma base de clientes, os ciclos de relacionamentos e, principalmente, a evasão de clientes, compõem objetivos prioritários deste trabalho. Outros grandes motivadores para configuração desta pesquisa consistem em questionamentos como: Como promover a rentabilidade de um cliente? Por que a rentabilidade do cliente diminui? O que significa *churn* no contexto bancário? Por que ocorre o fenômeno do *churn*? Como evitar a ocorrência do *churn*? Que entrega de valor é necessária para reter um cliente? Há equilíbrio entre a extração e a entrega de valor na relação com o cliente? O que move um cliente a migrar para um concorrente? A satisfação elimina a evasão? Qual a relação entre satisfação e lealdade? Como promover a lealdade? Qual a influência dos custos de

mudança em relação à lealdade? Como promover o alongamento do ciclo de vida dos clientes com a empresa? O cliente emite sinais antes de migrar para um concorrente? Estes sinais que podem sinalizar a ocorrência do *churn* podem ser capturados na forma de dados? Quais são esses dados? Como utilizar estas informações para promover a retenção? Existem ações que podem reverter a tendência de perda ou abandono do cliente? Quais as melhores práticas de retenção? Quanto gastar com cada cliente de forma a retê-lo? Quais estratégias de retenção devem ser adotadas? Entre outras.

Desta forma, em decorrência do impacto financeiro que a perda de cliente representa para a organização, esta dissertação busca aprofundar análise teórica e prática com vistas a melhorar a compreensão sobre modelagens preditivas de *churn*. A base teórica que ampara a presente pesquisa abrange linhas emergentes do Marketing de Relacionamento, notadamente pelas lentes e metodologias do *Customer Value Theory* (KUMAR, 2018) e *Customer Lifetime Value* (GUPTA; LEHMANN, 2005). Este estudo procurou investigar a seguinte questão-problema: Como identificar os clientes de uma instituição financeira que tendem ao fenômeno do *churn*?

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Comparar modelos preditivos para identificação de clientes de uma instituição financeira que tendem ao fenômeno do *churn*.

1.2.2 Objetivos Específicos

- a) Definir parâmetros que caracterizem o *churn* no contexto bancário, inclusive parcial e em nível de produto;
- b) mapear variáveis preditoras do fenômeno do *churn*;
- c) modelar a probabilidade de ocorrência do *churn*, em nível parcial e total, em amostras de teste;
- d) avaliar o modelo em amostras de validação;
- e) identificar clientes que tendem ao fenômeno do *churn*.

1.3 DELIMITAÇÃO DA PESQUISA

Este estudo restringe-se ao contexto institucional do Banco do Brasil S.A., não podendo ter seus parâmetros e diretrizes de pesquisa considerados imediatamente generalizáveis. Buscou-se identificar clientes que tendem ao fenômeno do *churn*, ainda que em nível parcial, configurado pela redução da quantidade de produtos observada pelos clientes, através das abordagens de análise econométrica e estatística com capacidade preditiva pelas técnicas da Regressão Logística, Regressão Logística Stepwise, Random Forests e Redes Neurais, sobre uma amostra de 2343 clientes pessoa física fornecida pelo Banco do Brasil. Portanto, todos resultados, inferências e conclusões restringem-se a estes atributos.

1.4 JUSTIFICATIVA E RELEVÂNCIA

Conhecer com antecedência quais clientes estão em iminente situação de *churn* permite que sejam implementadas ações proativas para reverter esta tendência. Este estudo contribui para formação do conhecimento na área, uma vez que à exceção de Botelho e Tostes (2010), cujo contexto de pesquisa é uma operadora de cartões de crédito, são escassos os registros teóricos sobre abordagens preditivas para subsidiar a gestão da retenção em instituições financeiras brasileiras e, de uma forma geral, das empresas no Brasil, o que, dentre outros fatores, pode ser característico de uma cultura incipiente de abordagens preditivas voltadas a retenção.

O estudo de campo, a ser conduzido numa grande instituição financeira brasileira, uma das 100 maiores do mundo, cuja base de clientes compõe um montante superior a 65 milhões de clientes, revela um contexto raramente explorado em pesquisas científicas, o que o destaca.

A pesquisa de Reichheld e Sasser (1990) mostrou que, naquele contexto de uma agência bancária, a redução de 5% na taxa de *churn* poderia representar 85% de incremento no resultado financeiro ao longo de alguns anos. Evidentemente, em um cenário com alta competitividade, esses números são muito representativos e reforçam a importância de gerir o fenômeno do *churn*.

Conforme Gupta e Lehmann (2005), em uma abordagem de marketing ao amparo das métricas do cliente, quando as empresas buscam elevar o patamar de

seus resultados, as estratégias organizacionais convergem normalmente para três pontos críticos:

- a) gestão do cliente (ações para promover a margem de contribuição do cliente);
- b) aquisição de clientes (abrange custo de aquisição e taxa de aquisição/conversão);
- c) retenção de clientes (abrange custo de retenção e taxa de retenção).

Em relação a essas três perspectivas, há uma questão chave que as instituições financeiras têm se dedicado a encontrar respostas: onde é mais adequado concentrar suas esforços e ações para melhorar (ou manter) os resultados? A gestão do cliente compõe uma perspectiva que está no dia-a-dia das empresas, no cerne de suas estratégias, contemplando vertentes entre as duas acepções de valor do cliente (TOLEDO; MORETTI, 2016) - entregar valor para o cliente e recolher valor do cliente. É improvável, na ampla maioria das organizações, que esta perspectiva esteja sendo relegada, pelo contrário. No entanto, a aquisição de clientes as vezes implica a necessidade de investir muito para conquistar cada um, e os resultados com essas ações são inferiores. Kurtz e Clow (1998) demonstraram que o custo de aquisição de um cliente representa até 5 vezes mais do que o custo para reter um cliente. Nesse sentido, é mais do que adequado convergir os esforços para estratégias de retenção, devendo as organizações contemplar, dentre suas práticas mais importantes, a efetiva gestão do *churn*.

A previsão da ocorrência do *churn* é uma informação relevante e ainda pouco explorada pelas organizações, que devem abordar o *churn* de forma sistêmica (VELOSO, 2013). Isto é, com agilidade capaz de mapear o cliente *churner* antes de consumada a evasão e ainda em tempo de agir para reverter esta tendência. A compreensão, por parte das instituições, deste fenômeno do *churn*, pode representar grande oportunidade. Com este propósito, de forma construtiva, objetiva-se aprofundar a análise deste problema e a proposição de métodos para combater ou reduzir o *churning*.

Pela análise econométrica da amostra recebida da instituição financeira, conforme Quadro 1 a seguir, foi observada uma oscilação média de -14,65% sobre a Margem de Contribuição gerada pela amostra de 2343 clientes em relação aos valores observados pelos mesmos clientes 12 meses antes (ex. Ref. out/18 sobre out/17). Esta comparação foi possível nos meses com mais de uma posição anual.

Quadro 1 - Evolução da Margem de Contribuição na amostra de trabalho

Evolução da Margem de Contribuição (MC) na amostra de trabalho			
Mês	MC Total	MC Média por cliente	Oscilação sobre observado 12 meses antes
OUT/17	R\$ 1.047.027	R\$ 446,87	Sem base de comparação
NOV/17	R\$ 1.111.118	R\$ 474,23	
DEZ/17	R\$ 976.492	R\$ 416,77	
JAN/18	R\$ 1.047.399	R\$ 447,03	
FEV/18	R\$ 976.658	R\$ 416,84	
OUT/18	R\$ 940.592	R\$ 401,45	-10,17%
NOV/18	R\$ 877.247	R\$ 374,41	-21,05%
DEZ/18	R\$ 861.243	R\$ 367,58	-11,80%
JAN/19	R\$ 884.130	R\$ 377,35	-15,59%
FEV/19	R\$ 833.761	R\$ 355,85	-14,63%
			Δ Média = -14,65%

Fonte: Elaborado pelo autor (2019).

Ainda que nossa observação delimita-se a uma pequena amostra de 2343 clientes que serviram de base para o quadro 1, tendo o resultado financeiro do Banco do Brasil apresentado patamares crescentes ao longo do tempo, afere-se que a tendência decrescente de margem de contribuição apresentada pela base de clientes pré-existente é recorrentemente suprida pela aquisição de novos clientes, que não é o foco deste estudo. Se projetarmos esta redução de 14,65% aos indicadores de resultado do Banco do Brasil, pautando somente pela Margem Financeira Bruta (MFB) gerada pela base de clientes no ano de 2018, que foi R\$49,6 bilhões (BANCO DO BRASIL, 2019), o fenômeno da desidratação pode representar, neste componente do resultado, cerca de R\$7 bilhões somente em 12 meses.

A perda de negócios e clientes, objeto deste estudo, têm impacto no Valuation⁵ das organizações. Através do Customer Equity⁶, por exemplo, uma das principais métricas do marketing, em cálculo elaborado pela abordagem proposta por Rust, Lemon e Zeithaml (2004), por meio de simulador online⁷, pautado também somente pela Margem Financeira Bruta observada nas demonstrações de Resultado do Banco do Brasil em 2018, desconsiderando outros componentes/receitas não recorrentes, verifica-se um Valuation que alcança R\$308 bilhões, conforme quadro 2. Assumimos como referência para Retenção Estimada um indicador conservador de 93%, aderente

⁵ Valuation: Métricas de avaliação e valoração de empresas.

⁶ Customer Equity: Técnica de valuation que reconhece aos clientes toda a fonte de rendimentos de uma empresa, ponderado o ciclo de vida do cliente com a empresa. O Customer Equity dimensiona, em valor presente, todas as receitas atuais a futuras de toda a base de clientes da organização (RUST; LEMON; ZEITHAML, 2004).

⁷ Simulador online de Customer Equity, nos parâmetros de Rust et al. (2004), disponível em: <<https://www.clv-calculator.com/free-excel-template-for-calculating-customer-equity/>>.

a Van den Poel e Larivière (2004), o que configura um índice de *churn* de 7%. Neste exemplo hipotético, a perda de clientes anualmente pode representar, para o Banco do Brasil, mais de R\$21 bilhões⁸. Essa questão norteou as definições da presente pesquisa, reforçando ainda mais a necessidade de gerir o fenômeno do *churn* e da desidratação.

Quadro 2 - Simulador Customer Equity

Simulador de Customer Equity	
Informações atuais sobre a base de clientes	Customer Equity
Margem Financeira Bruta Consolidada (R\$bi)	49,625
Retenção estimada	93%
Taxa de desconto (Ref. SELIC/COPOM vigente)	6,5%
CUSTOMER EQUITY (*) (**) (R\$bi)	308,287
(*) Em relação a base de clientes pré-existente, desconsiderada aquisição;	
(**) Ponderado a um prazo de 15 anos.	

Fonte: Elaborado pelo autor (2019).

A condução dos trabalhos, com vistas a contribuir para o gerenciamento do *churn*, estará pautada em compreender este fenômeno, que poderá nortear ações com vistas a manutenção da lucratividade dos clientes e capaz de aumentar a fidelidade do cliente, exatamente em aderência a Lejeune (2011), que enfatizava o propósito organizacional de tratar os clientes de tal forma que eles permaneçam clientes ainda que um concorrente proponha condições mais vantajosas.

Uma abordagem ao gerenciamento do *churn* é essencialmente proativa, em que inicialmente deve-se ser capaz de se prever o *churn* e depois planejar uma estratégia apropriada para evitá-la, afirmam KhakAbi, Gholamian e Namvar (2010). Destaca, ainda, que um instrumento capaz de auxiliar nas duas etapas mencionadas, elevando a precisão e a velocidade da previsão, além de contemplar a proposição de estratégias pré-definidas para clientes/segmentos específicos é a mineração de dados.

A mineração de dados é o processo que usa inteligência estatística, matemática, inteligência artificial e técnicas de aprendizado de máquina para extrair e identificar informações úteis e, subsequentemente, obter conhecimento de grandes bases de dados. (KHAKABI; GHOLAMIAN; NAMVAR, 2010). As técnicas de mineração de dados ajudam as organizações a obter conhecimento valioso de seus enormes bancos de dados de clientes.

⁸ O produto entre Customer Equity e índice de churn: (R\$bi) 308,287 x 7% = R\$21,58 (Bi)

A condução da pesquisa estará pautada pela busca de elementos que contribuam para melhor compreensão do fenômeno do *churn*, identificação de variáveis preditoras, modelagem e avaliação de técnicas preditivas, e identificação dos clientes que tendem ao fenômeno do *churn*. Espera-se viabilizar a extração de diretrizes que possam contribuir para melhor gestão das carteiras de clientes.

Entende-se que a presente pesquisa poderá contribuir com subsídios e elementos generalizáveis a quaisquer organizações, inclusive não atuantes no mercado financeiro, de forma a respaldar a adoção de estratégias de retenção por meio da gestão do fenômeno do *churn*.

1.5 ESTRUTURA DE TRABALHO

Esta dissertação está estruturada em cinco capítulos, contando com esta introdução, que é o primeiro. O segundo capítulo contempla o referencial teórico que serviu de base para realização desta pesquisa. No terceiro, é abordada a metodologia que foi utilizada. No quarto capítulo são analisados os resultados visando responder aos objetivos. Por fim, o capítulo cinco apresenta as considerações deste estudo.

2 REFERENCIAL TEÓRICO

2.1 MARKETING DE RELACIONAMENTO

O Marketing de Relacionamento surge pela adoção de um novo paradigma, o da troca relacional. Contempla a necessidade de mudança de foco, em que as empresas deixam de estar orientadas a transações de curto prazo amparadas no paradigma tradicional da transação direta, cujo foco é a maximização da venda, e caminham para um objetivo de construção de um relacionamento de longo prazo (KOTLER, 1991). O consumidor deixa de ser compreendido como um sujeito anônimo e independente.

Payne (2006) sugere três características distintivas do marketing de relacionamento em relação ao marketing transacional. Já a primeira é a ênfase na retenção de clientes e a ampliação do "valor vitalício" dos clientes por meio de estratégias voltadas para a retenção. A segunda é o reconhecimento de que as empresas precisam desenvolver relacionamentos se quiserem obter sucesso a longo prazo. A terceira característica do marketing de relacionamento é que o marketing precisa ser visto como uma responsabilidade de toda a empresa ou multifuncional, e não apenas uma preocupação do departamento de marketing, contemplando uma ideia holística do Marketing de Relacionamento, em que todas as áreas, pessoas, processos e operações devem estar integradas.

Guimarães, Milani e Crescitelli (2005) salientam que Marketing de relacionamento não é simplesmente cadastrar pessoas em uma base de dados e enviar-lhe carteirinhas. Inclusive o relacionamento pode ficar comprometido se o cliente descobrir que o atendimento e condições que está tendo não são tão exclusivas, seja nas vendas, ações promocionais ou ofertas. Ainda, a construção de um relacionamento duradouro não pode estar baseada em descontos financeiros. Neste caso, qualquer concorrente que ofereça alguma vantagem similar poderá fazer com que o cliente mude de posição. Também, condicionar o cliente apenas a recompensas em decorrência do volume e frequência de compras é uma forma inadequada de estabelecer o relacionamento. É essencial saber utilizar o banco de dados para alicerçar as ações de relacionamento, construindo reações satisfatórias dos clientes além da questão financeira, contemplando perspectivas emocionais e

comportamentais, superando expectativas e promovendo um sentimento de valorização. Esse é o caminho para gerar um vínculo mais sustentável.

Portanto, conforme Barreto, Crescitelli e Figueiredo (2015) destacam, o papel do Marketing de Relacionamento é ser 'uma alternativa capaz de desenvolver valor superior por meio do atendimento das necessidades dos clientes e do oferecimento de produtos adequados a cada cliente individual'. Isto é, trata de atrair, manter e aumentar as relações com clientes, em consonância a Payne (2006). De forma extrema, além disso, também proporciona mapear clientes que talvez seja desejável, inclusive, encerrar a relação (GRÖNROOS, 2000).

Aderente ao propósito de elevar o patamar de fidelidade dos clientes e baseado nos princípios de marketing de relacionamento, com as novas demandas do mercado e novas tecnologias, o gerenciamento do relacionamento com os clientes foi levado a um novo e mais complexo nível. Conforme Payne (2006), o CRM (Customer Relationship Management) é uma resposta a esse ambiente novo e mais desafiador e simplifica: - "o CRM é marketing de relacionamento com base em informações". Destacam-se outros pontos conforme abaixo:

- clientes compõem ativo para organização;
- a maior utilização de tecnologia na gestão e maximização do valor da informação;
- reconhecimento dos benefícios do uso proativo de informações;
- aceitação da necessidade de trade-off entre entrega e extração de valor para o cliente;
- o desenvolvimento de abordagens de marketing individual (um a um).

O CRM consiste na sistematização de ações interativas relacionais e transacionais para fortalecer o vínculo com os clientes e, naturalmente, rentabilizar a relação pela percepção de maior entrega de valor na própria relação, o que contribui para elevação da margem de contribuição do cliente e a extensão do ciclo de relacionamento (PAYNE, 2006).

O CRM surge, portanto, como uma ferramenta estratégica eficaz que emprega tecnologias de informação avançadas para desenvolver, manter e aprimorar o relacionamento com o cliente (LIN et al., 2009). Estes autores salientam que o foco

em gerar valor para o cliente é a semente do CRM, que define e identifica os clientes mais importantes, satisfaz suas necessidades e expectativas, eleva o relacionamento de longo prazo e molda a fidelidade do cliente. O ambiente de concorrência hostil e agressivo no mundo contemporâneo impõe que o CRM é imprescindível, porque as empresas precisam manter seus clientes existentes e simultaneamente atrair novos clientes potenciais. Inobstante, o CRM eleva ao topo de suas virtudes as suas competências preditoras, propiciando a implementação de estratégias de retenção para blindar a evasão de clientes marcados com grande potencial de evasão.

O aprendizado contínuo através do maior conhecimento do cliente e incremento de lucratividade de longo prazo constituem-se, talvez, nos dois aspectos mais relevantes do CRM. Essas estratégias proporcionam elevar maior participação nos hábitos de consumo do cliente (share of wallet), ao invés de maior participação de mercado (BOTELHO; TOSTES, 2010). Deixa-se de procurar o cliente certo para seu produto e passa-se a procurar o produto certo para o seu cliente. Evidentemente o foco está na maior entrega de valor para o cliente e elevar a satisfação.

Entregar valor na relação cliente-banco, retomando o contexto do presente estudo, é, portanto, imprescindível para evitar que o assédio dos concorrentes possa acarretar uma possível evasão. Isso encontra consonância com as definições de Valor para o Cliente de Zeithaml (1988). Além disso, inobstante, é necessário ser efetivamente percebido pelo cliente nesta entrega, conforme ressaltam Mello e Leão (2008).

Ainda, se faz necessário examinar os significados do constructo “valor do cliente”. Uma acepção refere-se ao valor que uma empresa fornece aos seus clientes, no sentido de valor para o cliente. Outra interpretação é no sentido do valor que os clientes fornecem ou representam à empresa, em uma conotação mais financeira, em consonância ao apresentado por Gupta e Lehmann (2005).

O foco deste trabalho concentra-se em mitigar o *churn*. Sob a perspectiva do Marketing de Relacionamento, um dos fatores-chaves em relação a ocorrência do *churn* é a lealdade (fidelidade) do cliente. Pensar que a lealdade estará consolidada apenas por que os clientes estão satisfeitos é um grande engano. Reichheld (1993) demonstrou que, a depender do ramo de atividade, de 65% a até 85% dos clientes que trocaram de fornecedor manifestaram satisfação em relação ao fornecedor antigo.

Barreto, Crescitelli e Figueiredo (2015) apresentam como produto final do Marketing de Relacionamento, juntamente com a lealdade, o *Customer Lifetime Value*

– CLV, porém este em um nível de 2º grau. Isto porque o Marketing de Relacionamento não possui relação direta com o CLV. O MR proporciona elevação dos níveis de fidelidade, já a fidelidade sim está diretamente ligada ao CLV. Clientes fiéis tem maior horizonte de relacionamento, consomem com maior frequência e em maior valor. Ainda em sua abordagem, Barreto, Crescitelli e Figueiredo (2015) levantaram os elementos iniciais do MR com impacto direto sobre fidelidade, destacando-se a satisfação, qualidade, comprometimento, confiança, personalização e custos de mudança. A quadro 1 consolida a conceituação destes constructos, fazendo-se necessária para avançarmos sobre a perspectiva de fidelidade.

Quadro 3 - Principais constructos do Marketing de Relacionamento com impacto em fidelidade

Constructos	Definição
Satisfação	Satisfação é a sensação prazerosa de que o acumulado de eventos de consumo com um dado produto ou marca supriu algum objetivo ou desejo. Ocorre satisfação quando um produto recebe expectativa positiva do consumidor, atende a essa expectativa e recebe afeto positivo. A satisfação pode se dar tanto em relação ao produto em si, quanto à relação do consumidor com a empresa/funcionários.
Qualidade	O quanto um produto (tanto o produto em si quanto a forma como é executado e entregue – relação entre fornecedor e cliente) se aproxima do ideal e apresenta baixa ocorrência de erros.
Comprometimento	Constructo subdivide-se em: Afetivo: O comprometimento afetivo ocorre quando o consumidor de fato gosta de dado fornecedor, e por isso resiste a abandoná-lo. Calculativo: O comprometimento calculativo ocorre quando o consumidor sente que encerrar o relacionamento com o fornecedor pode acarretar sacrifícios econômicos ou sociais.
Confiança	Confiança é a intenção de uma parte de se colocar de forma vulnerável às ações de outra parte. Ocorre confiança quando se percebe que a outra parte possui habilidade (competência), benevolência (tendência de fazer o bem para a outra parte envolvida) e integridade (respeita princípios e valores).
Personalização	Personalização é o esforço da empresa fornecedora para que o produto esteja o mais próximo possível das necessidades individuais de cada cliente. Personalização pode ocorrer tanto no produto em si quanto no contato/comunicação cliente X prestador.
Custos de mudança	Procedimento: Custos que podem deter os consumidores a mudar para outro fornecedor em razão do tempo demandando pela troca de fornecedor. Financeiro: Custos que podem deter os consumidores a mudar para outro fornecedor em razão de perdas financeiras (como pontos de fidelidade) ou custos únicos de inicialização (como matrículas ou adaptações de ativos específicos). Relacionamento: Custos que podem deter os consumidores a mudar para outro fornecedor em razão de perdas emocionais relacionadas à troca de fornecedor.

Fonte: Elaborado pelo autor, com base nos conceitos apresentados por Barreto et al. (2015).

Steinmann, Murphy e Mehta (2016) ponderam alguns pontos sobre lealdade. Embora existem outras variações conceituais para tipos de lealdade, destacam-se duas formas principais: lealdade comportamental e lealdade atitudinal. A lealdade comportamental denota que alguns clientes são leais porque precisam. É o caso de você que compra na padaria localizada na esquina de sua casa porque é o único estabelecimento nas proximidades. Já na lealdade atitudinal/emocional, os clientes são leais porque amam a marca ou produto. Salieta-se que é possível que você tenha lealdade atitudinal em relação a sua padaria, talvez seja até provável. Porém a configuração da lealdade se dá mais por aspectos de conveniência. Logo, para as empresas e marcas, é sempre preferível despertar em seus clientes a lealdade atitudinal, por várias razões: seu cliente será menos suscetível a concorrentes; disposição em pagar preços mais elevados; maior probabilidade de defender “suas” marcas de preferência; Um dos grandes pensadores do Marketing, Phillip Kotler sintetiza ao afirmar que se você criar um caso de amor com seus clientes, eles próprios farão sua publicidade.

Steinmann, Murphy e Mehta (2016) salientam que, para se sustentar no longo prazo, as empresas precisam compreender em profundidade os conceitos de evasão (*churn*) e retenção. Entender o porquê e com que frequência os clientes se mantem leais e permanecem consumindo os produtos e serviços da empresa passa a ser um fator chave. Por maior que seja seu mercado, a conquista de clientes é excessivamente onerosa. Está claro, portanto, a necessidade de preservar a base de clientes. Para superar elevadas taxas de evasão a empresa depende da lealdade de seus clientes, logo é imprescindível compreender o significado dessa palavra.

Kumar, Pozza e Ganesh (2013) realizaram revisão bibliográfica e estudo de campo com objetivo de aprofundar esta relação entre satisfação e lealdade, trazendo importantes questões que devem ser consideradas na análise da eficácia da satisfação do cliente para explicar e prever a fidelidade do cliente. É recorrente em todo mundo empresas investindo pesadamente em satisfação do cliente, na esperança de aumentar a lealdade e, conseqüentemente, a lucratividade. Pautados por muitos estudos acadêmicos sobre esta relação, evidenciaram, entretanto, alguns achados contrastantes. Após a análise, pontos relevantes desta revisão e generalizações empíricas que a acompanham, seguem conclusões que guardam pertinência à presente dissertação:

- no geral, existe uma relação positiva entre a satisfação do cliente e a lealdade;
- a variância, no entanto, explicada apenas pela satisfação é bastante pequena – apenas 8%;
- modelos holísticos que englobam outras variáveis relevantes como moderadores (custos de troca, nível de competição, ambiente/canal transacional), mediadores (confiança, comprometimento, valor relacional), variáveis antecedentes ou todos os três são melhores preditores de lealdade do que modelos com apenas satisfação do cliente;
- a inclusão dessas variáveis aumenta a variância explicada, em média, para 34% (54% para lealdade atitudinal e 15% para lealdade comportamental, respectivamente);
- o relacionamento de satisfação e lealdade tem o potencial de mudar ao longo do ciclo de vida do cliente;
- embora a satisfação do cliente tenha uma relação positiva com o boca-a-boca (recomendações pessoais), outras variáveis relacionadas, como comprometimento, confiança e envolvimento do produto, servem como melhores preditores;
- a satisfação do cliente, por si só, nem sempre resulta em retenção, alongamento do tempo de vida do cliente e recorrência de consumo. As variáveis de cliente, relacionamento e mercado geralmente desempenham um papel significativo de moderação;
- a relação entre a satisfação do cliente e a compra cruzada é caracterizada por resultados contrastantes. O nível de agregação usado para analisar os dados pode afetar a força do relacionamento.

Santos e Porto (2014) destacam que a lealdade consiste em um profundo comprometimento do consumidor pela recompra ou favorecimento a uma marca ou produto, apesar do contexto situacional e dos esforços de marketing com potencial de causarem um comportamento de troca.

Payne e Frow (2005) sugerem uma estrutura conceitual para o CRM e salientam os processos-chave para sua efetividade, destacando o processo de

criação de valor, ampliando o entendimento do CRM por seu papel na melhoria do valor do cliente e, como resultado, do valor para o acionista.

Para Kotler (2013), os mercados não são homogêneos e é imprescindível que as empresas o entendam, de forma a atender necessidades e desejos específicos. Grupos heterogêneos devem ser agrupados a partir de uma ou mais características, de forma a torná-los homogêneos. Por aprofundar o conhecimento sobre cada segmento de cliente, a organização estará mais preparada para entregar valor na relação. Em toda literatura da gestão contemporânea ressalta-se a necessidade de ser eficiente na alocação de recursos. Evidencia-se a necessidade de segmentação da base de clientes e a oferta de um atendimento compatível com o potencial de cada perfil.

Peppers e Rogers (1997), são autores seminais do chamado marketing um-a-um. Desenvolveram uma tipologia para auxiliar as empresas a classificar seus clientes e identificar estratégias relevantes para cada segmento de clientes. Sugerem classificar os clientes em cinco grupos, em termos de seu valor vitalício potencial e propõem que a ênfase seja colocada nos três grupos conforme abaixo:

- os MVCs ou clientes mais valiosos são aqueles clientes que têm os maiores valores hoje. Representam o núcleo dos negócios atuais de uma empresa e o objetivo do CRM em relação a eles é a retenção de clientes;
- os STCs ou clientes de segundo nível representam os clientes com o maior potencial não realizado. Podem ser uma fonte maior de lucratividade para a empresa e o objetivo do CRM em relação a eles é o crescimento e ampliação do volume de negócios;
- BZs ou abaixo de zero clientes são clientes que provavelmente nunca irão gerar lucro suficiente para justificar a despesa envolvida em atendê-los. Em relação a estes, o objetivo do CRM é minimizar gastos e até propor o desinvestimento.

Para Song et al. (2017), de forma convencional, o CRM é normalmente baseado em modelos de RFM (Recency, Frequency e Monetary). Pela necessidade de as organizações segmentarem clientes, Hughes (1994), de forma simples e bastante efetiva, propôs um modelo analítico, RFM (recency, frequency & monetary), que diferencia clientes a partir de três atributos: recência, frequência e valor

monetário, sobre os quais os clientes são classificados por seu comportamento de consumo histórico, formando grupos de clientes (segmentos). Esse modelo soluciona alguns objetivos recorrentes das organizações como classificar os clientes quanto ao retorno e a intensidade de relacionamento, priorizar clientes para ações de marketing direto ou ações de recompensa, gerenciar a rentabilidade dos clientes e analisar as tendências de crescimento ou retração da base de clientes. A abordagem RFM, apenas com essas variáveis, gera subsídios para conhecer os clientes que gastam mais dinheiro e criam o maior valor para as empresas. Apresentamos a seguir o detalhamento das variáveis do RFM e como elas devem ser utilizadas.

- R - Recência, que se refere ao intervalo entre o momento em que o último comportamento de consumo acontece o presente. Quanto menor este intervalo, maior será o R.
- F - Frequência, representa o número de transações em um período específico. Quanto maior a frequência, maior é o F.
- M - Valor Monetário, que se refere ao padrão de consumo em um período específico. Quanto maior o valor gasto pelo cliente, maior será o M.

Fader, Hardie e Lee (2005) salientam que, em toda literatura de marketing, as medidas de comportamento passado dos clientes são, majoritariamente, os principais indicadores de seu comportamento futuro, manifestando críticas sobre esse fato, porém ressaltam que modelar receitas futuras adequadamente representa uma barreira significativa. Fader, Hardie e Lee (2005) apresentaram um modelo integrado de RFM e CLV, ponderando a complementaridade que a vinculação dos dois paradigmas proporciona.

Uma tipologia de valor do cliente como essa é útil também para ajudar uma organização a tomar decisões apropriadas sobre como ela pode utilizar melhor seus recursos. Em aderência a isso e em consonância ao proposto por Kumar et al. (2008), é essencial que as ferramentas de CRM, através do *Customer Lifetime Value*, subsidiem decisões consistentes ao longo do tempo conforme ponderações a seguir:

- a) Quais clientes devo selecionar (segmentar/escolher) ?
- b) Qual nível de recursos a serem alocados aos clientes selecionados ?
- c) Quais clientes devem ser estimulados a aumentar sua rentabilidade futura?

A profunda compreensão da relação entre a execução de ações pela empresa e a lucratividade do cliente são fundamentais para garantir o sucesso das decisões acima. Essa conexão entre CRM e CLV, conforme resultados empíricos apresentados por Kumar et al. (2008), em virtude da redistribuição de recursos baseado no CLV, em estudo piloto com 35.000 clientes da IBM, levou a realocação de recursos para cerca de 14% dos clientes, gerando um incremento na receita de cerca de 20 milhões de dólares, representando um expressivo aumento de dez vezes no retorno financeiro da campanha, sem quaisquer outras alterações no nível de investimento em marketing.

Em relação a isso, Payne (2006) pondera que a empresa precisa ter uma visão clara da lucratividade gerada por seus principais clientes e segmentos de clientes, e principalmente considerar sua lucratividade futura, necessitando identificar o potencial de cada um. Gupta e Lehmann (2005) consideram que há muito ceticismo sobre a utilidade das métricas financeiras no marketing, especialmente quando se tem dificuldade em dimensionar o retorno de investimentos em seus próprios programas, como de publicidade, por exemplo. Gupta e Lehmann (2005) defendem a adoção do que chamam das métricas do cliente. Com este objetivo, aprofundaremos a análise dos conceitos ao amparo do *Customer Lifetime Value*.

2.2 CUSTOMER LIFETIME VALUE

O movimento em direção a uma abordagem de marketing centrada no cliente, juntamente com a crescente disponibilidade de dados de transações do cliente, convergiu pesquisadores e organizações a um grande interesse para as abordagens do *Customer Lifetime Value* (CLV) (FADER; HARDIE; LEE, 2005).

Conforme apresentado por Dwyer (1989), *Customer Lifetime Value* refere-se a um constructo importante para a tomada de decisões dos gestores, notadamente por auxiliar a definição de orçamentos e projetos relativos a programas de aquisição e retenção de clientes. Berger (1998) evidenciou um conceito de CLV como sendo a soma dos fluxos de caixa acumulados - descontados usando o custo médio ponderado do capital - de um cliente durante toda sua vida útil com a empresa. Berger (1998) ainda apresentou sentido ao CLV como a essência do marketing a partir de duas definições de Kotler e Armstrong (1996), conforme abaixo:

Em última análise, o marketing é a arte de atrair e manter clientes lucrativos (Kotler & Armstrong, 1996). Uma empresa não deve tentar perseguir e satisfazer todos os clientes. O que torna um cliente rentável? Kotler e Armstrong (1996) definem um cliente lucrativo como 'uma pessoa, família ou empresa cujas receitas excedem, por um montante aceitável, sobre os custos da empresa em atrair, vender e prestar serviços'. Esse excesso é chamado de valor da vida útil do cliente (CLV). (BERGER, 1998).

O *Customer Lifetime Value* tem como objetivo principal compreender a dinâmica que ocorre dentro da base de clientes e a mensuração do valor do ciclo de vida de cada cliente para organização (GUPTA; LEHMANN, 2006). Veja que, às vezes, a organização pode estar expandindo sua base de clientes, em número absoluto, porém deixando evadir seus clientes mais rentáveis. Portanto, conforme o mesmo autor, o CLV apresenta-se como uma ferramenta que facilitará a tomada de decisões sobre onde e como investir, seja para adquirir ou reter um cliente, onde potencialmente resultará o maior retorno.

O CLV fornece informações consistentes com a perspectiva de clientes serem considerados ativos de longo prazo. Protege a organização de uma visão míope de entregar resultado imediato, as vezes somente de curto prazo e insustentável, o que pode significar o que se chama de "lucros ruins" (STEINMANN; MURPHY; MEHTA, 2016) Isso ganha relevância quando estamos inseridos num contexto de um Banco, em que aproximadamente 60% do resultado recorrente é oriundo de operações de crédito. Logo, mais do que perder clientes, no mercado financeiro, isso pode significar contabilizar prejuízos sobre o capital emprestado.

Bancos e outras empresas que possuem bases ampliadas de clientes, em diferentes estágios de relacionamento, e continuam a atrair clientes novos, podem ter dificuldade de visualizar que a receita recorrente da base consolidada (antiga) pode estar diminuindo. Portanto, a perda de clientes altamente rentáveis ou a redução de sua margem de contribuição, cujo valor vitalício é relevante para organização, pode estar ocultada por novas receitas, oriunda de novos clientes/operações, inclusive, de curto prazo. Por si só, isso já evidencia a necessidade de utilização das Métricas do Cliente, em nível individual de cada cliente, conforme Kumar e Reinartz (2016). Mas ainda, o CLV não trata de sugerir recusar estas receitas de curto prazo, obviamente, porém evidencia a necessidade de, concomitante à importância da aquisição de novos clientes, mitigar a evasão de clientes rentáveis. Percebe-se que muitas empresas não investem o necessário para retenção de clientes cuja aquisição consome tantos recursos (STEINMANN; MURPHY; MEHTA, 2016).

Kumar e Pansari (2016) apresentaram achados em que a cultura e a conjuntura econômica de um país têm impacto direto nos hábitos de consumo e margem de contribuição, exercendo ação moderadora sobre o CLV. Essa informação contribui para definição gerencial de escolha alocação de recursos de marketing, para o caso de empresas multinacionais, por exemplo.

A abordagem do CLV consiste em dimensionar o valor do ciclo de vida de cada cliente para organização, trazendo a valor presente todas receitas recorrentes futuras ponderadas pelos índices de retenção, isto é, enquanto o cliente está ativo. Quando consolidamos o valor de toda base de clientes, sob esta métrica do CLV, a valor presente, configura-se um indicador que se chama Customer Equity, na definição de Rust, Zeitham e Lemon (2001). Há duas maneiras de relacionar esses indicadores. A partir do CLV de todos os clientes, em nível individual, mas também pode-se calcular o Customer Equity diretamente pelos indicadores consolidados de toda organização, também trazendo a valor presente o montante de receitas recorrentes totais e futuras ponderadas pelos índices de retenção. Neste caso, no sentido inverso, do topo para a base, a partir do Customer Equity é passível obter um indicador denominado CLV Médio.

O *Customer Lifetime Value* entrega algumas perspectivas muito importantes para as organizações. Compromete a empresa a colocar o cliente no centro de sua estratégia. Além disso, pela sua definição, concentra-se na rentabilidade a longo prazo ao invés de unicamente acompanhar os índices de participação de mercado ou volume de vendas. Outro ponto importante é que o CLV permite que uma empresa avalie o valor de clientes individuais e direcione-os por meio de ofertas personalizadas. Além disso, conforme Gupta e Lehmann (2006), a melhoria na tecnologia da informação e a disponibilidade de transações em nível de cliente permite que as empresas realizem análises detalhadas em vez de medidas agregadas baseadas em pesquisas, como satisfação.

Ainda além das métricas tradicionais de Finanças e do Marketing, centradas em vendas, lucro, satisfação do cliente e aceitação da marca, pelas quais não se pode necessariamente aferir a qualidade dos resultados de uma empresa, no esteio do Marketing de Relacionamento, as Métricas do cliente, contemplando uma perspectiva de longo prazo, evidenciam vantagens a partir de uma abordagem centrada na rentabilidade por cliente e no valor vitalício do cliente. Gupta e Lehmann (2006) ponderam que o *Customer Lifetime Value* – CLV (Valor Vitalício do Cliente) preenche

essa lacuna, torna o marketing mais responsável e também fornece uma ferramenta fundamental para ajudar as empresas. O CLV surge com o propósito de colocar o cliente no centro da estratégia e tangibilizar o valor dessa relação empresa-cliente, que por definição consiste no valor presente de todos os lucros futuros gerados pelo cliente ao longo da vida de relacionamento com a empresa. Essa dinâmica permite que uma empresa possa investir em seus clientes, proporcionando maior valor para eles, justificadamente para obter um retorno maior deles através de lucros atuais e futuros para a empresa, devendo os programas serem avaliados em termos de seu impacto nas Métricas do Cliente ao amparo do CLV, e os gastos com marketing sendo vistos como investimentos, ao invés de despesas.

Gupta e Lehmann (2005) elencam a importância da empresa se posicionar de forma a estar voltada para os clientes, que devem ser vistos como principal ativo das organizações, e, por isso, todas as decisões estratégicas devem ser tomadas em convergência a esta visão cliente, notadamente com o objetivo de identificar e dimensionar previamente qualquer impacto.

Steinmann, Murphy e Mehta (2016) propõem uma reflexão sobre uma hipótese de que sua empresa aumentou em 10% sua receita recorrente, obtendo um desempenho superior a performance do setor de atuação e questiona: - Parece satisfatório? Até pode ser. No entanto, conforme sugere, é imprescindível conhecer com profundidade mais alguns detalhes que compõem esse indicador, como por exemplo:

- a) Que porcentagem de clientes aumentou o valor de seus contratos?
- b) Que setores/segmentos apresentaram os maiores índices de *churn*?
- c) Quais as taxas de retenção e crescimento por segmento?
- d) Em que porcentagem reduzimos os descontos na primeira renovação?
- e) Qual é a diferença média entre o tamanho do contrato atual e o tamanho do contrato original, de todos os clientes que são clientes há mais de três anos?

Gerenciar uma empresa de maneira adequada pressupõe conhecer esses detalhes em relação a cada transação de maneira crítica, insistem Steinmann, Murphy e Mehta (2016).

O aspecto comportamental das relações entre clientes e os bancos evidencia uma realidade: grandes bases de clientes em que uma parte mantém variados níveis

de interações recorrentes e outra parte que compõe um estoque de clientes inativos. Em consonância está o proposto por Olsen e Parker (2008), em que destacam a necessidade das organizações ter que fazer escolhas ao definir um modelo de atendimento que se compatibilize ao potencial de negócios de cada cliente. Configura-se artificialmente um contexto de limitação de oferta de atendimento, dedicando padrão de entrega de serviços compatível com o retorno esperado de cada base de clientes, também no sentido de desonerar e evitar gastos que podem não gerar retorno. Portanto ressaltada a característica preditiva do *Customer Lifetime Value*, inobstante, é relevante sua aplicação de forma a subsidiar diretrizes para alocação de recursos em segmentos de clientes.

Diversos são os modelos de CLV, porém normalmente convergem, basicamente, aos mesmos parâmetros. A seguir, um modelo genérico e básico apresentado por Berger e Nasr (1998).

$$CLV = \sum_{i=1}^n \frac{(R_i - C_i)}{(1 + d)^{i-0.5}}$$

Onde: i = o período do fluxo de caixa da transação do cliente;
 R_i = receita do cliente no período i ;
 C_i = custo total de geração da receita R_i no período i ;
 n = o número total de períodos de vida projetada do costume
 d = taxa de desconto/custo de capital

Neste modelo, assume-se que todos os fluxos de caixa ocorrem no final de um período de tempo. Outros modelos de CLV incluíram dados de forma contemplar melhor as perspectivas dos gastos com aquisição e retenção de clientes. Nesse sentido, as variáveis preponderantemente necessárias para proceder com o cálculo do CLV, conforme Braun e Schweidel (2011) a seguir, bem como sua equação proposta:

$$CLVi = \sum_{j=T+1}^{T+N} \frac{p(Buyij = 1)XMC}{(1 + r)^{j-T}} - \frac{CMij}{(1 + r)^{j-T}}$$

- a) custo de aquisição: gastos com marketing e investimentos de atração para adquirir clientes dividido pela quantidade de clientes novos;
- b) índice de *churn*: percentual de clientes que encerram seu relacionamento;
- c) taxa de desconto: custo de capital;
- d) gastos com retenção: montante gasto em determinado período para manter um cliente existente;
- e) tempo;
- f) margem de contribuição (Receita x Margem de lucro).

Para Payne (2006), o papel e a importância relativa da aquisição de clientes varia consideravelmente de acordo com a situação específica da empresa. Obviamente que um novo participante no mercado estará focado principalmente na aquisição de clientes, enquanto uma empresa estabelecida estará mais preocupada com a retenção de clientes. O processo de aquisição de clientes geralmente tem seu enfoque em problemas como:

- a) adquirir clientes a um custo menor
- b) adquirir mais clientes
- c) adquirir clientes mais atraentes
- d) adquirir clientes utilizando novos canais.

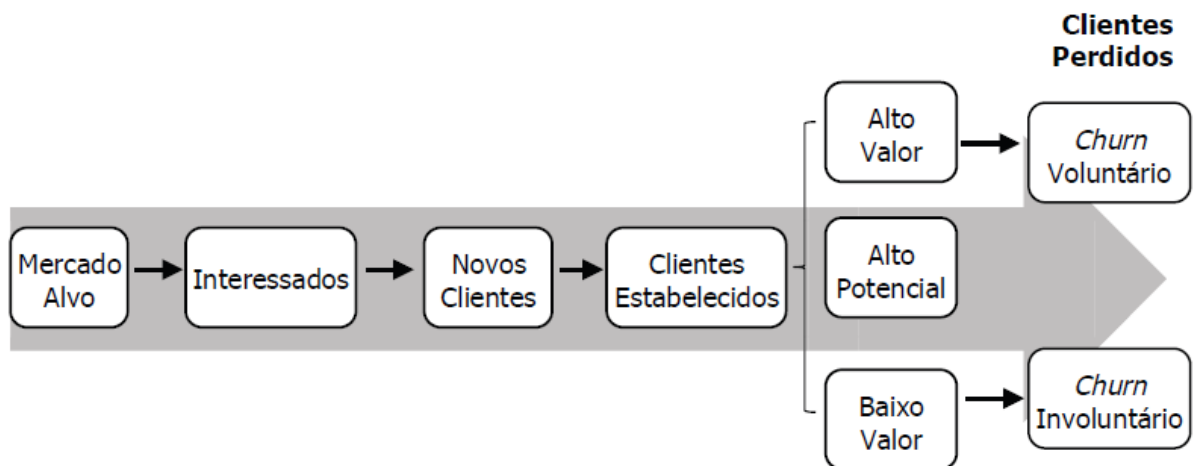
O ponto de partida para entender o valor do cliente na perspectiva da organização é determinar os custos de aquisição de clientes existentes nos principais canais usados pela empresa e identificar como esses custos variam em diferentes segmentos de clientes. Payne (2006) pondera que a análise do CLV de determinado segmento de cliente, deve contemplar as despesas com aquisição exatamente dispensada a este segmento específico.

Payne (2006) apresenta resultados de diversos estudos apontando que muitas empresas se concentram mais em aquisição de novos clientes do que no desenvolvimento de sua base de clientes existente. Em uma pesquisa no setor financeiro do Reino Unido, analisando o valor investido em marketing em diversas empresas, mostrou que os gastos com aquisição de clientes (48%) foram mais que o dobro do gasto com retenção de clientes (22%). Em outra pesquisa sobre as práticas de marketing contemplando 225 organizações do Reino Unido, de diversos setores da indústria, os resultados foram semelhantes - 41% gastos em aquisição de clientes, e apenas 23% foram gastos em retenção de clientes. A grande maioria das empresas

pesquisadas nesta amostra estava em indústrias maduras. Evidencia-se a baixa compreensão sobre o valor da retenção de clientes.

A relação cliente-empresa abrange diferentes estágios e é importante a compreensão destas etapas, pondera Veloso (2013), conforme figura 1. Clientes são entidades dinâmicas e ao longo da relação com a empresa passam por várias fases, podendo o seu comportamento variar entre cada uma delas.

Figura 1 - Estágios da relação cliente-empresa



Fonte: Veloso (2013), adaptado de Berry e Linoff (2011).

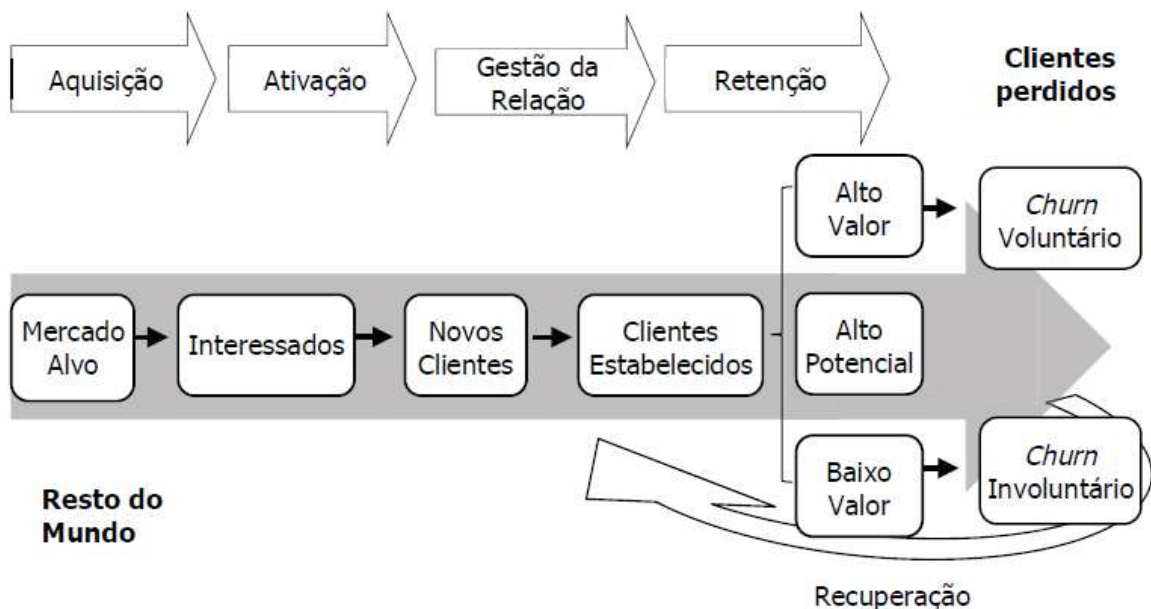
Veloso (2013) destaca os quatro estágios fundamentais do ciclo de vida de um cliente:

- a) potenciais: estão no mercado onde a empresa atua, mas não são clientes;
- b) interessados: estão no mercado e demonstram interesse na empresa;
- c) ativos: estão atualmente a adquirir serviços ou produtos ou manter. Podem ser divididos em dois subgrupos:
 - novos clientes: aqueles que recentemente adquiriram pela primeira vez um produto ou serviço;
 - estabelecidos: relacionamento contínuo, os que regressam e voltam a adquirir produtos ou serviços e espera-se que a relação se vá aprofundando.
- d) perdidos: aqueles que já foram clientes, mas que por algum motivo deixaram de ser.

Para cada estágio de relacionamento há uma estratégia mais adequada para gestão do relacionamento por parte da empresa. Veloso (2013) define processos de negócios, conforme abaixo e demonstrado na Figura 2.

- a) **Aquisição:** processo de captar clientes potenciais. Feito através de campanhas de publicidade, boca-a-boca ou marketing direcionado;
- b) **ativação:** processo de oficializar como clientes as pessoas que demonstraram interesse em adquirir produtos ou serviços;
- c) **gestão (desenvolvimento) da relação:** processo que busca potencializar o valor dos clientes para a empresa;
- d) **retenção:** processo com vistas a reter clientes estabelecidos. Processo crítico para a maioria das empresas;
- e) **recuperação:** tentativa de recuperar bons clientes que abandonaram recentemente a empresa para a concorrência, normalmente contemplando a oferta de incentivos.

Figura 2 - Processos de negócios



Fonte: Veloso (2013), adaptado de Berry e Linoff (2011)

Em virtude do relevante impacto que a evasão de clientes representa para as organizações, esta seção aprofundará a temática do *churn* em continuidade a presente pesquisa.

2.3 CHURN

"*Churn*" refere-se ao ato de clientes que têm a intenção de sair e se tornar cliente de uma empresa concorrente (GHORBANI; TAGHIYAREH, 2009). Há uma preocupação crescente e emergente pelas empresas com a perda de clientes, que é o que caracteriza o *churn* (VELOSO, 2013). Fato é que a evasão, perda ou abandono de um cliente pode representar significativo impacto no nível de receitas das organizações. Maior será este impacto, portanto ainda mais relevante, se considerarmos as abordagens do CLV e métricas centradas no cliente.

Nas últimas duas décadas, profissionais e acadêmicos prestaram considerável atenção ao *churn* e retenção de clientes, seus antecedentes e consequências, principalmente devido ao impacto da retenção no valor da vida útil do cliente e, conseqüentemente, no resultado final da empresa (NITZAN; LIBAI, 2011). Por causa das implicações a uma melhor compreensão e previsão do *churn*, Nitzan e Libai (2011) sugerem que as organizações necessitam de um CRM efetivo e sistêmico.

Conforme Glady et al. (2009), o *churn* configura-se com aquele cliente em que o CLV é decrescente ao longo do tempo. Caracterizado pela evasão ou abandono, embora as razões para a ocorrência do *churn* sejam tão diversas, complexas e específicas, entender por quê isso acontece e quando será mais provável que isso venha a acontecer pode permitir que a atuação proativa da empresa consiga reduzir significativamente a taxa de evasão.

Para Glady, Baesens e Croux (2009), o *churn* é um termo relacionado ao marketing que caracteriza um consumidor que está indo de uma empresa para outra. Como cliente, ele ainda tem um relacionamento com a empresa focal, mas irá para o concorrente no futuro próximo. Se a empresa quiser impedi-lo de sair, é necessária uma ação de retenção. A população de interesse é, portanto, os clientes que já foram adquiridos. Naquela análise, os mesmos autores ponderam que abordagens absolutas para identificar *churners* podem ocultar desertores parciais de grande valor para organização, logo consideram mais apropriada uma definição de status de *churn* baseado em mudanças no volume de negócios do cliente. Ponderam que poderá ser tarde demais esperar que o cliente encerre seu relacionamento com a empresa para tomar qualquer medida preventiva.

Enquanto a taxa de *churn* (*churn rate*) é caracterizada pela quantidade de clientes *churners* em relação a toda base de clientes da empresa, a taxa de retenção

é exatamente o inverso, uma vez que consiste na quantidade de clientes que permaneceu cliente em relação a toda a base de clientes. A taxa de *churn*, conforme Jahanzeb e Jabeen (2007), pode ser calculada através da seguinte equação:

$$\text{Taxa de } churn = (C0 + A1 - C1) / C0$$

Onde:

C0, é o número de clientes no início do período;

A1, os novos clientes brutos durante o período.

C1, o número de clientes ao final do período;

Também é possível proceder com o cálculo do *churn* sob a perspectiva do montante de receita que a empresa deixa de arrecadar em virtude do *churn*, isto é, qual o montante de receita mensal dos clientes que evadiram em relação ao montante de receitas de toda base de clientes.

Em artigo publicado na Harvard Business Review, Reichheld e Sasser Jr. (1990) apresentaram fortes evidências da importância da retenção de clientes e redução da ocorrência de *churn*. A depender do ramo de atividade, apenas 5% de redução na taxa de *churn* podem elevar em até 95% os lucros.

Complementando esta visão, Rust e Zaharik (1993) enfatizaram que a taxa de retenção é o componente mais importante para uma empresa elevar sua rentabilidade e consolidar sua participação de mercado. A pesquisa apresentou que o principal impulsionador da taxa de retenção é a satisfação do cliente, que assume um papel central, mostrando a relação desta com a lealdade individual, taxa de retenção agregada, participação de mercado e, também, indicadores de resultado. Com este propósito de elevar lucratividade, Rust e Zaharik (1993) utilizaram modelo que indica quanto e onde deve ser investido para elevar os índices de satisfação do cliente, mas também aonde podem ser cortados gastos que terão mínimo impacto neste indicador, podendo gerar ainda mais lucratividade.

A pesquisa de Reichheld e Sasser (1990), que foi desenvolvida no contexto de uma agência bancária, demonstrou que uma redução de 5% do *churn* é capaz de aumentar os lucros da unidade em até 85%, relativizado pelo tempo. Destaca-se, portanto, a urgência em rapidamente identificar, analisar e definir estratégias com vistas a melhorar os índices de *churn* e retenção. A previsão do *churn*, portanto, é

imprescindível, inclusive, para mensurar o valor agregado da base de clientes, e serve para diagnosticar sua saúde financeira (FADER; HARDIE, 2010).

Van den Poel e Larivière (2004), mostraram que pequenas alterações nos índices de retenção podem resultar em significativos impactos financeiros para as organizações. Em estudo no contexto bancário, apresentaram o impacto que a elevação de apenas 1% no índice de retenção representa nos resultados ao longo do tempo, conforme Tabela 1 a seguir.

Tabela 1 - Impacto da retenção sobre a base de clientes

Taxa de Retenção	Número de clientes								Quantidade de clientes adicional em relação ao índice de retenção de 93%
	Início	2º ano	3º ano	4º ano	5º ano	10º ano	15º ano	25º ano	
93%	1.000.000	930.000	864.900	804.357	748.052	520.411	362.044	175.223	-
94%	1.000.000	940.000	883.600	830.584	780.749	572.995	420.523	226.500	51.277
Δ (94%/93%)		1,08%	2,16%	3,26%	4,37%	10,10%	16,15%	29,26%	
95%	1.000.000	950.000	902.500	857.375	814.506	630.249	487.675	291.989	116.766
Δ (95%/93%)		2,15%	4,35%	6,59%	8,88%	21,11%	34,70%	66,64%	
96%	1.000.000	960.000	921.600	884.736	849.347	692.534	564.673	375.413	200.190
Δ (96%/93%)		3,23%	6,56%	9,99%	13,54%	33,07%	55,97%	114,25%	

Fonte: Adaptado de Van den Poel e Larivière (2004).

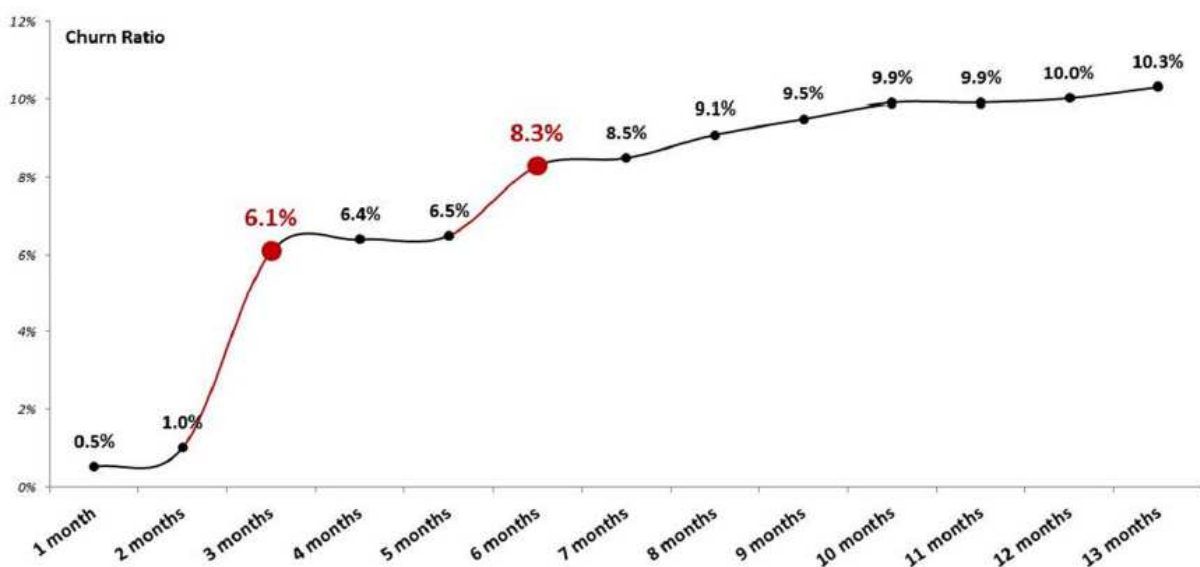
O setor bancário de varejo caracteriza-se por clientes que permanecem com uma empresa por muito tempo (MUTANEN; NOUSIAINEN; AHOLA, 2010). Geralmente os clientes concentram suas operações financeiras em uma instituição e não a abandonam frequentemente. Isso produz um ambiente estável para o gerenciamento do relacionamento com o cliente, sob a ótica da empresa. Embora essa perspectiva contínua do relacionamento com os clientes, a perda de receita por eventual evasão/abandono de clientes pode ser enorme.

Clientes que estão há mais tempo com uma organização tem menos chance de evasão, e maior é a probabilidade de eles permanecerem. Os clientes, com o passar do tempo, tendem a gastar mais e clientes de longo prazo também são menos sensíveis a preço. Cheney (2016) demonstra que, com o passar do tempo, o

relacionamento se fortalece. Ainda, destaca o maior desafio das organizações em relação a evasão dos clientes de forma a elevar os índices de retenção: o tempo. A necessidade de superar a retenção reativa e assumir uma estratégia de retenção proativa. Em suma, envidar ações efetivas para reter os clientes antes que eles tomem a decisão de evadir e identificar os clientes com alta probabilidade da evasão.

Depren (2018) destaca outro ponto relevante no ciclo de relacionamento com o cliente quando, em decorrência de algum atrito ou insatisfação, o cliente manifesta reclamação perante os canais de ouvidoria da organização. Nesta pesquisa quantitativa, com 1000 clientes de um banco situado na Turquia, foi abordada a relação financeira futura e a satisfação geral dos clientes sobre a condução do processo de reclamação, através dos índices de *churn*. Conforme figura 3, destaca-se a expressiva evolução do nível de *churn* entre 2 e 3 meses após a reclamação, evidenciando uma janela de tempo passível de desenvolver ações para melhorar a satisfação destes clientes previamente a consumação da evasão. Ou seja, torna-se necessário que a instituição envide ações para manter seu cliente nos primeiros 2 meses.

Figura 3 - Taxa de *churn* em clientes com ocorrência canais de Ouvidoria



Fonte: Depren (2010).

Ainda mais nestes momentos, em que a perspectiva de evasão é eminente, investir em consistentes ações de retenção é fundamental. Kurtz e Clow (1998) apresentaram casos em que o custo de retenção é até cerca de 5 vezes inferior ao

custo de aquisição de um cliente. Para redução dos índices de *churn*, além do papel relevante de ações de pós-venda, destaca-se a necessidade de um efetivo sistema de CRM. Assim, impõe-se a natural conexão às diretrizes do CLV.

As ferramentas de CRM precisam estar comprometidas com às estratégias de retenção, de forma a prestar direcionamentos de atuação e suportar a implementação dessas estratégias. Mais do que simplesmente listar potenciais clientes, a eficiência do CRM combinado com as estratégias de retenção precisa contemplar a mensuração da redução efetiva dos índices de *churn*. Diante disso, este estudo consiste em uma pesquisa operacional para identificar clientes com alto risco de *churn*.

O fenômeno do *churn* está diretamente associado ao tempo em que o cliente mantém esta posição (de cliente) perante uma organização. Por aderência aos objetivos, algumas conceituações serão adotadas dos parâmetros definidos em Veloso (2013), conforme a seguir:

O termo *churn* é definido neste trabalho como “a quebra da relação entre uma empresa e um cliente”. Do mesmo modo, se denomina *churner* o cliente para o qual ocorreu a evento de *churn* e *churning* o ato de um cliente incorrer em *churn*. (...) A taxa de *churn* é um dado estatístico sobre a quantidade de *churn* num determinado período de tempo (PAGE et al., 1996). Por exemplo, se não se levar em consideração os clientes adquiridos, numa empresa que tem 100 clientes, dos quais, num período de 1 ano, 10 abandonam a empresa, a taxa de *churn* anual é na ordem dos 10%. (VELOSO, 2013).

Ghorbani e Taghiyareh (2009) salientam a importância do gerenciamento do *churn*, que consiste na prévia identificação dos *churners* e a implementação de ações proativas para mantê-los na condição de clientes. Ponderam ainda que o primeiro passo para combater o fenômeno do *churn* consiste em investigar as suas causas. Os fatores causadores do *churn* podem ter diversas origens, para Veloso (2013), dentre outros, destacam-se:

- a) serviço prestado ao cliente;
- b) qualidade;
- c) preço;
- d) funcionalidade;
- e) conveniência.

Veloso (2013) destaca a necessidade de classificar os tipos de *churn*. Pondera que é perigoso tratar da mesma forma clientes que a empresa quer manter e clientes

que não são importantes para a empresa. Imagine deliberar uma campanha para recuperar clientes evadidos, e em meio aos beneficiados constarem clientes porventura inadimplentes. Para Hadden et al. (2007), seguem os tipos de *churn*:

- a) *Churn* involuntário: Cliente tem papel passivo. O término da relação se dá por ação ou posição da empresa, seja por término ou não cumprimento de contrato ou por inadimplência de cliente. Sob a perspectiva da gestão para preservação do relacionamento, não se evidencia interesse das empresas em identificar previamente este tipo de *churn*;
- b) *Churn* voluntário: Mais difícil identificar, pois a decisão de quebra do contrato parte do cliente. Normalmente ocorre por motivações financeiras, quando concorrentes oferecem condições mais atrativas, ou ainda relacionadas a fatores como qualidade, mau atendimento, disponibilidade entre outros. A preocupação das empresas está concentrada neste tipo de *churn*;
- c) *Churn* parcial: Comum no mercado bancário. Quando o cliente transfere parte de seus negócios/operações para um concorrente, configurando a desidratação em relação ao volume de negócios que mantinha. Mais perceptível, este tipo de *churn* pode ser sinalizador de um *churn* total.

Burez e Van Den Poel (2008) conceituaram mais um tipo de *churn*, sobre uma perspectiva importante ao mercado bancário – *churn* financeiro, cuja razão da evasão é decorrente da incapacidade financeira. Para este caso, as abordagens estão voltadas sob a perspectiva da política de crédito e avaliação de risco de cada instituição.

Nesse sentido, um processo de gerenciamento de *churn* envolve a construção de um modelo de previsão. A ampla maioria de técnicas estão baseadas no histórico de ocorrências de *churn* passadas e a identificação de variáveis-chave, que têm maior influência no processo do *churn*. Estes modelos preditores do *churn* são então usados para identificar e classificar em nível individual de cliente o risco potencialmente alto de evasão e propicia a adoção de ações de retenção apropriadas (LEE et al., 2012).

Burez e Van Den Poel (2007) apresentaram importantes achados em uma base de clientes de uma Operadora de TV por assinatura. Após a aplicação do modelo preditivo, dentre o grupo de clientes com provável ocorrência do *churn*, foram

realizadas 3 ações distintas para testar o impacto de cada ação sobre o índice de retenção. Além do grupo controle (Grupo 0), em que não foi realizada nenhuma ação, houve um grupo que foi convidado a responder um questionário de satisfação (Grupo A), outro grupo recebeu pares de ingressos de cinema (Grupo B) e outro grupo ganhou convites para um evento exclusivo (Grupo C). O trabalho evidenciou que os três tipos de ações reduzem significativamente a taxa de *churn* em relação ao grupo de controle. Enquanto o grupo controle teve 15,67% de taxa de *churn*, o grupo A teve 10,94% (4,72% a menos do que o Grupo Controle, reduzindo a quantidade de *churners* em 30,18%). Dar tickets ou convidar clientes para um evento exclusivo também reduz o *churn*, mas em menor grau, 12,12% e 12,87%, o que ainda representa redução na quantidade de *churners* em 22% e 17%, respectivamente. Essa evidência demonstra que os clientes foram mais sensíveis à sinalização, por parte da empresa, de preocupação por sua satisfação do que receber benefícios.

Hadden et al. (2007) salientam que os esforços de gerenciamento do *churn* não devem se concentrar em toda a base de clientes, notadamente porque não são todos os clientes que vale a pena manter, e ainda porque reter clientes custa dinheiro; Logo, dispôr recursos a clientes que não têm intenção de evadir representa unicamente ônus. Neslin et al. (2006) consideram que, normalmente, as ações de retenção envolvem investimento financeiro, que, para se justificar, deve gerar retorno. Eles desenvolveram um modelo para dimensionar o retorno real da implementação de campanhas de retenção, da seguinte forma:

$$\Pi = N\alpha[\beta\gamma(CLV - c - \delta) + \beta(1-\gamma)(-c) + (1 - \beta)(-C - \delta)] - A$$

Em que:

N = o número total de clientes;

α = a fração de clientes incluídos no programa de gerenciamento de *churn*;

β = a fração de clientes “ α ” abrangidos que são prováveis *churn*;

δ = o custo dos incentivos/benefícios para reter os clientes;

γ = a fração de *churners* potenciais que decidem permanecer por causa do incentivo (ou seja, a taxa de sucesso do incentivo);

c = o custo de entrar em contato com um cliente para oferecer-lhe o incentivo;

CLV = valor da vida útil do cliente (ou o valor para a empresa se o cliente for retido);

A = os custos fixos da execução do programa de gerenciamento de *churn*.

Esta equação expressa inicialmente $\beta\gamma(CLV - c - \delta)$ a contribuição do lucro da fração dos clientes contatados que iriam evadir e decidiram permanecer pelos benefícios recebidos, a fração de clientes cujo programa não obteve efeito $\beta(1-\gamma)$, além do custo por benefícios entregues a clientes que aceitaram recebê-los mas não iriam evadir $(1 - \beta)$. O termo β reflete a precisão do modelo e pode ser expresso da seguinte forma:

$$\beta = \lambda\beta_0$$

Em que β_0 é a fração de clientes *churn* dentre o total de clientes da empresa e λ é o lift (refere-se às ocorrências de *churn* dos clientes abrangidos pelo programa de retenção, comparativamente aos demais).

Kumar et al. (2016) ressalta que satisfação não significa lealdade nem retenção, logo configura-se a complexa necessidade de identificar outros marcadores para prever a ocorrência do *churn*. Construir um modelo eficaz de predição do *churn*, através das diversas técnicas dispostas atualmente, tornou-se uma questão de foco para empresas e de pesquisa nos últimos anos, o que será abordado na continuação deste referencial teórico.

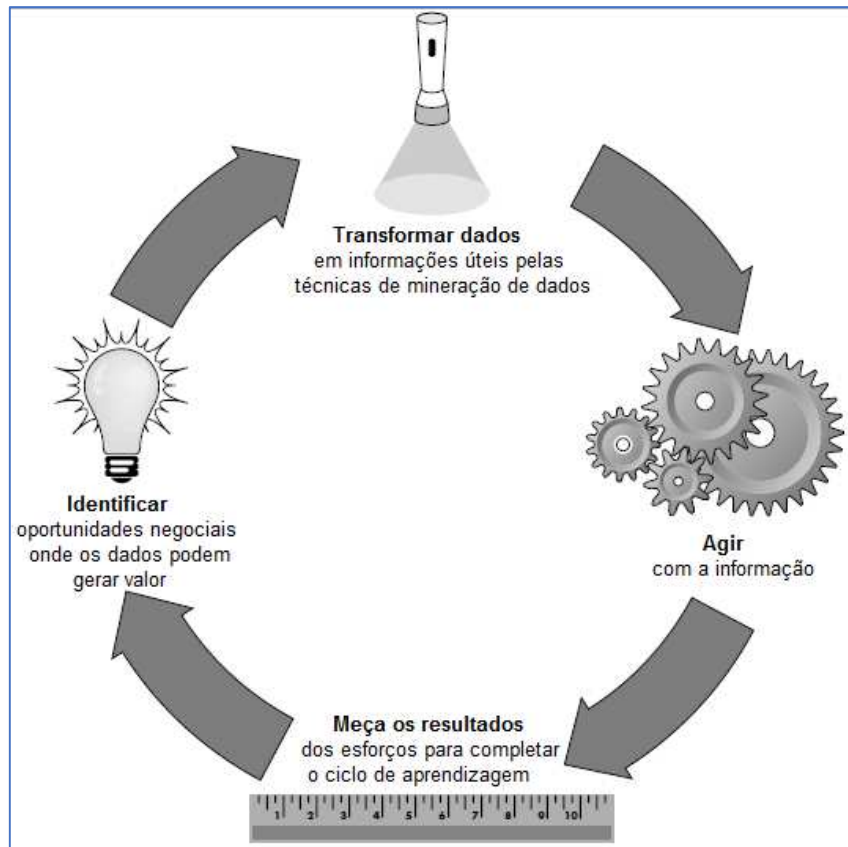
2.3.1 Abordagens preditivas aplicáveis ao fenômeno do *churn*

Atualmente, a maioria das estratégias de gerenciamento de *churn* é baseada em data warehousing (banco de dados), mantendo e gerenciando dados históricos e muito detalhados obtidos de muitos dispositivos de computação e canais, bem como técnicas de mineração de dados que enriquecem e transformam dados em informações significativas para tomadores de decisões (ADEBIYI, 2016).

Barry e Linoff (2011) conceituam que *Data Mining* - mineração de dados - é um processo de negócios para explorar grandes quantidades de dados para descobrir padrões e regras significativos. Estes autores ponderam que, por muitos anos, esse mesmo conceito foi aplicado a outros nomes como *Business Intelligence*, *Knowledge Discovery from Database – KDD*, *Predictive Analytics*, entre outros, todos com a mesma finalidade. Ainda, salientam que as empresas estão se movendo em direção à meta de compreender cada cliente individualmente e usar essa compreensão para tornar a relação com seus clientes melhor, com menos atritos, mais proativa, e principalmente, mais rentável.

A mineração de dados compreende ferramentas e técnicas que oferecerão suporte às empresas centradas no cliente, e de forma integrada aos sistemas de CRM, eleva o aprendizado e conhecimento sobre o cliente. Barry e Linoff (2011) salientam as etapas do ciclo aprendizagem em mineração de dados conforme Figura 4, a seguir:

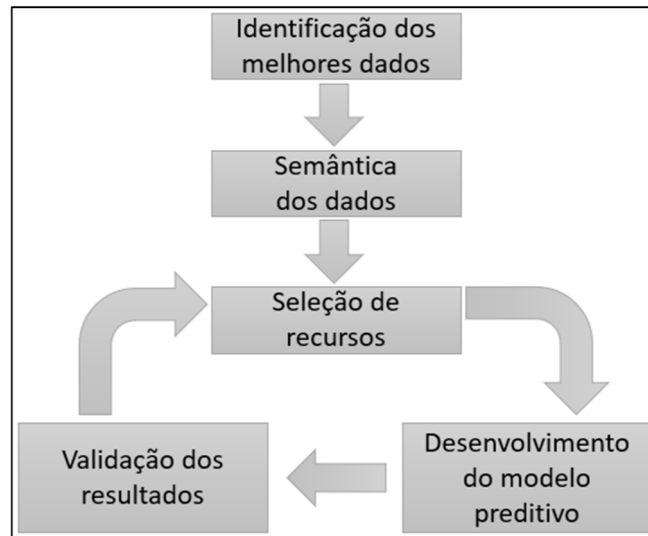
Figura 4 - Ciclo de aprendizagem em Data Mining



Fonte: Adaptado de Barry e Linoff (2011).

As técnicas preditivas de mineração de dados usam dados do passado para prever o que provavelmente acontecerá no futuro. Berry e Linoff (2011) ponderam que a construção de um modelo preditivo requer separação no tempo entre as entradas ou os preditores do modelo e a saída do modelo, a coisa a ser prevista. Se esta separação não for mantida, o modelo não funcionará.

Datta et al. (2001) apresentaram um modelo de cinco estágios para desenvolvimento de um framework preditivo do *churn* conforme Figura 5 e ponderações a seguir:

Figura 5 - Modelo de framework preditivo de *churn*

Fonte: Datta et al. (2001).

- a) Identificação dos melhores dados: Diferentes combinações de dados contêm diferentes poderes analíticos. É necessário identificar os dados que melhor se adaptam ao tipo de análise que está sendo executada e a finalidade pretendida;
- b) estabelecimento de semântica de dados: Semântica de dados é o processo de entender o contexto dos dados em um banco de dados. Os campos e dados armazenados em um banco de dados geralmente podem ser visualizados como uma coleção de palavras. Essas palavras podem parecer autoexplicativas e não ambíguas, no entanto, é comum que os dados sejam difíceis de interpretar;
- c) seleção de recursos: Processo de identificação dos campos que são os melhores para previsão. Etapa importante porque auxilia na ênfase de dados e exclusão de variáveis redundantes. Compreende a fase da pesquisa de seleção podendo ser dos tipos ótima, heurística ou aleatória e a fase de avaliação, com as categorias filtro e wrapper;
- d) desenvolvimento do modelo preditivo: Um modelo preditivo é definido como aquele que pega padrões que foram descobertos no banco de dados e prevê o futuro. As técnicas mais importantes e populares de modelagem preditiva são Regressão Logística, Árvores de Decisão, Redes Neurais, *Support Vector Machine* e Random Forests, cada uma com suas peculiaridades, vantagens, entregas, limitações e restrições;

- e) validação dos resultados: São vários os métodos documentados para validar um modelo de perda de clientes. Notadamente, destaca-se a Validação Cross-fold (70/30 ou 75/25), em que a amostra divide-se entre conjunto de treinamento e conjunto de validação.

A fim de compreender como diferentes estudos conceberam seus modelos de previsão do *churn*, segue Quadro 4, contemplando a síntese teórica, contexto e técnicas de análise preditiva constantes deste referencial teórico e empírico.

Quadro 4 - Síntese teórica de estudos/modelos preditivos do *churn*

Autores	Ano	Objeto de pesquisa e contexto	Técnicas
Eiben A.E., Koudijs A.E., Slisser F.	1998	Comparação de técnicas de modelagem, base de 500 mil clientes de instituição financeira	Regressão logística, programação genética, análise de dados aproximados, CHAID
Madden G., Savage S.J., Coble-Neal G.	1999	Desenvolvimento de um modelo de predição de <i>churn</i> e análise de drivers de <i>churn</i> em uma base de 200 mil assinantes de um provedor de internet	Modelo binomial probit
Datta P., Masand B., Mani D.R., LiB.	2000	Descrição e aplicação do sistema automatizado (CHAMP) de modelagem de previsão de <i>churn</i> de usuários de operadora de telefonia celular - base de 5 milhões de clientes	Redes neurais e árvores de decisão
Mozer M.C., Wolniewicz R., Grimes D.B., Johnson E., Kaushansky H.	2000	Predição da probabilidade de <i>churn</i> , otimização de incentivos de retenção e estimativa de ganhos de lucro sobre uma base de 47 mil clientes de empresa telecom.	Regressão logística, árvore de decisão, redes neurais
Wei C.P., Chiu I.T.	2002	Aplicação do algoritmo C4.5 para criar um modelo de previsão de <i>churn</i> , sobre uma base de 114 mil clientes de empresa Telecom	Árvores de decisão (C4.5)
Au W.H., Chan K.C.C., Yao X.	2003	Aplicação de uma nova técnica de mineração de dados para prever probabilidades de <i>churn</i> , base de 100 mil clientes de empresa Telecom	Árvores de decisão (C4.5), rede neural, mineração de dados por aprendizado evolutivo
Hwang H., Jung T., Suh E.	2004	Modelo de preditivo de <i>churn</i> agregado a um modelo de CLV. Base de 16384 clientes de empresa telecom	Regressão logística, árvores de decisão, redes neurais

Autores	Ano	Objeto de pesquisa e contexto	Técnicas
Buckinx W., Van den Poel D.	2005	Comparação de técnicas para predição de deserção parcial, foco em clientes lucrativos em um ambiente não contratual. Base de 158.884 clientes de varejista do setor de bens de consumo.	Regressão logística, redes neurais, florestas aleatórias
Larivière B., Van den Poel D.	2005	Investigação de variáveis explicativas e métodos de modelagem para previsão de <i>churn</i> em uma base de 100.000 clientes de uma instituição financeira.	Regressão logística, regressão linear, florestas aleatórias, florestas de regressão
Lemmens A., Croux C.	2006	Aplicação de técnicas de bagging e boosting para melhorar o poder preditivo dos modelos de previsão de rotatividade. Base de 100.462 clientes de uma empresa Telecom.	Regressão logística, árvores de decisão com <i>bagging</i> e <i>boosting</i> .
Neslin S.A., Gupta S., Kamakura W., Lu J., Mason C.H.	2006	Análise dos resultados de um torneio de modelagem de predição de <i>churn</i> . Base de dados com 51.306 clientes.	Regressão logística, árvore de decisão, redes neurais, análise discriminante, análise de cluster e Bayes
Hung S.Y., Yen D.C., Wang H.Y.	2006	Estudo comparativo e aplicação de métodos de modelagem de predição de <i>churn</i>	Árvore de decisão, redes neurais (em segmentos agrupados)
Burez J., Van den Poel D.	2007	Desenvolvimento de modelo de previsão de <i>churn</i> com 24.185 clientes, e realizada ações de retenção.	Regressão logística (com cadeias de Markov), florestas aleatórias.
Burez J., Van den Poel D.	2008	Estudo sobre métodos de amostragem, métricas e métodos de avaliação e técnicas de modelagem e controle. Teve por contexto 2 bancos, 1 telecom, 1 Jornal, 1 TV por assinatura e 1 supermercado.	Regressão logística e florestas aleatórias
Kumar D.A., Ravi V.	2008	Estudo extensivo para comparar os resultados de diferentes técnicas de amostragem e modelagem para prever a evasão de clientes de cartões de crédito. Base de 14.814 clientes bancários.	Regressão logística, arvores de decisão, redes neurais, Máquina de Vetores de Suporte (SVM)
Lima E., Mues C., Baesens B.	2008	Incorporação de conhecimento de domínio em modelos de previsão de <i>churn</i> . Base de 5.000 e 15.000 clientes de 2 empresas do ramo de telecomunicação	Regressão logística e árvores de decisão

Autores	Ano	Objeto de pesquisa e contexto	Técnicas
Coussement K., Van den Poel D.	2008	Aplicação da máquina de vetores de suporte na previsão de rotatividade em uma assinatura de jornal	Regressão logística, máquina de vetores de suporte, florestas aleatórias
Botelho, D; Tostes, F. D.	2010	Modelou a probabilidade de clientes abandonarem o relacionamento. Base de 100.000 clientes de operadora de cartão de crédito	Regressão logística
Farquad, M; Ravi, V; Raju, SB	2014	Modelou a predição de ocorrência do <i>churn</i> em uma base de 14.814 clientes de operadora de cartão de crédito.	Maquina de Vetores de Suporte (SVM)
Thomas L, Tang L, Fletcher M, Marshall A, Thomas L, Pan J.	2014	Uso de informações de comportamento para aprimorar a modelagem de ocorrências de <i>churn</i> . Base de 19774 clientes de empresa de serviços financeiros.	<i>Probit - hazard model</i>
Gür Ali, Özden, and Umut Arıtürk	2014	Configuraram modelo de previsão de <i>churn</i> dinâmico. Testaram empiricamente em uma base de 7204 clientes <i>private bank</i> .	Regressão logística e análise de sobrevivência
Zorić, Alisa Bilal	2016	Estudo de caso com foco na identificação de clientes com risco de evasão e se valem a pena reter. Base de 1866 clientes bancários.	Redes neurais

Fonte: Elaborado pelo autor (2019).

Todos estes estudos constantes da quadro 4 têm em comum o propósito de entregar informações gerenciais, que, por seu caráter preditivo, antecipadamente podem revelar aqueles clientes cuja evasão pode ser considerada uma tendência. Evidenciam-se variadas técnicas de mineração de dados para realizar este propósito, sendo as mais referenciadas a Regressão Logística, Random Forests e as Redes Neurais, as quais serão melhor exploradas no capítulo de método de pesquisa, uma vez que estas metodologias serão desenvolvidas na presente pesquisa. O capítulo de método, a seguir, vai explorar as demais etapas de condução desta pesquisa.

3 MÉTODO DE PESQUISA

Para o desenvolvimento desta pesquisa, fez-se necessário a adoção de metodologia adequada para alcançar os objetivos propostos. Este capítulo, portanto, tem por objetivo esclarecer os procedimentos, definições e escolhas adotados para sua realização, em consonância aos preceitos do processo formal e sistemático do método científico e do processo de pesquisa, apresentando seu delineamento e técnicas utilizadas.

3.1 DELINEAMENTO DA PESQUISA

Em relação às definições de método e metodologia, Sampieri (2013) descreve que o método compreende no caminho ou forma para se alcançar determinado objetivo, enquanto a metodologia consiste nos procedimentos e regras utilizadas por determinado método. Para Sampieri (2013), “as premissas de ambos os paradigmas podem conviver ou se entrelaçar e combinar com teorias substantivas; portanto, integrar os métodos quantitativos e qualitativos não só é possível, como é conveniente”. Nesse sentido, esta pesquisa esteve pautada em procedimentos de natureza essencialmente quantitativa, à exceção de uma etapa inicial de caráter qualitativo, caracterizada por entrevistas informais com profissionais da instituição financeira para alinhamento e definição de atributos de pesquisa, notadamente em relação a caracterização do *churn* no contexto bancário e o mapeamento de variáveis explicativas para as análises preditivas.

Quanto à classificação das pesquisas, aderente ao exposto por Vergara (2013), apresentam-se de diferentes maneiras, cada qual tendo suas especificidades, classificando-se quanto aos meios e aos fins. Em relação ao meio, a pesquisa pode ser classificada como *ex post fact*, uma vez que os dados que serão utilizados são secundários, extraídos dos sistemas internos do Banco, não sendo possível manipular ou controlá-los. Em relação a finalidade, conforme Vergara (2013), esta pesquisa classifica-se como descritiva e aplicada. Aderente a Botelho e Tostes (2010), a pesquisa empírica é descritiva, com uso de dados secundários coletados longitudinalmente. Através do cumprimento do objetivo geral, que compreende comparar modelos preditivos, configura-se o carácter descritivo desta pesquisa. Classifica-se por aplicada em virtude de ser passível de aplicações práticas, uma vez

que, a partir dos achados desta pesquisa, poderá ser desenvolvida campanha de retenção juntos aos clientes previstos a desidratar, podendo ainda subsidiar a adequação das estratégias institucionais de retenção, bem como propiciar a elevação sustentável da rentabilidade da base de clientes. Também se observa uma perspectiva exploratória nesta pesquisa, ponderado não se parte de um modelo teórico a ser testado, pelo contrário, utilizam-se conjuntos de variáveis que são candidatos a explicar o fenômeno analisado.

3.2 ETAPA QUALITATIVA

De forma a extrair diferentes percepções e com vistas a permitir ao pesquisador a obtenção de elementos com visões complementares, a etapa qualitativa foi imprescindível para importantes deliberações da pesquisa. Para Gil (2008), entrevista é uma técnica em que o investigador se apresenta frente ao investigado e lhe formula perguntas, com o objetivo de obtenção dos dados que interessam à investigação. Dentre os tipos de entrevista, Gil (2008) salienta que a entrevista informal deve ter a mínima estrutura possível, distingue-se da simples conversação pelo objetivo de coleta de dados, informações ou percepções. Além disso, ressalta que esta é recomendada para estudos exploratórios, que neste estudo está configurado pelo emergente tema abordado.

Portanto, a efetiva definição das variáveis a serem analisadas foi precedida de entrevistas informais com gestores e funcionários do Banco do Brasil, atuantes nos pilares Estratégico (Diretorias), Tático (Superintendências) e Negocial (Agências e Escritórios), o que agregou complementaridade nas percepções sobre os fatores preponderantes correlacionados à evasão de clientes. Todos os profissionais entrevistados possuem ampla experiência no segmento bancário e notoriamente demonstraram compreender o fenômeno do *churn* e da desidratação (*churn* parcial). Buscou-se elementos que poderiam trazer maior clareza acerca da definição e escolha das variáveis, bem como ponderações em relação a disponibilidade de dados. O quadro 5 apresenta, de forma consolidada, a função, dependência, pilar/segmento de atuação, tempo de banco e formação.

Quadro 5 - Caracterização dos entrevistados na etapa qualitativa

Função	Dependência	Pilar/Segmento	Tempo de banco (anos)	Formação
Gerente de Soluções	Diretoria de Clientes	Estratégico	26	Doutorado
Superintendente Comercial	Superintendencia Varejo e Governo RS	Tático	31	Pós-graduação
Gerente de Negócios	Superintendencia Varejo e Governo RS	Tático	18	Graduação
Gerente Geral de Agência	Escritório Exclusivo Porto Alegre	Negocial PF	33	Mestrado
Gerente Geral de Agência	Agência Empresa Porto Alegre	Negocial PJ	17	Mestrado
Gerente Geral de Agência	Agência Estilo Nilo Peçanha	Negocial Alta Renda	19	Pós-graduação
Gerente Geral de Agência	Escritório Estilo Digital Porto Alegre	Negocial Alta Renda	18	Pós-graduação
Gerente Geral de Agência	Agência Canoas	Varejo	17	Pós-graduação
Gerente de Relacionamento	Escritório Private Porto Alegre	Negocial Private	11	Pós-graduação
Gerente de Relacionamento	Agência Estilo Moinhos de Vento	Negocial Alta Renda	12	Pós-graduação
Gerente de Relacionamento	Agência Igara	Negocial PF	6	Pós-graduação
Assistente de Negócios	Agência Igara	Varejo	18	Pós-graduação
Escriturário	Agência Igara	Varejo	19	Pós-graduação

Fonte: Elaborado pelo autor (2019).

A condução desta etapa, de cunho exploratório/descritivo com o objetivo de colher *insights*, foi pautada por questões abertas e não estruturadas. O propósito foi atender aos objetivos “a” e “b” do presente trabalho, no sentido de capturar entendimentos sobre constructos e parâmetros que caracterizem o *churn* no contexto bancário, em nível parcial e de produto (objetivo a) e, ainda, o mapeamento das variáveis que podem ser consideradas preditoras do fenômeno do *churn*, passíveis de coleta e análise (objetivo b).

Ponto crítico para subsequente análise dos dados é o mapeamento das variáveis explicativas, uma vez que as técnicas preditivas indicam suas influências no fenômeno observado do *churn*, estritamente limitadas às variáveis ou conjuntos de variáveis escolhidas (VANDECRUYS et al, 2008). Ressalte-se, neste sentido, a conceituação de Wazlawick (2014), ponderando que “uma variável é um nome que se

dá a um fenômeno que pode ser medido e que varia conforme a medição. Se não variasse, seria uma constante e não teria maior interesse para pesquisa.”

Subramanya e Somani (2017) apresentaram relevante contribuição às metodologias de análise preditiva do *churn* ao demonstrar a melhoria da eficiência preditiva em modelos usando um conjunto de recursos estendido e algoritmos de seleção de recursos. Foi utilizada uma combinação de dados não estruturados e dados estruturados do data warehouse do cliente para extrair vários conjuntos de recursos relacionados a clientes, cujos dados podem ser estáticos, categóricos ou quantitativos, conforme definem:

- a) demográficas: recursos específicos do cliente, indicando a localização geográfica do cliente, renda, setor de atuação;
- b) informações do cliente: esses recursos representam informações de perfil que podem contêm detalhes como nível de negócios, tipo de contrato com o varejista e número de contas de faturamento;
- c) receitas do cliente (Vendas): Esses recursos representam agregados de receita, comportamento de compra, periodicidade, incluindo métricas para as últimas tendências em compras;
- d) consumo de produtos: Esses recursos agregam o padrão de compra do cliente no nível de categoria para cada produto, permitindo, assim, a descoberta de qualquer categoria de produto específica que eleve o potencial de evasão de clientes;
- e) freqüência: Esses recursos analisam vários recursos convencionais e longitudinais relacionados ao comportamento de compra e periodicidade do cliente;
- f) comportamentais : Esses são recursos extraídos da interação do cliente com a organização. Eles contêm recursos que indicam como os clientes passam o tempo na website/app, navegação e visualizações de produto etc.;
- g) experienciais: Recursos que definem métricas de experiência do cliente descrevendo a experiência de compra, caso negativa pode afastar o cliente.

A consolidação das contribuições alcançadas durante a etapa qualitativa adotou a classificação de dados em aderência ao exposto por Subramanya e Somani (2017), conforme quadro 6 a seguir.

Quadro 6 - Proposta de variáveis para análises preditivas

Categoria dos dados	Variáveis	Justificativa
Comportamentais	Quantidade de acessos ao Home banking / App	Variáveis comportamentais permitem compreender perfil comportamental do cliente, cuja evidência em nível isolado pode representar informação importante.
	Quantidade de acessos ao App Mobile	
	Histórico de atendimento presencial	
Consumo de produtos	Seguro patrimonial (Veículo/residencial)	Produtos e serviços com caráter fidelizador, que podem evidenciar configuração de barreiras de troca (custo de mudança).
	Seguro pessoal	
	Cartão de Crédito - Possui adicional	
	Consórcios	
	Serviço de SMS	
	Quantidade de Convênios de Débito Automático	Produtos e serviços com caráter fidelizador, também podem evidenciar barreiras de troca, mas além disso a eventual redução do volume em operações (aplicação / captação) pode ser sintomáticos de um processo de <i>churn</i> .
	Investimentos financeiros (Fundos, CDB, Poupança, LCI/ LCA)	
	Previdência privada (PGBL/ VGBL)	
	Custódia de ações e títulos	
	Cartão de Crédito - Volume de fatura paga	
Crédito - Empréstimos BB	Financiamento imobiliário - BB	
Crédito - Financiamento Veículo BB		
Demográficos	Idade	Variáveis demográficas cuja ponderação em nível individual pode revelar possíveis concentrações de ocorrência do <i>churn</i> . Evidência e validação em nível isolado pode gerar relevantes informações.
	Gênero	
	Rendimentos	
	Escolaridade	
	Atividade profissional	
	Ramo de atividade profissional	
	Estado social	
	UF de residência	
Domicílio pessoal/profissional no município da agência		
Experienciais	Ocorrência de reclamação em canais de ouvidoria	Variáveis importantes para mapear ocorrências de atrito, normalmente associadas ao <i>churn</i> .
	Avaliação de atendimento humano (SMS)	
	Histórico de ocorrência de fraude eletrônica (conta / cartão)	
Frequência	Cartão de Crédito - Quantidade de transações	Variáveis cuja redução da quantidade de transações pode ser sintomática de um processo de <i>churn</i> .
	Quantidade de transações Home banking	
	Quantidade de transações App Mobile	
	Diversificação da cesta de produtos do cliente - evolução +/-	

Categoria dos dados	Variáveis	Justificativa
Informativos	Conta corrente - Saldo médio credor	Variáveis informativas cuja ponderação em nível individual pode revelar possíveis marcadores de ocorrência do <i>churn</i> . Evidencia e validação em nível isolado pode gerar relevantes informações.
	Cheque Especial - Saldo médio devedor	
	Cartão de Crédito - Endividamento SFN	
	Cartão de Crédito - % anuidade paga	
	Crédito - Empréstimos SFN	
	Crédito - Financiamento Veículo SFN	
	Financiamento imobiliário - SFN	
	Recebe proventos	
	Livre opção bancária - LOB VEM	
	Livre opção bancária - LOB VAI	
	TED para conta de mesma titularidade em outra IF	
	Histórico de transação recusada (cartão ou movimentação)	
	Tempo de relacionamento BB	
	Cadastro atualizado	
	Limite de Crédito	
	Risco do Limite de Crédito aceitável (Rating até C)	
	Anotações restritivas (Total fontes externas)	
	Encarteiramento compatível ao segmento do cliente	
	Pacote de Serviços compatível ao segmento do cliente	
	Interação relacional ativa BB dentro do prazo de rotação	
Cadastro - relacionamentos pessoais cliente BB		
Segmentação BB		
Receitas do cliente	Margem de contribuição consolidada	Eventuais alterações podem ser características da ocorrência do <i>churn</i> . Sendo possível, pretende-se estratificar variáveis em nível de produto.
	Margem de contribuição - Produtos de Aplicação	
	Margem de contribuição - Produtos de Captação	
	Margem de contribuição - Produtos de Serviços	
	Pacote de Serviços	

Fonte: Elaborado pelo autor (2019).

3.3 ETAPA QUANTITATIVA

Tendo por norte o objetivo de responder a questão “Como identificar os clientes de uma instituição financeira que tendem ao fenômeno do *churn*?”, foram desenvolvidas metodologias de pesquisa de natureza quantitativa, amparada em técnicas de análise estatísticas e econométricas com capacidade preditiva, consistentes com o propósito desta pesquisa, respaldadas pelo referencial teórico previamente exposto. Seguem alguns fundamentos desta etapa, que contempla informações sobre a base de dados, modelos e métodos preditivos, além das métricas utilizadas para avaliação dos resultados.

3.3.1 Universo e amostra

Notadamente, a indústria bancária é reconhecida pela densidade de dados a que tem acesso, contemplando normalmente, para cada cliente, centenas ou até milhares de variáveis históricas de dados. Não apenas por questões estratégicas para não expôr o padrão de consumo e comportamento dentro suas bases de clientes, mas também decorrente de questões legais como, por exemplo a legislação atinente ao sigilo bancário, as instituições financeiras reservam com muito zelo o acesso aos dados e informações armazenadas em seus sistemas corporativos de bases de dados. No Banco do Brasil não é diferente, sendo que, por regra, todos dados são para uso estritamente interno. A solicitação de dados ao Banco do Brasil, almejava a obtenção de uma amostra abrangente, representativa do universo de clientes da instituição, sendo que os dados solicitados foram pautados pelas diretrizes constantes da etapa qualitativa, tendo sido acolhida e atendida em caráter de absoluta excepcionalidade, uma vez que o autor é funcionário da Instituição e seu vínculo com este Programa de Pós-graduação é patrocinado pela instituição, devendo as pesquisas serem desenvolvidas em seu contexto.

O universo da pesquisa é representado por toda a base de clientes do Banco do Brasil S.A., notadamente expressando um montante de mais de 65 milhões de clientes, entre clientes Pessoa Jurídica e Pessoa Física, porém a amostra foi selecionada somente dentre a população de clientes Pessoa Física.

A base efetivamente fornecida pela Instituição, em resposta à solicitação apresentada pelo pesquisador, abrange 2.938 clientes. Os dados sobre esses clientes

contemplam uma posição mensal histórica para cada uma das 118 variáveis capturadas, sobre até 17 meses anteriores a Fevereiro de 2019. Como nem todos os clientes constavam desde o início, os trabalhos foram desenvolvidos considerando apenas os clientes para os quais foi possível observar o comportamento ao longo de 12 meses, o que delimitou a amostra efetiva em 2343 clientes. Um ponto a ser destacado refere-se às características altamente heterogêneas da amostra, contemplando diversos perfis sóciográficos de clientes, dentre os quais foram observadas, por exemplo, 118 distintas atividades profissionais dentre clientes com idades entre 23 a até 95 anos, o que pode elevar o nível de complexidade para as análises econométricas.

O fornecimento dos dados pelo Banco do Brasil ao pesquisador ocorreu por meio de planilha eletrônica. A amostra recebida, em padrão anonimizado (sem a identificação nominal dos clientes), contemplou apenas parcialmente as variáveis solicitadas previamente. Contudo, o Banco do Brasil disponibilizou informações referentes a variáveis que não haviam sido solicitadas, sendo que essa também foram objeto da análise. Os dados recebidos constavam em padrões diversos de acordo com suas características, sendo 33 variáveis em padrão quantitativo, 30 variáveis em padrão categórico e 55 em padrão dummy (tomando valor igual a 1 quando há presença da variável em questão, ou 0 quando não há), conforme Quadro 17 constante do Apêndice A. No Quadro 17, do Apêndice A, demonstra-se parcialmente o padrão dos dados recebidos. As 39.216 linhas e 116 colunas de variáveis, sendo que as duas primeiras colunas se referem a identificação anonimizada do cliente e o período sobre os quais as demais variáveis explicativas (colunas) fazem referência.

3.3.2 Tratamento dos dados

O tratamento estatístico dos dados e o desenvolvimento dos modelos preditivos foram realizados por meio do software RSTUDIO (RStudio Team, 2015), que é uma ferramenta de desenvolvimento para o ambiente de computação estatística em R, que conforme R Core Team (2019), é uma linguagem de programação especializada em manipulação, análise e visualização gráfica de dados, sendo considerado um dos melhores ambientes computacionais para as finalidades do presente estudo (ÇAVDAR; FERHATOSMANOĞLU, 2018).

Vandecruys et al. (2008) salientam que as etapas de pré-processamento de dados, como seleção de variáveis de entrada e amostragem, podem ter um impacto significativo no resultado final, possivelmente até em uma extensão maior do que a escolha da técnica de modelagem. Para Agresti e Finlay (2017), a estratégia de seleção das variáveis explicativas é um fator crítico para a efetividade de um modelo. Pode parecer óbvio incluir todo preditor potencialmente útil, porém, infelizmente, de modo geral isso é inadequado, especialmente quando o número de preditores é grande. Dispor de variáveis extras que acrescentam pouco poder de previsão em um modelo, notadamente pela sobreposição de outras variáveis, gera desvantagens, tornando o modelo de mais difícil interpretação, o que pode resultar em erros padrão inflados das estimativas dos parâmetros, impossibilitando a avaliação das contribuições parciais das variáveis que são efetivamente importantes. Ainda, Agresti e Finlay (2017) salientam que é desejável que as variáveis explicativas estejam correlacionadas com a variável dependente, mas que apresentem baixa ou nenhuma correlação entre si.

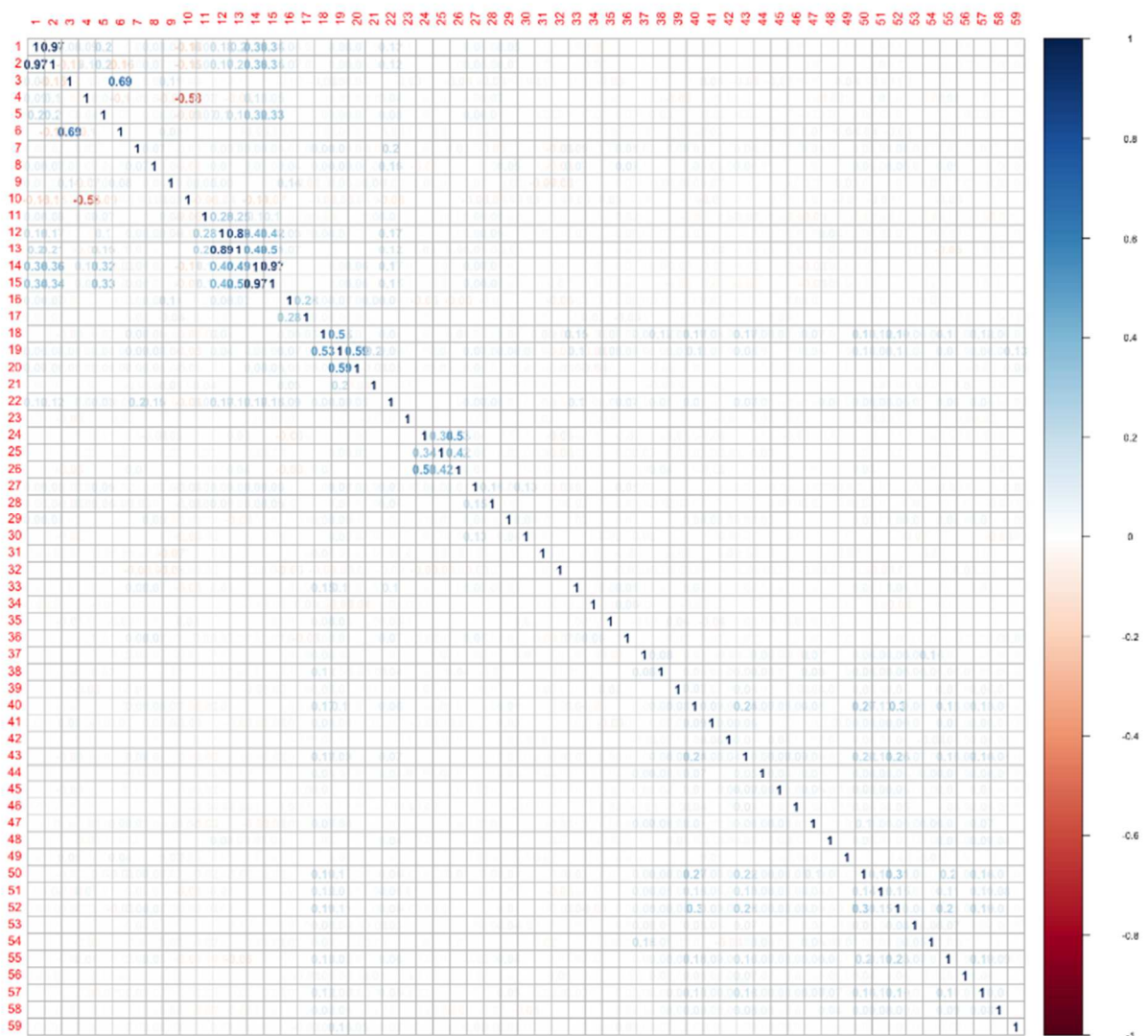
Hadden et al. (2007) ressaltam a importância de uma cuidadosa análise prévia dos dados, sugerindo exclusão de itens redundantes ou com pouco poder informativo, o que agiliza o processo analítico. Neste sentido, de forma a otimizar os processos analíticos foi efetuada uma delimitação inicial das variáveis em aderência ao sugerido por Agresti e Finlay (2017), notadamente as que se caracterizavam por um perfil meramente informativo e de reduzido perfil preditivo.

Em relação as variáveis explicativas, ainda que tenham se apresentado algumas ocorrências de dispersão, não foram identificados padrões inusitados ou outras incoerências. Identificou-se potencial erro de captura da base mensal nos períodos de 06/2018 e 07/2018. No período 05/2018 havia 2499 clientes, e 08/2018 eram 2433. As posições de 06/2018 e 07/2018 agregadas totalizaram 2456 (826 e 1630, respectivamente), evidenciando, inclusive, uma posição intermediária entre as posições 05/2018 e 08/2018. Compreendemos tratar-se, pela visão agregada, de apenas 01 mês de referência. Sendo as variáveis medidas por sua oscilação entre os meses observados, e tendo a calibração ocorrida pela oscilação observado entre o período inicial ao final, não vislumbramos fragilidade que poderia acarretar em prejuízo para as análises. Foi optado por dar continuidade aos trabalhos de análise nos termos dispostos uma vez que articular nova captura junto a instituição financeira poderia ser oneroso em relação ao tempo.

Os dados representativos de valores monetários, em reais, não foram objeto de ajuste ou correção. A inflação acumulada no período que compreendeu as análises, entre novembro de 2017 a janeiro de 2019, alcançou 4,26%, conforme indicador INPC/IBGE.

Além disso, de forma a mitigar ocorrências de multicolinearidade entre as variáveis independentes, em aderência ao sugerido por Agresti e Finlay (2017), foi realizada verificação de correlação dentre as variáveis residuais utilizadas, cuja demonstração gráfica a partir da mensuração do grau de relação entre cada par de variáveis consta na Figura 6, a seguir. Por meio dessa análise, repetida em cada período analisado, foram eliminadas da análise as variáveis que apresentavam mais de 80% de correlação com alguma outra variável.

Figura 6 - Verificação de correlação



Fonte: Software RStudio (2019).

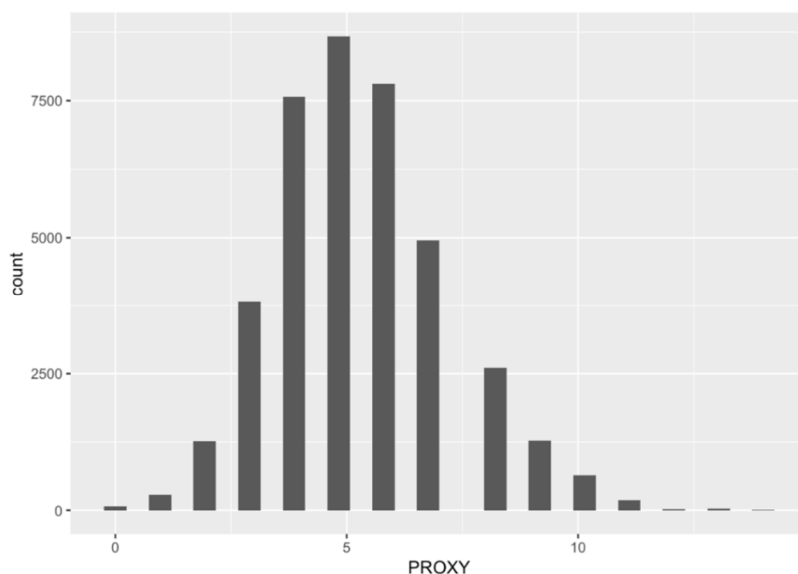
3.3.3 Construção da variável dependente

A definição da variável dependente, que em um nível conceitual objetiva medir a desidratação dos clientes, foi realizada em uma sequência de etapas, de forma a alcançar melhor aderência a realidade do contexto bancário, cujas relações negociais cliente-banco abrangem, normalmente, diversos grupos de produtos. O cliente que possui relacionamento bancário, normalmente além da conta corrente, também possui produtos de investimento, operações de crédito, seguros pessoais e patrimoniais, dentre outros.

Ainda que para este estudo busquemos identificar aqueles casos de *churn* total, envidamos esforços para mapear as ocorrências intermediárias de desidratação, haja vista que as ocorrências de *churn* parcial são fenômenos mais comuns e já sinalizam a perda de negócios, aderente aos conceitos de Hadden et al. (2007). Para esta finalidade, construímos como marcador (variável dependente Y) que expressa a desidratação que buscamos analisar, definido pela variável PROXY, representada pela quantidade de produtos observado, dentre 18 distintos produtos de diversos grupos, sem distinção de peso, conforme etapas que seguem.

- 1) Inicialmente verificamos a quantidade agregada de produtos possuídos pelos clientes da amostra em cada período de corte, conforme histograma a seguir.

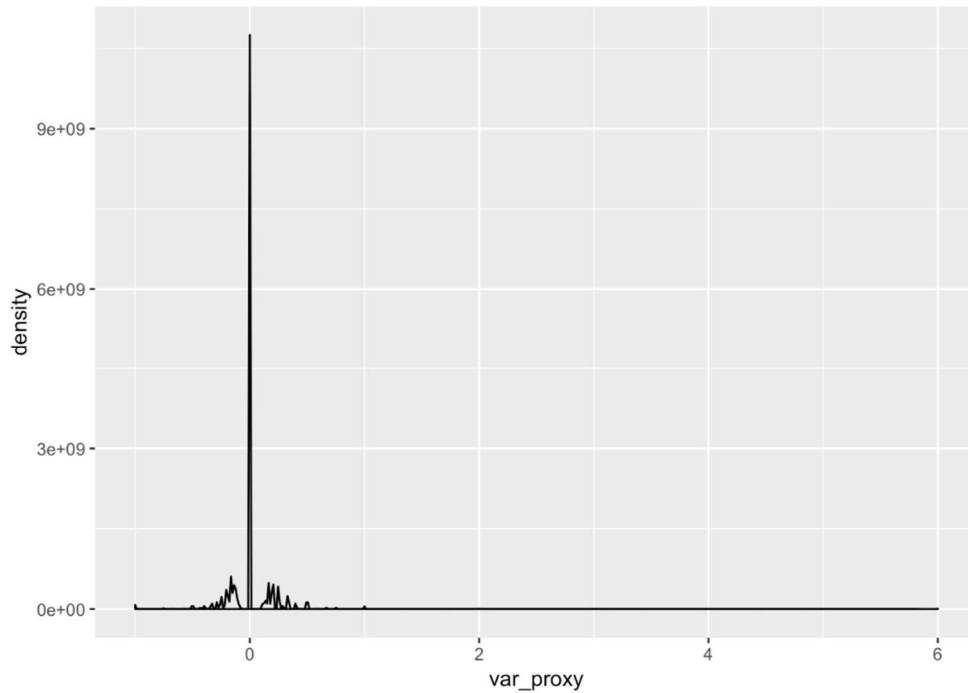
Figura 7 - Histograma da quantidade agregada de produtos



Fonte: Elaborado pelo autor por meio do software RStudio (2019).

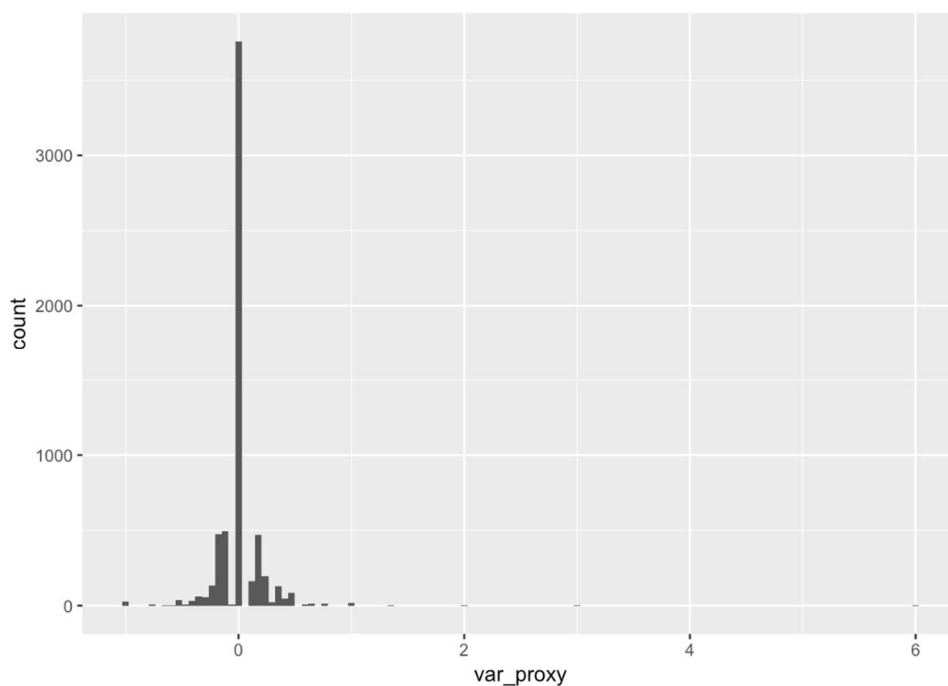
- 2) Posteriormente, foi computada a diferença entre a quantidade de produtos possuídos no início do intervalo de medição e no final do intervalo de medição, tendo sido obtido no gráfico de densidade a seguir.

Figura 8 - Gráfico de densidade da variação da quantidade de produtos



Fonte: Elaborado pelo autor por meio do software RStudio (2019).

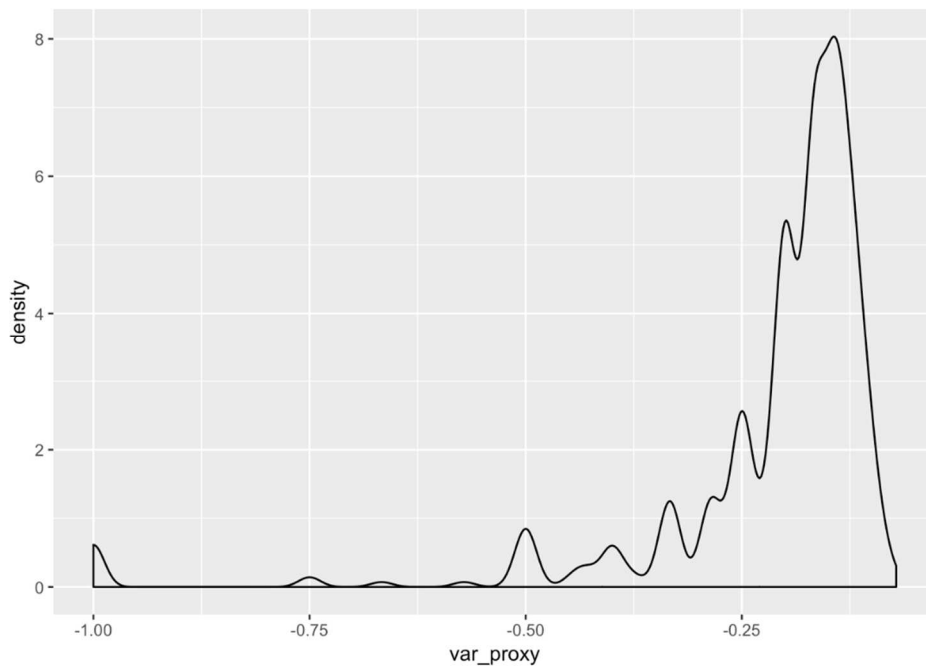
Figura 9 - Variação da quantidade de produtos (contagem)



Fonte: Elaborado pelo autor por meio do software RStudio (2019).

Evidencia-se uma distribuição centrada em praticamente Zero. O grupo de clientes que tiveram oscilação negativa, configurado pela redução da quantidade de produtos, portanto, estão posicionados à esquerda da posição 0 do eixo X. O gráfico de densidade a seguir detalha o comportamento do grupo de clientes com oscilação negativa da quantidade de produtos possuídos no intervalo.

Figura 10 – Gráfico de densidade da oscilação negativa de produtos.



Fonte: Elaborado pelo autor por meio do software RStudio (2019).

3) Pela análise descritiva do grupo que teve variação de proxy negativa, conforme Quadro 7, evidencia-se uma mediana de -16,67%. A mediana é um indicador de referência que separa a metade maior e a metade menor de uma amostra. Agresti e Finlay (2017) ressaltam que a mediana é mais apropriada do que a média quando as distribuições expressam assimetria. Porém, considerando a distribuição estatística dos clientes que apresentaram variação negativa na quantidade de produtos possuídos no intervalo, optou-se por considerar como cliente desidratado aquele cliente cuja variação na variável proxy foi mais intensa (em módulo) que a média dos clientes cujas variações foram negativas.

Quadro 7 - Output variável proxy

##	var_proxy
##	Min. : -1.00000
##	1st Qu.: -0.22222
##	Median : -0.16667
##	Mean : -0.21477
##	3rd Qu.: -0.14286
##	Max. : -0.07143

Fonte: Gerado pelo software RStudio (2019).

- 4) A partir das definições acima, arbitrou-se a condição de “cliente desidratado” àqueles que se enquadravam na classificação definida. A variável dependente (Y), de forma categórica, foi expressa conforme abaixo:

$$y_t = \begin{cases} 0, & (\text{não desidratado}) \\ 1, & (\text{desidratado}) \end{cases}$$

A Prevalência é uma métrica que indica a quantidade total de casos observados em uma amostra. Por entender que a desidratação é um fenômeno mais comum do que o *churn* total, foi considerada que a referida calibração é adequada aos objetivos propostos.

Ainda, seguem algumas considerações sobre a definição da variável dependente. Notadamente, ainda que outras variáveis pudessem também ser “marcadoras” do fenômeno do *churn*, como por exemplo as relacionadas a Margem de Contribuição, na prática isso não se mostrou adequado. O indicador Margem de Contribuição fornecido pela Instituição Financeira reflete a efetiva rentabilidade do cliente para organização e expressa em nível individual questões diversas que sensibilizam este indicador, como por exemplo o fluxo de PCLD⁹. Em muitos casos a margem de contribuição expressa uma oscilação negativa, o que poderia ser interpretado como característico do *churn*, mesmo em situações que o cliente não está desidratando ou evadindo, pelo contrário. Por exemplo, quando um cliente contrata operação de crédito, por questões fiscais, o Banco terá de contabilizar/provisionar em suas demonstrações contábeis uma fração do valor do empréstimo a título de PCLD, o que reduz a margem de contribuição. Outro exemplo é quando o cliente

⁹ PCLD: Provisão de Crédito para Liquidação Duvidosa é uma rubrica redutora do ativo circulante, o que gera efeito negativo no resultado contábil das organizações. Instituições Financeiras necessitam provisionar suas operações de crédito.

transfere/gasta as milhas oriundas de programa de benefícios de cartão de crédito. A contabilização da despesa decorrente do custo das milhas também deduz a margem de contribuição. Essas situações, dentre outras, poderiam ser interpretados como sintomáticas do *churn* sob a perspectiva da margem de contribuição, porém, na prática, expressam clientes que absolutamente estão fortalecendo seu vínculo.

3.3.4 Modelos e métodos preditivos para o fenômeno do *churn*

Conforme evidenciado no Quadro 4, diversas são as técnicas passíveis de serem implementadas para execução de modelos preditivos ao *churn*, porém algumas se destacam por seus resultados, notadamente a Regressão Logística, Florestas Aleatórias (Random Forests) e Redes Neurais Artificiais, sendo estas as escolhidas para consecução dos trabalhos. A definição da escolha das técnicas de mineração de dados a ser utilizada é um trade-off entre a precisão e a interpretabilidade que se pretende obter, conforme Veloso (2013). Neslin et al. (2006) ponderam que técnicas de mais fácil interpretação podem ter uma performance preditiva não tão boa. Modelos como Random Forests e Redes Neurais Artificiais podem, por vezes, apresentar melhores resultados nas análises preditivas, mas são mais sensíveis a base de dados. Outras técnicas consideradas de mais fácil interpretação, como Regressão Logística, permitem interpretar as razões e sintomas do *churn*.

Berry e Linoff (2011) ressaltam a importância de modelar o *churn* com resultado binário. Modelos preditivos de *churn* com desfecho binário, como as que serão desenvolvidas neste estudo, têm um horizonte de tempo definido, o que atende ao melhor aproveitamento dos resultados da pesquisa. Os dados históricos do cliente são combinados com um marcador, a variável dependente, que define cada cliente como ativo ou não. Após isso, os algoritmos de classificação irão indicar a probabilidade de *churn* dos clientes analisados. Um horizonte de tempo não pode ser muito curto, pois, neste caso, não restará tempo para atuar com ações de retenção.

A limitação do acesso a dados, às vezes, pode ser um fator restritivo para desenvolvimento de algumas técnicas. Vandecruys et al. (2008) realizaram extensiva revisão bibliográfica e elencaram alguns achados relativos a diferentes técnicas de predição do *churn*. Os autores ponderam que o impacto das características de um conjunto de dados no desempenho de uma técnica de mineração de dados é o mesmo para todas as técnicas. Comparando os resultados e rankings de técnicas aplicadas

em uma variedade de conjuntos de dados, por outro lado, permite avaliar e estudar o efeito das características dos dados no desempenho de uma técnica. Vandecruys et al. (2008) também salientam que a etapa de pré-processamento de dados, como seleção de variáveis de entrada e amostragem, pode ter um impacto significativo no resultado final, possivelmente até em uma extensão maior do que a escolha da técnica de modelagem.

Aderente a estas considerações, tendo por objetivo entender qual técnica preditiva poderia melhor interpretar o fenômeno que buscamos compreender, responder ao problema de pesquisa, bem como avaliar o nível de performance entre as diferentes técnicas em um contexto bancário, foram desenvolvidas análises pelas abordagens da Regressão Logística, Random Forests, Redes Neurais Artificiais, e, ainda, a Regressão Logística com seleção de variáveis independentes do modo Stepwise. Com vistas a melhor compreender as características conceituais destas técnicas, em aderência a Khakabi, Gholamian e Namvar (2010), seguem suas definições.

3.3.4.1 Regressão Logística

A regressão logística é uma técnica estatística clássica usada para classificação. É especificamente adaptada à situação em que a variável dependente é dicotômica (ou binária). Na regressão logística, a variável dependente é o logaritmo natural dos odds de ocorrência de um fenômeno (probabilidade de ocorrência de um evento dividida pela probabilidade de não ocorrência). O modelo logístico é adequado para definir a relação entre a probabilidade de um cliente evadir/abandonar $p(x)$ e um vetor de características do cliente $[x_1, \dots, x_k]$, definido pela função *logit* na expressão a seguir:

$$\log \left\{ \frac{\pi(x)}{1 - \pi(x)} \right\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Essa variável dependente, para ser interpretada sob a forma de probabilidade de ocorrência do evento analisado (sendo a probabilidade por definição limitada entre 0 e 1), deve ser transformada na equação abaixo:

$$\pi(x) = e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)} / [1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}]$$

Outros pontos elencados por Buckinx e Van Den Poel (2005) sobre a metodologia da regressão logística, que a consideram adequada à predição do *churn*, por cinco razões principais:

- a) a modelagem é bem conhecida, conceitualmente simples, frequentemente usada em marketing, especialmente no nível do consumidor individual;
- b) a facilidade de interpretação do *logit*;
- c) a modelagem *logit* mostrou fornecer bons e robustos resultados em estudos gerais preditivos (respaldado no referencial teórico);
- d) mais especificamente nas abordagens de métricas do marketing, foi demonstrado por vários autores que a modelagem *logit* pode até mesmo superar métodos mais sofisticados;
- e) enquanto a análise de sobrevivência nos daria o tempo estimado para o evento (neste caso, *churn*), estamos interessados apenas na probabilidade estimada do evento dentro de um dado período.

3.3.4.2 Regressão Logística Stepwise

A técnica de seleção de atributos denominada Stepwise (passo-a-passo) complementa a abordagem prevista no método da Regressão Logística, ponderado que a função *logit* é muito sensível a qualidade dos dados. Se os dados contiverem informações irrelevantes, a Regressão Logística poderá produzir resultados menos precisos. Portanto, a seleção de recursos Stepwise é frequentemente usada em combinação com a regressão logística, como um pré-processamento. Trata-se de uma ferramenta automatizada usada nos estágios exploratórios da construção de modelos para identificar um subconjunto útil de preditores, através de rotinas que adicionam ou removem sistematicamente as variáveis, passo-a-passo, de acordo com critérios específicos. O processo continua até que cada variável remanescente explique uma quantia parcial significativa da variabilidade em *y*.

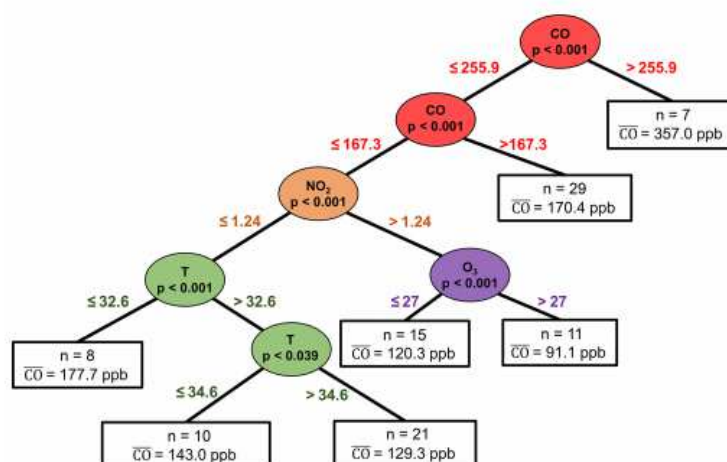
Agresti e Finlay (2017) ressaltam, por isso, que no modelo final da Regressão Stepwise todas as variáveis remanescentes tendem a ser significativas. Entretanto, quando duas variáveis são altamente correlacionadas, é possível que apenas uma permaneça no modelo, mesmo sendo a outra importante. Logo, por seu caráter automatizado, as rotinas da metodologia Stepwise desconsideram o conhecimento

especializado que o analista possa ter sobre dados, podendo o modelo selecionado não ser o melhor sob alguma perspectiva prática.

3.3.4.3 Random Forests

Random forests ou florestas aleatórias cultivam muitas árvores de decisão (Decision Tree). Este método realiza classificações a partir de um vetor de entrada, que é analisado em cada uma das árvores da floresta, conforme representação da Figura 11. Cada árvore dá uma classificação e dizemos que a árvore vota para essa classe. A floresta escolhe a classificação com maior número de votos (sobre todas as árvores da floresta). Essa metodologia é resistente à falta de dados e pode manipular elevadas quantidades de atributos, sendo considerada facilmente interpretável em termos de qualidade preditiva.

Figura 11 - Representação gráfica - Random Forests



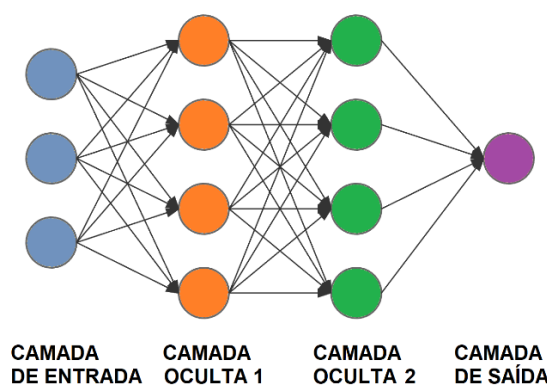
Fonte: Zimmermann et al. (2018).

3.3.4.4 Redes Neurais

Os algoritmos de Redes Neurais Artificiais (Artificial Neural Networks) são classificados como *Deep Learning* (aprendizado profundo), um ramo de aprendizado da máquina (*machine learning*) com grandes capacidades de processamento (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Redes Neurais Artificiais é uma modelagem representativa de um sistema de computador projetado para emular a

maneira como o cérebro humano opera. É constituído por um conjunto de neurônios organizados em camadas. Cada neurônio em uma camada é conectado a outros neurônios na camada resultante por meio de várias arestas ponderadas. Essas relações entre dados aparentemente desconexos podem gerar insights e informações, conforme demonstra a figura 12. Os algoritmos de aprendizagem profunda (Deep learning) podem ser muito superiores aos métodos tradicionais de regressão e classificação (por exemplo, regressão linear e logística), devido à capacidade de modelar interações e não-linearidades que, de outras formas, não seriam detectadas.

Figura 12 - Representação figurativa de Redes Neurais



Fonte: Elaborado pelo autor (2019).

3.3.5 Métricas de avaliação de modelos de classificação

De acordo com Botelho e Tostes (2010), com vistas a minimizar o risco de decisões errôneas na classificação dos clientes, muitos métodos estatísticos têm sido utilizados e descrevem a habilidade do modelo de classificação. A avaliação da qualidade preditiva pode ser avaliada através de medidas como acurácia, sensibilidade, especificidade. A etapa de validação do modelo preditivo, a partir dos resultados obtidos, contemplará, portanto, mais alguns passos, imprescindíveis para compreender a efetividade de cada modelo.

Castro e Braga (2011) ponderam que, tradicionalmente, a métrica usada na avaliação e seleção de modelos de classificação é a acurácia, estimada em relação a um dado conjunto de teste. Justifica-se esse critério uma vez que se busca a menor probabilidade de erro global. Porém, ressaltam, entretanto, que uma maneira mais eficaz de se avaliar um dado classificador é através da distinção dos erros (ou acertos) cometidos para cada classe. Isso pode ser obtido descrevendo o desempenho a partir

de uma matriz de confusão, que consiste em uma matriz interpor os dados reais e os valores preditos pelo modelo, conforme exemplo a seguir:

Quadro 8 - Matriz de confusão para um classificador binário

Matriz de confusão		Previsto pelo Modelo	
		Não <i>churner</i> (N)	<i>Churn</i>
Situação real	Não <i>churner</i> (N)	Verdadeiro Negativo (TN)	Falso Positivo (FP)
	<i>Churner</i> (P)	Falso Negativo (FN)	Verdadeiro Positivo (TP)

Fonte: Elaborado pelo autor 2019).

Conforme Castro e Braga (2011), ao longo da diagonal principal (em azul), estão representadas as predições corretas do modelo: verdadeiros positivos (TP) e verdadeiros negativos (TN); os elementos fora dessa diagonal representam os erros cometidos pelo modelo: falsos positivos (FP) e falsos negativos (FN).

Luque et al. (2019) destacam importantes métricas que podem ser extraídas da matriz de confusão, conforme abaixo e quadro resumo a seguir, que irão compor os demonstrativos dos resultados:

$$\text{Sensibilidade ou Taxa de Verdadeiros Positivos:} = \frac{TP}{TP+FN}$$

$$\text{Especificidade ou Taxa de Verdadeiros Negativos:} = \frac{TN}{TN+FP}$$

$$\text{Acurácia ou Taxa de Casos Corretamente Classificados:} = \frac{TN+TP}{TN+TP+FN+FP}$$

$$\text{Precisão ou Taxa de Positivos Previstos:} = \frac{TP}{TP+FP}$$

Quadro 9 - Resumo da Matriz de Confusão e métricas para avaliação de modelos.

Matriz de Confusão e métricas derivadas para avaliação de modelos de classificação.							
Matriz de confusão		Previsto pelo Modelo		Sensibilidade	Especificidade	Precisão	Acurácia
		Não churner	Churner				
Situação real	Não churner (N)	Verdadeiro Negativo (TN)	Falso Positivo (FP)	$= \frac{TP}{TP + FN}$	$= \frac{TN}{TN + FP}$	$= \frac{TP}{TP + FP}$	$= \frac{TP + TN}{TP + TN + FP + FN}$
	Churner (P)	Falso Negativo (FN)	Verdadeiro Positivo (TP)				

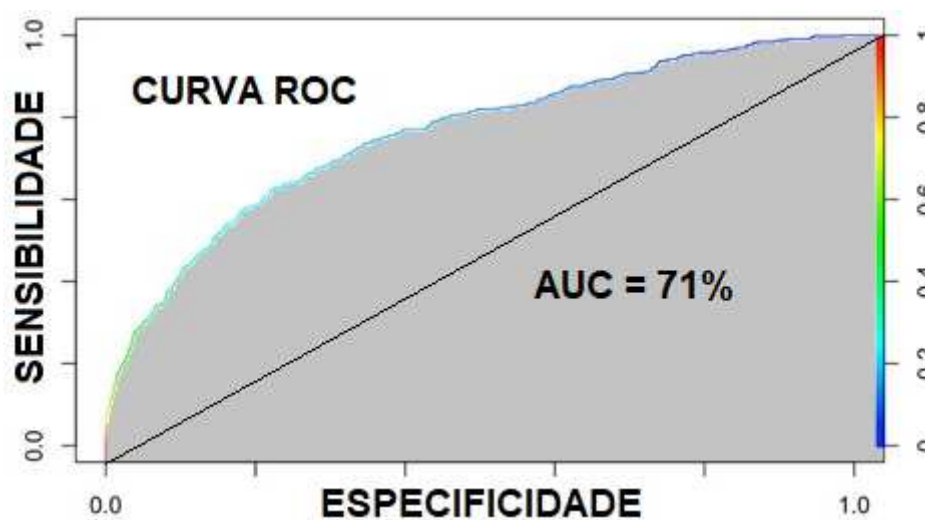
Fonte: Elaborado pelo autor (2019)

Amparados por Veloso (2013), Castro e Braga (2011) e Luque et al. (2019), seguem alguns comentários sobre as métricas acima elencadas e suas aplicações para avaliação de resultados em modelos de classificação. A Acurácia é, normalmente, a medida de performance mais usada. Entretanto, ela pode não ser adequada a alguns problemas. No caso da previsão do *churn*, o principal objetivo é identificar o máximo de casos de *churn*, ou seja, os falsos positivos e os falsos negativos têm impactos claramente diferentes. Por exemplo, se em dado conjunto de dados possui 90% de casos negativos (não *churn*) e 10% positivos (*churn*) e o modelo classifica todos como negativos, a precisão do modelo será de 90%, o que poderia ser considerada uma boa performance, apesar de absolutamente inócua, uma vez que não identificou nenhum caso de *churn*. A Sensibilidade é a proporção de exemplos positivos que foram corretamente classificados como positivos. A Especificidade é o contrário, mede a proporção dos exemplos negativos corretamente classificados como negativos. Já a Precisão refere-se à proporção de casos corretamente classificados como positivos em relação ao total de casos classificados como positivos (corretos + incorretos). Derivada da Precisão existe a métrica Prevalência Indicada, que se refere ao número de casos positivos (verdadeiros + falsos) indicados pelo modelo dentro da amostra. Já a Prevalência refere-se ao número de casos positivos reais observados dentro de toda amostra.

Também derivada da Matriz Confusão, o trade off da relação entre os indicadores de Sensibilidade e Especificidade, em um plano gráfico, gera o que se denomina de Curva ROC (Receiver Operating Characteristic). Conforme Castro e Braga (2011), a curva ROC fornece uma estimativa da capacidade discriminativa do classificador em termos das probabilidades de erro, refletindo os erros de classificação. Portanto, controla a fração de exemplos corretamente classificados versus fração de exemplos incorretamente classificados.

A AUC – Area Under Curve fornece uma métrica para avaliar, em média, qual o melhor algoritmo, sendo quanto maior a área entre a Curva ROC e a diagonal principal, melhor a performance do modelo (Veloso, 2013). Por exemplo, uma AUC de 0,71 significa que há 71% de chances de que o modelo seja capaz de distinguir entre classe positiva ou classe negativa, conforme figura a seguir:

Figura 13 - Curva ROC e AUC



Fonte: Elaborado pelo autor (2019).

O Coeficiente de Concordância de Kappa é utilizado para descrever a concordância entre duas ou mais avaliações sobre uma mesma amostra. Já o AIC (Critério de Informação de Akaike), é uma medida relativa da qualidade de ajuste de um modelo estatístico estimado. Sob este critério, o melhor modelo será o que apresentar o menor valor de AIC.

O teste Kolmogorov-Smirnov (KS), conforme Botelho e Tostes (2010), consiste em uma estatística não paramétrica com o objetivo de testar se as funções de distribuição de dois grupos são iguais. Essa métrica avalia o desempenho do modelo e dimensiona sua capacidade de distinguir o status dos clientes entre ativos e inativos, representando a diferença máxima entre as distribuições acumuladas dos clientes ativos e inativo. O Teste K-S é um dos métodos mais úteis e difundidos para a comparação de duas amostras, já que é sensível a diferenças tanto no local, como na forma das funções de distribuição acumulada empírica das duas amostras.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Este capítulo apresenta a análise dos dados e os resultados alcançados na consecução da etapa quantitativa da pesquisa.

4.1 ANÁLISE DESCRITIVA DOS DADOS

As estatísticas descritivas (medidas que resumem as informações contidas nos dados) das principais variáveis explicativas são apresentadas na TABELA 2, a seguir, delimitadas aos meses do período constante entre novembro de 2017 a janeiro de 2019. Estes dados, fornecidos pelo Banco do Brasil, contemplaram o grupo de clientes que estava presente a partir do primeiro período de análise. Os casos de desidratação (*churn* parcial) e *churn* total, foram mantidos nas estatísticas, ainda que não expressassem valores. Neste grupo, foram observados 307 casos de clientes que deixaram de apresentar margem de contribuição por decorrência do cancelamento de produtos nestes 15 meses, o que pode caracterizá-las como *churn* total.

Tabela 2 - Análise descritiva das variáveis explicativas

Estatísticas descritivas							
Variáveis explicativas	N (qtde)	Média	Desvio Padrão	Mínimo	Percentil (25%)	Percentil (75%)	Máximo
PROXY - Variável dependente	35231	5,46	1,79	0,00	4,00	7,00	14,00
MARGEM.DE.CONTRIBUICAO	35231	421,07	1431,34	-72998,62	86,27	452,26	95712,48
MARGEM.CARTÃO.DE.CRÉDITO	33218	72,67	351,92	-16626,89	0,00	125,15	11832,64
MARGEM_CXQ_ESP	33222	32,12	181,36	-899,07	0,00	0,00	5699,46
MARGEM_EMPRESTIMO	33222	45,65	389,74	-22378,40	0,00	0,00	26323,37
VALOR PACOTE DE SERVIÇOS	35231	54,24	14,11	0,00	58,28	58,28	79,79
INDICADOR_POSSE_ROTATIVO	32895	0,83	0,37	0,00	1,00	1,00	1,00
VLR_RENDIMENTO	35231	15074,41	26945,55	0,00	3668,37	17795,81	611464,70
VLR_TOT_CHQ_ESP	31254	-319,01	1565,74	-44001,92	-0,96	0,00	0,00
ENDIVIDAMENTO.SFN	35231	108688,50	241090,20	0,00	0,00	105820,60	5392764,00
ENDIVIDAMENTO.BB	35231	48769,01	85892,45	0,00	5104,57	56179,54	3081992,00
VLR_TOT_EMP	33218	21884,95	79344,95	0,00	0,00	0,00	2986824,00
VLR_CDC_TOTAL	12953	29247,09	64570,35	0,00	0,00	25037,49	531410,30
VLR_CDC_EMP	12586	28163,75	63692,53	0,00	0,00	21258,80	529545,50
VLR_INVESTIMENTO	33222	218176,60	410663,60	0,00	3551,52	259647,00	6704983,00
VLR_TOT_POUPANCA	33218	19890,32	105979,50	0,00	0,00	4317,18	4248967,00
QTDE_INTERACOES_AG	33221	19,53	20,16	0,00	4,00	30,00	236,00
QTDE_INTERACOES	33221	68,74	61,89	0,00	23,00	99,00	759,00
QTDE_INTERACOES_MOBILE	33221	19,46	30,77	0,00	0,00	28,00	382,00
QTDE_PGTOS_INTERNET	33221	0,76	2,31	0,00	0,00	0,00	40,00
IDADE	35231	56,24	11,22	23,00	48,00	63,00	95,00
QTDE_SIMULA_CDC	33221	0,24	0,43	0,00	0,00	0,00	1,00
QTDE_SIMULA_CDC_RENOV	33221	0,06	0,23	0,00	0,00	0,00	1,00
QTDE_SIMULA_FIN_IMOB	33221	0,06	0,24	0,00	0,00	0,00	1,00
QTDE_SIMULA_FIN_VEICULO	33221	0,01	0,09	0,00	0,00	0,00	1,00
RENDA_PRESUMIDA	33272	6919,97	5075,07	0,00	2875,00	10200,00	30000,00
SEGMENTO_CLIENTE	35231	2222,57	217,52	2012,00	2112,00	2112,00	2712,00
USO_AG	33221	0,89	0,31	0,00	1,00	1,00	1,00
USO_ATEND_INTERNET	33221	2,12	0,87	0,00	2,00	2,00	5,00
USO_CALL_CENTER	33221	0,05	0,22	0,00	0,00	0,00	1,00
UTLZ_CNL_INTERNET	33221	0,84	0,37	0,00	1,00	1,00	1,00
UTZ_CART_SUPERMERCADO	26855	0,70	0,46	0,00	0,00	1,00	1,00
UTZ_CART_RESTAURANTE	26855	0,58	0,49	0,00	0,00	1,00	1,00
UTZ_CART_SERVICOS	26855	0,57	0,50	0,00	0,00	1,00	1,00
UTZ_CART_VAREJO	26855	0,57	0,49	0,00	0,00	1,00	1,00
UTZ_CART_COMBUSTIVEL	26855	0,52	0,50	0,00	0,00	1,00	1,00
UTZ_CART_DROGARIA	26855	0,52	0,50	0,00	0,00	1,00	1,00
UTZ_CART_VESTUARIO	26855	0,46	0,50	0,00	0,00	1,00	1,00
UTZ_CART_CONSTRUCAO	26855	0,37	0,48	0,00	0,00	1,00	1,00
VLR_SAQUE	33221	1410,65	4122,67	0,00	0,00	1100,00	88000,00

Elaborado pelo autor (2019).

4.1.1 Teste K-S

O Teste Kolmogorov-Smirnov testa se as funções de distribuição de dois grupos são iguais. Tendo as variáveis explicativas sido parametrizadas para serem medidas por sua variação entre cada período de corte analisado, somente estas oscilações que são capturadas e analisadas, em relação a si mesmo, dentro de cada grupo e de cada variável. Em relação ao indicador p-value do KS Test, somente quando se apresenta inferior a 0,05 que os grupos, em relação a determinada variável, podem ser considerados estatisticamente diferentes. Por exemplo, observa-se que o grupo considerado desidratado alcançou perfis que não podem ser considerados estatisticamente diferentes (p-value > 0,05), decorrente das variáveis Idade (1,00), Renda Presumida (0,37) e VLR Rendimento (0,18), evidenciando uniformidade na ocorrência da desidratação.

Em relação a variável Margem de Contribuição o grupo desidratado ($D = 1$) teve variação média de -R\$261,61, já o grupo retido evoluiu sua margem em +R\$27,00; O grupo desidratado, na média, reduziu seu volume de endividamento junto ao Banco do Brasil em -R\$2988,58, enquanto o grupo retido variou positivamente em R\$2685,51; Em relação a variável "Endividamento SFN¹⁰, evidencia que o grupo desidratado aumentou suas operações em R\$699,53 junto aos demais bancos. Já o grupo retido, reduziu o saldo de operações externas em -R\$1073,23;

Conclusivamente evidenciamos dois pontos: O grupo retido está elevando o volume de suas operações no BB ao mesmo tempo em que reduz o volume de operações junto a outros bancos. Já no grupo desidratado ocorre o contrário, enquanto reduz suas operações no BB, aumenta o volume junto aos demais bancos. Aderente ao referencial teórico, objetivamente temos configurada ocorrência de *churn*, bem como comprova-se a adequada configuração da modelagem proposta.

¹⁰ Endividamento SFN: indicador fornecido pelo Banco Central do Brasil (BACEN) que indica o saldo das operações de cada cliente junto a outras instituições do Sistema Financeiro Nacional (SFN).

Tabela 3 - Teste Kolmogorov-Smirnov (KS)

Teste Kolmogorov-Smirnov. D = 0 é o grupo de clientes retidos. D = 1 é o grupo de clientes que desidratou.									
Variáveis explicativas	Média		Desvio Padrão		Máximo		Mínimo		p-value - KS.Test
	D = 0	D = 1	D = 0	D = 1	D = 0	D = 1	D = 0	D = 1	
PROXY	0,39	-0,28	1,03	1,31	8,00	4,00	-4,00	-7,00	0,00
MARGEM.DE.CONTRIBUICAO	27,00	-261,61	1021,63	3022,42	26771,34	5266,67	-35691,41	-73231,28	0,00
MARGEM.CARTÃO.DE.CRÉDITO	51,34	24,03	353,78	485,19	4848,07	11244,88	-3709,01	-2911,89	0,00
MARGEM_CXQ_ESP	-2,39	-27,11	153,99	308,16	1957,60	2114,32	-2904,31	-5367,36	0,58
MARGEM_EMPRESTIMO	12,44	-7,91	592,69	436,00	26323,37	6595,57	-13036,07	-6548,31	0,00
VALOR PACOTE DE SERVIÇOS	1,30	-0,10	8,38	9,67	58,28	58,28	-58,28	-58,28	0,73
INDICADOR_POSSE_ROTATIVO	0,01	-0,04	0,22	0,25	1,00	1,00	-1,00	-1,00	0,20
VLR_RENDIMENTO	1587,31	956,62	17573,87	15058,79	318658,10	116632,95	-211515,90	-240443,40	0,18
VLR_TOT_CHQ_ESP	-38,86	79,31	1469,97	2423,16	22120,77	30118,95	-23018,46	-31698,34	0,39
ENDIVIDAMENTO.SFN	-1076,23	699,53	132141,86	171557,44	2062551,80	2292322,04	-3664872,54	-2294864,05	0,00
ENDIVIDAMENTO.BB	2685,51	-2988,58	31437,58	78385,65	493810,69	427931,30	-259723,50	-1390238,57	0,00
VLR_TOT_EMP	781,72	-3861,85	27628,86	71544,95	397435,22	394690,21	-267281,59	-1373048,19	0,00
VLR_CDC_TOTAL	1241,93	-1453,45	20785,81	30502,93	397435,22	394690,21	-200047,51	-263254,39	0,00
VLR_CDC_EMP	1411,44	-500,57	20362,62	28017,72	397435,22	394690,21	-200047,51	-224957,83	0,00
VLR_INVESTIMENTO	22857,88	-16848,85	192955,85	190073,09	3959801,42	1247864,66	-3819351,04	-1433956,16	0,00
VLR_TOT_POUPANCA	3017,22	2983,46	55160,44	34154,96	2003198,35	700049,55	-924778,78	-131659,10	0,02
QTDE_INTERACOES_AG	1,78	1,39	13,02	15,70	139,00	150,00	-92,00	-85,00	0,00
QTDE_INTERACOES	0,40	-1,34	32,21	38,13	185,00	277,00	-349,00	-257,00	0,00
QTDE_INTERACOES_MOBILE	-4,84	-5,65	20,34	20,60	133,00	74,00	-203,00	-140,00	0,00
QTDE_PGTOS_INTERNET	0,00	-0,11	1,73	1,64	27,00	11,00	-15,00	-12,00	0,34
IDADE	1,08	1,08	0,28	0,27	2,00	2,00	1,00	1,00	1,00
NRO_CONVENIO_13SAL	1588,01	908,67	35868,60	40711,73	367383,00	353322,00	-360239,00	-357632,00	1,00
NRO_CONVENIO_CONSIG	383,50	880,71	22220,76	27204,76	368142,00	352905,00	-351257,00	-225742,00	1,00
NRO_CONVENIO_CRED_SAL	712,07	1638,50	32294,10	47195,49	353321,00	353321,00	-363808,00	-353113,00	1,00
QTDE_SIMULA_CDC	-0,02	-0,02	0,46	0,47	1,00	1,00	-1,00	-1,00	1,00
QTDE_SIMULA_CDC_RENOV	0,01	-0,01	0,22	0,24	1,00	1,00	-1,00	-1,00	0,99
QTDE_SIMULA_FIN_IMOB	0,01	0,01	0,26	0,29	1,00	1,00	-1,00	-1,00	1,00
QTDE_SIMULA_FIN_VEICULO	0,00	0,00	0,10	0,12	1,00	1,00	-1,00	-1,00	1,00
RENDA_PRESUMIDA	1121,74	1285,85	3482,44	3694,47	30000,00	30000,00	0,00	0,00	0,37
SEGMENTO_CLIENTE	-25,73	-15,50	167,93	153,45	600,00	590,00	-600,00	-590,00	0,80
USO_AG	-0,01	-0,04	0,32	0,33	1,00	1,00	-1,00	-1,00	0,95
USO_ATEND_INTERNET	-0,05	-0,08	1,01	0,85	4,00	4,00	-4,00	-4,00	0,51
USO_CALL_CENTER	-0,01	-0,01	0,27	0,32	1,00	1,00	-1,00	-1,00	0,96
UTLZ_CNL_INTERNET	-0,03	-0,03	0,30	0,30	1,00	1,00	-1,00	-1,00	1,00
UTZ_CART_AEREO	0,02	0,02	0,36	0,40	1,00	1,00	-1,00	-1,00	0,97
UTZ_CART_ALUG_CAR	0,02	0,01	0,35	0,41	1,00	1,00	-1,00	-1,00	0,55
UTZ_CART_AUTOPECAS	0,00	0,00	0,37	0,39	1,00	1,00	-1,00	-1,00	1,00
UTZ_CART_COMBUSTIVEL	-0,02	-0,02	0,41	0,46	1,00	1,00	-1,00	-1,00	0,52
UTZ_CART_CONSTRUCAO	0,02	0,01	0,40	0,42	1,00	1,00	-1,00	-1,00	0,91
UTZ_CART_DECORACAO	0,01	0,01	0,32	0,33	1,00	1,00	-1,00	-1,00	1,00
UTZ_CART_DROGARIA	0,00	-0,02	0,45	0,47	1,00	1,00	-1,00	-1,00	0,73
UTZ_CART_ENTRETENIMENTO	0,01	0,00	0,34	0,36	1,00	1,00	-1,00	-1,00	0,98
UTZ_CART_ESTACIONAMENTO	-0,01	0,00	0,29	0,33	1,00	1,00	-1,00	-1,00	0,98
UTZ_CART_HOSPITAL	0,00	0,00	0,37	0,37	1,00	1,00	-1,00	-1,00	1,00
UTZ_CART_HOTEL	0,01	-0,03	0,40	0,45	1,00	1,00	-1,00	-1,00	0,03
UTZ_CART_JOALHERIA	0,00	0,00	0,29	0,30	1,00	1,00	-1,00	-1,00	1,00
UTZ_CART_PEDAGIO	0,01	0,01	0,15	0,13	1,00	1,00	-1,00	-1,00	1,00
UTZ_CART_RESTAURANTE	0,01	0,02	0,42	0,46	1,00	1,00	-1,00	-1,00	0,60
UTZ_CART_SERVICOS	0,04	0,01	0,47	0,50	1,00	1,00	-1,00	-1,00	0,31
UTZ_CART_SUPERMERCADO	0,00	-0,01	0,39	0,37	1,00	1,00	-1,00	-1,00	1,00
UTZ_CART_TELEMARKETING	0,01	-0,03	0,46	0,46	1,00	1,00	-1,00	-1,00	0,80
UTZ_CART_TURISMO	0,01	0,01	0,30	0,33	1,00	1,00	-1,00	-1,00	1,00
UTZ_CART_VAREJO	-0,02	-0,04	0,48	0,49	1,00	1,00	-1,00	-1,00	0,92
UTZ_CART_VEICULOS	0,00	0,01	0,31	0,32	1,00	1,00	-1,00	-1,00	1,00
UTZ_CART_VESTUARIO	-0,01	-0,04	0,49	0,50	1,00	1,00	-1,00	-1,00	0,76
UTZ_ELETRO	0,02	0,01	0,45	0,49	1,00	1,00	-1,00	-1,00	0,56
VLR_SAQUE	29,71	170,95	3850,19	4459,11	48000,00	61500,00	-71000,00	-24100,00	0,14

Fonte: Elaborado pelo autor (2019).

4.2 CONSTRUÇÃO DOS MODELOS DE CLASSIFICAÇÃO

A condução dos trabalhos com o Software RStudio esteve aderente aos manuais operacionais do desenvolvedor, bem como amparado nas rotinas validadas e reconhecidas por Cirillo (2016). Com vistas a buscar a melhor capacidade preditiva e testar o modelo em diferentes períodos, foi adotado como critério de análise um horizonte temporal (período de observação) de 12 meses. Esta opção se justifica uma vez que cortes temporais menores, de 3 ou 6 meses, por exemplo, demonstraram capacidades explicativas inferiores, ponderada a restrição observada do tamanho da base de dados recebida. Além disso, os problemas de balanceamento de classes são majorados, o que normalmente reduz a capacidade preditiva dos modelos. O referencial teórico deste estudo, por seus resultados empíricos, evidencia que sobre amostras maiores pode-se adotar diferentes definições, inclusive com horizonte temporal inferior, sem prejuízo dos resultados (GLADY; BAESENS; CROUX, 2009). Em cada técnica de classificação foram desenvolvidas três análises, com período de observação de conforme abaixo:

- a) A1: janela temporal compreendida entre dezembro/2017 a novembro/2018;
- b) A2: janela temporal compreendida entre janeiro/2018 a dezembro/2018;
- c) A3: janela temporal compreendida entre fevereiro/2018 a janeiro/2019.

As variáveis explicativas independentes foram parametrizadas para serem medidas por sua variação entre cada período mensal. As análises contemplaram Validação Cross-fold (padrão 75/25), em que a amostra divide-se entre conjunto de treinamento (75%) e conjunto de validação (25%).

Em relação as métricas de avaliação dos modelos de classificação, ponderamos alguns pontos. Considerando o contexto e o objetivo de identificar a maior quantidade de clientes positivos corretamente classificados como positivos (*churn real*), a modelagem mais adequada a esta finalidade será a que possui maior Sensibilidade. Porém, ao conduzir uma campanha de retenção junto aos clientes previstos como *churn*, é relevante observar a Precisão de cada modelo, uma vez que um melhor índice de Precisão significa entregar benefícios/incentivos, normalmente financeiros, aos clientes que efetivamente devem ser reconhecidos para mitigar sua evasão. As próximas sessões irão apresentar os resultados observados em cada técnica e posteriormente a síntese comparativa dentre as diferentes técnicas.

4.3 REGRESSÃO LOGÍSTICA

A primeira técnica preditiva desenvolvida foi a regressão logística, aderente as premissas descritas no método. Segue Quadro 10 que apresenta a Matriz Confusão das 3 análises realizadas (A1, A2 e A3), contemplando a média da performance observada nas métricas derivadas da Matriz Confusão. Posteriormente consta o Quadro 11, que consolida os relatórios outputs das 3 análises gerados pelo software RStudio, bem como a média observada nos respectivos indicadores de performance.

Quadro 10 - Matriz de Confusão dos resultados da Regressão Logística

Matriz de Confusão e métricas derivadas para avaliação dos resultados - Regressão Logística								
Análise referênci	Matriz de confusão	Previsto pelo Modelo		Sensibilidade	Especificidade	Precisão	Acurácia	AUC
		Retido	Desidratado					
		Retido	Desidratado					
A1	Retido	1565	41	21,353%	97,447%	71,127%	80,135%	0,73312
	Desidratado	372	101					
A2	Retido	1608	38	18,329%	97,691%	67,521%	81,223%	0,74223
	Desidratado	352	79					
A3	Retido	1608	46	21,362%	97,219%	66,423%	81,683%	0,75677
	Desidratado	335	91					
Média da performance observada				20,348%	97,452%	68,357%	81,013%	0,74404

Fonte: Elaborado pelo autor (2019).

Quadro 11 - Consolidação dos resultados analíticos da Regressão Logística

Técnica de análise	Regressão Logística			
Análise de Referência	A1	A2	A3	Média
Intervalo de confiança (95%)	(0,7835 - 0,8183)	(0,7948 - 0,8288)	(0,7995 - 0,8332)	(0,7926 - 0,8268)
Acurácia	0,801300	0,812200	0,816800	0,810100
Sensibilidade	0,213530	0,183290	0,213620	0,203480
Especificidade	0,974470	0,976910	0,972190	0,974523
Precisão	0,711270	0,675210	0,664230	0,683570
Valor negativo previsto	0,807950	0,820410	0,827590	0,818650
Prevalência	0,227510	0,207510	0,204810	0,213277
Taxa de detecção total	0,048580	0,038040	0,043750	0,043457
Prevalência indicada	0,068300	0,056330	0,065870	0,063500
AUC (Área sob a curva ROC)	0,733116	0,742233	0,756774	0,744041
Kappa	0,249600	0,219100	0,248300	0,239000
P-Value	0,000797	0,013470	0,007239	0,007169

Fonte: Elaborado pelo autor, amparado nos resultados do software Rstudio (2019).

Os resultados da Regressão Logística, por meio dos seus coeficientes, estão contemplados no Apêndice E, abrangendo as 3 análises, o que permite compreender a influência de cada variável no modelo preditivo. Por exemplo, na Análise 1, verifica-se que o coeficiente negativo da variável independente “INDICADOR POSSE ROTATIVO”, que refere-se à existência de operação amparada em limite de crédito rotativo (cartão de crédito e/ou cheque especial), em perfil Dummy (1 se possui, e 0 caso contrário), indica que a existência da posse (1) diminui a probabilidade da desidratação, e esta relação é estatisticamente significativa à 1%.

4.4 REGRESSÃO LOGÍSTICA STEPWISE

Buscou-se por meio do método de seleção de atributos Stepwise melhorar a capacidade explicativa da Regressão Logística, entretanto os resultados, na maioria dos indicadores como acurácia, especificidade, valor negativo previsto, convergiram a patamares semelhantes aos observados na Regressão Logística, porém tendo havido piora em relação a sensibilidade do modelo, que apresentou 15,91% na Regressão Stepwise contra 20,34% na Regressão Logística. Foi observada uma pequena evolução em relação a precisão do modelo, o que é interessante, alcançando 70,27% na Regressão Stepwise, contra 68,35% na Regressão Logística. Além disso, o Critério de Informação de Akaike (AIC) observado na Regressão Stepwise também apresentou resultados ligeiramente melhores do que a Regressão Logística nas 3 análises efetuadas.

Em cada uma das 3 análises houve alterações no conjunto de variáveis preservadas pela seleção de atributos Stepwise, o que pode ser verificado no Apêndice E, que consolida também os coeficientes desta Regressão Logística Stepwise. A delimitação das variáveis pelo Stepwise faz com que apenas as mais importantes para o modelo permaneçam. Por si só, esta contribuição pode gerar agilidade e eficiência em outras análises, uma vez que pode subsidiar a dispensa da captura de variáveis, muitas vezes onerosas, que apresentaram baixa capacidade explicativa.

Segue Quadro 12 que apresenta a Matriz Confusão das 3 análises realizadas (A1, A2 e A3), contemplando a média da performance observada nas métricas derivadas da Matriz Confusão. Posteriormente consta o Quadro 13, que consolida os

relatórios outputs das 3 análises gerados pelo software RStudio, bem como a média observada nos respectivos indicadores de performance. O Apêndice D contempla os relatórios outputs das análises de Regressão Stepwise.

Quadro 12 - Matriz de Confusão dos resultados da Regressão Stepwise

Matriz de Confusão e métricas derivadas para avaliação dos resultados - Regressão Stepwise								
Análise referênci	Matriz de confusão	Previsto pelo Modelo		Sensibilidade	Especificidade	Precisão	Acurácia	AUC
		Retido	Desidratado					
		Retido	Desidratado					
A1	Retido	1573	33	16,702%	97,945%	70,536%	79,461%	0,71577
	Desidratado	394	79					
A2	Retido	1625	21	15,313%	98,724%	75,862%	81,416%	0,70097
	Desidratado	365	66					
A3	Retido	1617	37	15,728%	97,763%	64,423%	80,962%	0,72095
	Desidratado	359	67					
Média da performance observada				15,914%	98,144%	70,274%	80,613%	0,71256

Fonte: Elaborado pelo autor, amparado nos resultados do software RStudio (2019).

Quadro 13 - Consolidação dos resultados analíticos da Regressão Stepwise

Técnica de análise	Regressão Stepwise			
	A1	A2	A3	Média
Análise de Referência				
Intervalo de confiança (95%)	(0,7766 - 0,8118)	(0,7967 - 0,8307)	(0,7921 - 0,8263)	(0,7885 - 0,8229)
Acurácia	0,794600	0,814200	0,809600	0,806133
Sensibilidade	0,167020	0,153130	0,157280	0,159143
Especificidade	0,979450	0,987240	0,977630	0,981440
Precisão	0,705360	0,758620	0,644230	0,702737
Valor negativo previsto	0,799690	0,816580	0,818320	0,811530
Prevalência	0,227510	0,207510	0,204810	0,213277
Taxa de detecção total	0,038000	0,031780	0,032210	0,033997
Prevalência indicada	0,053870	0,041890	0,050000	0,048587
AUC (Área sob a curva ROC)	0,715770	0,700965	0,720953	0,712563
Kappa	0,200400	0,199000	0,187500	0,195633
P-Value	0,008115	0,007454	0,053530	0,023033

Fonte: Elaborado pelo autor, amparado nos resultados do software RStudio (2019).

4.5 RANDOM FORESTS

Os resultados verificados nas análises desenvolvidas com a técnica Random Forests demonstraram menor capacidade explicativa do que as abordagens de Regressão Logística, nas três análises de referência. Entretanto, verifica-se que sua performance evoluiu significativamente nas análises 2 e 3. A etapa de testes dos algoritmos, contempla automaticamente a avaliação do número de atributos de

entrada (testes lógicos para selecionar conjuntos), tendo sido escolhido pelo modelo conjuntos de 10 atributos (A1 e A3) e 20 (A2). Este e outros detalhes constam nos relatórios/outputs dispostos no Apêndice F.

Segue Quadro 14 que apresenta a Matriz Confusão das 3 análises realizadas (A1, A2 e A3), além da média da performance observada nas métricas derivadas da Matriz Confusão.

Quadro 14 - Matriz de Confusão dos resultados da Random Forests

Matriz de Confusão e métricas derivadas para avaliação dos resultados - Random Forests								
Análise referência	Matriz de confusão	Previsto pelo Modelo		Sensibilidade	Especificidade	Precisão	Acurácia	AUC
		Retido	Desidratado					
A1	Retido	391	10	11,017%	97,506%	56,522%	77,842%	0,54262
	Desidratado	105	13					
A2	Situação real	Retido	390	18,692%	94,891%	48,780%	79,151%	0,56791
		Desidratado	87					
A3	Retido	401	12	17,925%	97,094%	61,290%	80,925%	0,57509
	Desidratado	87	19					
Média da performance observada				15,878%	96,497%	55,531%	79,306%	0,56187

Fonte: Elaborado pelo autor, amparado nos resultados do software RStudio (2019).

A seguir consta o Quadro 15, que consolida os relatórios outputs das 3 análises gerados pelo software RStudio, bem como a média observada nos respectivos indicadores de performance.

Quadro 15 - Consolidação dos resultados analíticos da Random Forests

Técnica de análise	Random Forests			
Análise de Referência	A1	A2	A3	Média
Intervalo de confiança (95%)	(0,7402 - 0,8134)	(0,7539 - 0,8257)	(0,7728 - 0,8422)	(0,7556 - 0,8271)
Acurácia	0,778400	0,791500	0,809200	0,793033
Sensibilidade	0,110170	0,186920	0,179250	0,158780
Especificidade	0,975060	0,948910	0,970940	0,964970
Precisão	0,565220	0,487800	0,612900	0,555307
Valor negativo previsto	0,788310	0,817610	0,821720	0,809213
Prevalência	0,227360	0,206560	0,204240	0,212720
Taxa de detecção total	0,025050	0,038610	0,036610	0,033423
Prevalência indicada	0,044320	0,079150	0,059730	0,061067
AUC (Área sob a curva ROC)	0,542616	0,567911	0,575095	0,561874
Kappa	0,119100	0,176000	0,203800	0,166300
P-Value	0,400200	0,568700	0,241200	0,403367

Fonte: Elaborado pelo autor, amparado nos resultados do software RStudio (2019).

4.6 REDES NEURAIAS

A quarta modelagem desenvolvida foi através dos algoritmos de aprendizagem profunda das Redes Neurais. Trata-se de uma técnica que contempla rotinas automatizadas de inteligência artificial, através da metodologia homologada Keras¹¹, que abrange as etapas de pré-tratamento de dados até às definições de seleção de atributos e análise de resultados.

Antes de iniciar a análise, a base de dados foi separada em 2 grupos, para treino e validação. Esta é uma etapa essencial nesta metodologia, de forma a mitigar que o aprendizado fosse influenciado, o que faria ter um modelo cujos acertos convergissem a 100% no grupo de teste, sem lastro para validação. O Quadro 16 que apresenta a Matriz Confusão das 3 análises realizadas (A1, A2 e A3), além da média da performance observada nas métricas derivadas da Matriz Confusão.

Quadro 16 - Matriz de Confusão dos resultados das Redes Neurais Artificiais

Matriz de Confusão e métricas derivadas para avaliação dos resultados - Redes Neurais								
Análise referênci	Matriz de confusão	Previsto pelo Modelo		Sensibilidade	Especificidade	Precisão	Acurácia	AUC
		Retido	Desidratado					
		Retido	Desidratado					
A1	Retido	458	0	0,000%	100,000%	0,000%	88,247%	0,50075
	Desidratado	61	0					
A2	Retido	466	0	0,000%	100,000%	0,000%	89,788%	0,50972
	Desidratado	53	0					
A3	Retido	470	0	0,000%	100,000%	0,000%	90,385%	0,53136
	Desidratado	50	0					
Média da performance observada				0,000%	100,000%	0,000%	89,473%	0,51394

Fonte: Elaborado pelo autor (2019).

A seguir consta o Quadro 17, que consolida os relatórios outputs das 3 análises gerados pelo software RStudio, bem como a média observada nos respectivos indicadores de performance.

¹¹ Keras: Biblioteca de rede neural com algoritmos de código aberto. A abordagem utilizada apresentou 82% de acurácia em análise preditiva de churn, constando suas etapas no endereço eletrônico abaixo: <https://www.business-science.io/business/2017/11/28/customer_churn_analysis_keras.html>.

Quadro 17 - Consolidação dos resultados analíticos das Redes Neurais Artificiais

Técnica de análise	Redes Neurais Artificiais			
	A1	A2	A3	Média
Análise de Referência				
Intervalo de confiança (95%)	(*)	(*)	(*)	(*)
Acurácia	0,882466	0,897881	0,903846	0,894731
Sensibilidade	0,000000	0,000000	0,000000	0,000000
Especificidade	1,000000	1,000000	1,000000	1,000000
Precisão	Nula	Nula	Nula	Nula
Valor negativo previsto	0,882466	0,897881	0,903846	0,894731
Prevalência	0,117534	0,102119	0,096154	0,105269
Taxa de detecção total	0,000000	0,000000	0,000000	0,000000
Prevalência indicada	0,000000	0,000000	0,000000	0,000000
AUC (Área sob a curva ROC)	0,500752	0,509717	0,531362	0,513944
Kappa	0,000000	0,000000	0,000000	0,000000
P-Value	(*)	(*)	(*)	(*)

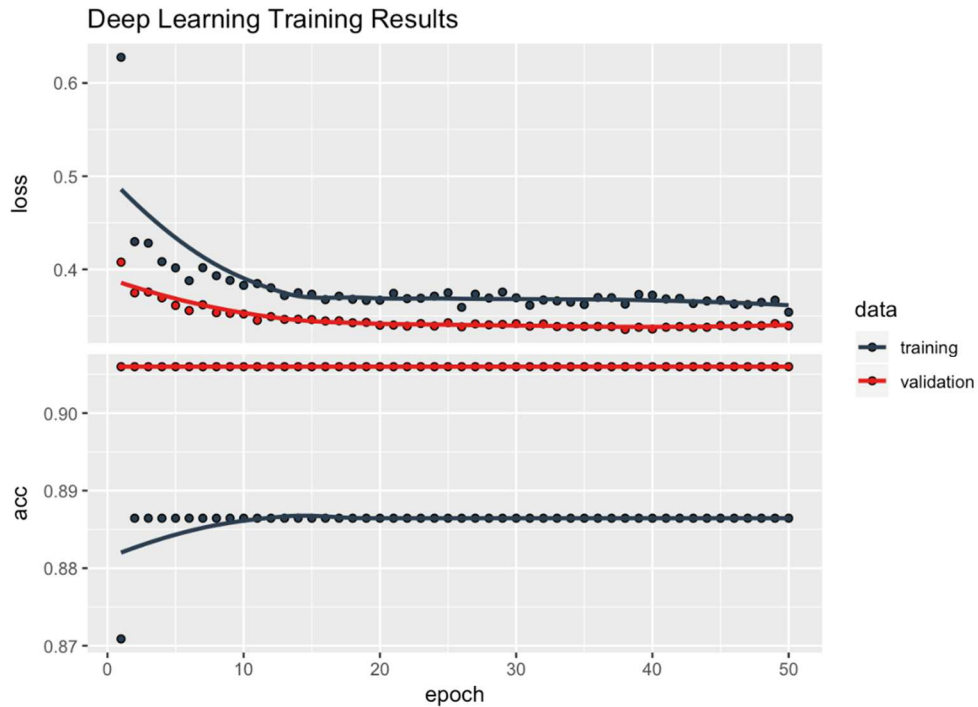
(*) não calculados pelo software

Fonte: Elaborado pelo autor, amparado nos resultados do software RStudio (2019).

Nos resultados observados na Análise A1, por exemplo, conforme Quadro 16, verificou-se uma acurácia de 88,247% e na etapa de validação alcançou 89,58%, entretanto a previsão do modelo indicou 100% de clientes retidos. A métrica Área sob a Curva ROC apresentou média de 0,5134 nas 3 análises, o que significa que há 51,34% do modelo classificar corretamente cada cliente, o que expressa um resultado insignificante quando os resultados possíveis são dicotômicos. Em relação a estes resultados obtidos com as Redes Neurais, observamos um índice de acurácia que alcança até 90% na Análise 3, sendo o maior dentre todas as abordagens realizadas, porém trata-se de um resultado absolutamente inócuo. Ressaltados os reconhecidos problemas de balanceamento de classes, veja que, ainda nesta Análise 3, conforme Prevalência, 9,61% do grupo de teste são considerados os verdadeiros “desidratados” e o restante (90,39%) são considerados os verdadeiros “retido”. Logo, tendo as Redes Neurais apresentado resultado que indica que 100% dos clientes serão retidos, sua acurácia efetivamente converge a esse patamar, convergindo a 90% de acerto nas previsões, porém este resultado não produz as informações que buscamos: quais sejam identificar os clientes que tendem a evadir (sensibilidade) e qual a precisão do modelo.

A demonstração gráfica a seguir, conforme Figura 12, demonstra a evolução do aprendizado do modelo no decorrer dos testes da Análise A3,

Figura 14 - Representação gráfica do resultado A3 das Redes Neurais



Fonte: Elaborado pelo autor, amparado nos resultados do software RStudio (2019).

Considerando que os resultados das Redes Neurais não alcançaram as expectativas, foi efetuada revisão de todos os parâmetros de análise. O software de trabalho RStudio não acusou nenhuma falha de rotina ou de processo durante a condução das Redes Neurais. Podemos considerar que os parâmetros dos dados já estavam validados no software, uma vez que subsidiaram as demais técnicas preditivas previamente desenvolvidas. Ponderado o nível de balanceamento de classes, acreditamos que o restrito tamanho da amostra, por sua heterogeneidade, foi o fator determinante para estes resultados. Em outras condições de análise acreditamos que os resultados devem evoluir.

4.7 SÍNTESE DOS RESULTADOS

O quadro abaixo apresenta a consolidação dos resultados médios observados nos principais indicadores das metodologias da Regressão Logística, Regressão Stepwise, Random Forests e Redes Neurais Artificiais.

Quadro 18 - Consolidação dos resultados analíticos das modelagens preditivas

Técnica de análise	Regressão Logística	Regressão Stepwise	Random Forests	Redes Neurais Artificiais
Análise de Referência	Média	Média	Média	Média
Intervalo de confiança (95%)	(0,7926 - 0,8268)	(0,7885 - 0,8229)	(0,7556 - 0,8271)	(*)
Acurácia	0,810100	0,806133	0,793033	0,894731
Sensibilidade	0,203480	0,159143	0,158780	0,000000
Especificidade	0,974523	0,981440	0,964970	1,000000
Precisão	0,683570	0,702737	0,555307	Nula
Valor negativo previsto	0,818650	0,811530	0,809213	0,894731
Prevalência	0,213277	0,213277	0,212720	0,105269
Taxa de detecção total	0,043457	0,033997	0,033423	0,000000
Prevalência indicada	0,063500	0,048587	0,061067	0,000000
AUC (Área sob a curva ROC)	0,744041	0,712563	0,561874	0,513944
Kappa	0,239000	0,195633	0,166300	0,000000
P-Value	0,007169	0,023033	0,403367	(*)

(*) não calculados pelo software

Fonte: Elaborado pelo autor, amparado nos resultados do software RStudio (2019).

Uma vez que este trabalho teve por objetivo identificar quais clientes tendem ao *churn*, notadamente, temos como principal métrica de avaliação dos modelos a Sensibilidade, pois esta expressa a quantidade de positivos corretamente classificados pelo modelo. A esta perspectiva, a regressão logística alcançou 20,34%, enquanto a Regressão Stepwise 15,91% e Random Forest 15,87%. Para a Regressão Logística, significa dizer, praticamente, que o modelo consegue identificar 1 de cada 5 clientes que irão desidratar.

Dos clientes previstos que irão desidratar, a Regressão Stepwise apresentou o melhor Precisão, com 70,27% de acerto. Se, por exemplo, for estruturada uma ação de retenção preventiva destinada aos clientes previstos para desidratar pelo modelo, 7 de cada 10 clientes selecionados efetivamente desidratam. Os outros 3 de cada 10 são os Falsos Positivos, erroneamente considerados desidratados, seriam abrangidos

pela ação de retenção sem necessidade, uma vez que o modelo já indica sua retenção, independente de quaisquer novos incentivos/benefícios. Logo, quanto maior a Precisão, menor será o nível de desperdício.

5 CONCLUSÕES

Neste capítulo serão apresentadas as conclusões observadas no decorrer desta pesquisa. Registramos inicialmente o cumprimento do objetivo geral desta pesquisa, bem como dos demais objetivos específicos definidos. Além disso, o desenvolvimento dos trabalhos proporcionou a apropriação e geração de conhecimento, em perspectivas teórica e gerencial, conforme apresentamos a seguir:

5.1 IMPLICAÇÕES TEÓRICAS

O referencial teórico fomenta a necessidade de as empresas atuarem voltadas para o cliente (GUPTA; LEHMANN, 2005; KUMAR, 2018). A este propósito, compreender os estágios de relacionamento, com métodos e processos que permitam seu acompanhamento e gerenciamento se faz necessário (PAYNE, 2006; NITZAN; LIBAI (2011). A perda de clientes é a principal barreira para uma empresa elevar sua rentabilidade (CHENEY, 2016; RUST; ZAHARIK, 1993; REICHHELD; SASSER, 1990). Prever quais clientes tendem a migrar para concorrentes, com o propósito de agir para evitar essa situação é questão que pode ser solucionada através de metodologias de análise preditiva, permitindo uma gestão proativa por parte das organizações (KHAKABI; GHOLAMIAN; NAMVAR, 2010; VANDECRUYS et al., 2008). Exceto Botelho e Tostes (2010), que teve por contexto uma operadora de cartões de crédito, são escassos os registros teóricos sobre abordagens preditivas em instituições financeiras brasileiras e, de uma forma geral, nas empresas brasileiras, o que, dentre outros fatores, pode ser característico de uma cultura incipiente de abordagens preditivas como suporte a retenção.

Em relação as definições de método, a construção da variável dependente mostrou-se assertiva e adequada para caracterização do *churn*, comprovado pelo comportamento dos diferentes grupos através Teste Kolmogorov-Smirnov, conforme descrito na seção 4.1.1.

O entendimento apropriado para consecução desta fase somente foi possível a partir das interações constantes da etapa qualitativa, que permitiu a ampliação de entendimentos sobre o processo de *churn* e o mapeamento de variáveis consideradas predictoras, em aderência ao pipeline de coleta de dados proposto por Subramanya e

Somani (2017), contemplando sete categorias de variáveis, gerando maior eficácia aos algoritmos de análise, o que contribui para formação de conhecimento na área.

A revisão das métricas de avaliação de modelos de classificação, ponderado o foco de pesquisa em identificar os verdadeiros positivos, nos permite manifestar entendimento dissonante a grande maioria do referencial teórico, conforme Castro e Braga (2011), que elenca a Acurácia como principal indicador para avaliação de modelos de classificação. A métrica prioritária deve ser a Sensibilidade, que reflete o quanto o modelo é eficaz para identificar corretamente os positivos. Porém, ao considerarmos que os modelos se destinam a identificação de clientes que irão compor campanhas de retenção, por exemplo, a escolha do modelo também deve se subordinar ao desempenho do indicador de Precisão, de forma a mitigar desperdícios.

Em relação a etapa quantitativa, a síntese comparativa das análises, que consolidou os resultados analíticos observados nas abordagens de Regressão Logística, Regressão Stepwise e Random Forests, também evidencia algumas contribuições. A seleção de atributos Stepwise agregou melhora de Precisão do modelo, entretanto não conseguiu o mesmo patamar de Sensibilidade observada na Regressão Logística, que se evidenciou como o melhor modelo dos que foram desenvolvidos. Acreditamos que em bases mais robustas de dados, que permitam ultrapassar as limitações ponderadas em relação a quantidade de clientes da amostra e a heterogeneidade verificada, provavelmente a performance de todos os modelos deve evoluir, notadamente os algoritmos de inteligência artificial de Random Forests e Redes Neurais, aderentes aos níveis de resultados observados por Veloso (2013), por exemplo.

5.2 IMPLICAÇÕES PRÁTICAS

Ponderado o perfil exploratório deste trabalho, uma vez que não se parte de um modelo teórico a ser testado, os resultados alcançados podem ser considerados satisfatórios. A condução desta pesquisa evidencia que realizar análises preditivas é algo factível, e estas abordagens podem ser desenvolvidas em empresas de diferentes segmentos e portes. A esta finalidade, evidencia-se a importância de as organizações possuírem um Banco de Dados organizacional que contemple um adequado nível de informação (dados) em nível individual. A diretriz elaborada na

etapa qualitativa, com proposta de sugestão de variáveis que possam subsidiar análises preditivas, pode ser uma referência para muitas empresas e organizações mapearem, coletarem e armazenarem outros dados relevantes a este fim.

Ponderadas as limitações elencadas na seção 5.3, notadamente em relação a heterogeneidade e tamanho da amostra, consideramos satisfatórios os resultados observados nas análises preditivas, que alcançaram um nível de sensibilidade superior a 20%. Para uma empresa como o Banco do Brasil, conforme informações apresentadas na seção 1.4, o cancelamento de negócios por abandono de clientes pode representar até R\$7 bilhões em um ano. Sobre este patamar, o modelo desenvolvido pode ser capaz de prever, em nível de cliente, perdas de negócios que alcançam o montante de R\$1,4 bilhão. A precisão da referida modelagem, no patamar de 68%, não configura um nível de desperdício excessivamente oneroso. Estas referências ganham proporções maiores sob as abordagens do *Customer Equity* e *Customer Lifetime Value*. Notadamente, são valores muito expressivos, o que evidencia a importância do desenvolvimento de estratégias de retenção, aderentes a Neslin et al. (2006), que subordina as campanhas de retenção ao efetivo retorno financeiro. Por exemplo, retomando o contexto bancário, as vezes o incentivo/benefício necessário para reter um cliente pode ser apenas uma flexibilização de tarifas (da conta ou cartão de crédito), o que pode configurar reduzidos custos com ação de retenção, porém com grande potencial de retorno.

5.3 LIMITAÇÕES DO MÉTODO

Os reconhecidos problemas de balanceamento de classes, caracterizado quando há densidades altamente desequilibradas, sendo a classe relativa aos eventos do *churn* muito pouco representada, normalmente são tratados a partir de metodologias de pré-processamento como a PSM (Propensity Score Matching). King e Nielsen (2016) ponderam que essas abordagens de pré-processamento frequentemente alcançam o oposto do objetivo pretendido, aumentando o desequilíbrio e a ineficiência. Também ressaltam que a fraqueza do PSM vem de suas tentativas de aproximar um experimento completamente randomizado, portanto, cego para a grande parte do desequilíbrio. Reconhecemos a limitação do tamanho da base

de dados, porém, em consonância a essa perspectiva, não foi efetuado pré-processamento.

Nem sempre a redução do volume de negócios é decorrente de *churn*. Em muitos casos, a redução ou cancelamento de produtos decorrem da perda do interesse/necessidade por parte do cliente. Há clientes que resgatam seus investimentos para investir na aquisição de um imóvel, por exemplo, outros que liquidam operações de crédito por não mais necessitarem. Em nível de análise quantitativa, com os dados coletados, não vislumbramos formas de distinção destas ocorrências, o que configura, também, uma limitação do método.

Uma limitação deste estudo refere-se à coleta de dados e ao padrão dos dados constantes do banco de dados do Banco do Brasil. Em face desta limitação, variáveis importantes deixaram de ser consideradas. O conjunto de variáveis não contempla, por exemplo, indicadores capazes de sinalizar o nível de satisfação, tampouco situações de atrito. Também, ainda que a desidratação possa ser considerada algo abstrato, inexistia variável com capacidade de ser marcadora deste fenômeno, exigindo que a variável dependente tivesse de ser desenvolvida.

O tamanho da amostra recebida representa limitação às análises. A partir da definição da variável dependente, foi observada uma taxa de Prevalência de 21%. Dentre os pouco mais de 2000 clientes que permaneceram em cada, portanto, somente cerca de 450 casos configuravam de desidratação. Ainda que 450 não necessariamente pode ser considerado um número inexpressivo, seria desejável um número superior, ponderada a quantidade de variáveis explicativas, o que certamente iria elevar a capacidade preditiva dos modelos.

A Prevalência é uma métrica que indica a quantidade total de casos observados em uma amostra. Pelos parâmetros acima, a prevalência observada com a condição desidratado alcançou cerca de 21% da amostra. Por entender que a desidratação é um fenômeno mais comum do que o *churn* total, foi considerada que a referida calibração é adequada aos objetivos propostos.

Ainda que as metodologias e técnicas possam ser replicadas em outras organizações e segmentos, inclusive não financeiras, o modelo desenvolvido não pode ser considerado generalizável, ressaltadas, ainda, as demais delimitações dispostas na seção 1.3.

5.4 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

A abordagem desta pesquisa, ao amparo do *Customer Value Theory* (KUMAR, 2018) e *Customer Lifetime Value* (GUPTA; LEHMANN, 2005), não caracteriza esgotamento, pelo contrário, muitas são as frentes necessárias para progredir nas organizações a uma cultura centrada no cliente. Ainda que as análises preditivas permitam identificar os clientes que tendem ao fenômeno do *churn*, além de subsidiar campanhas de retenção, acredito que deve ser buscado maior conhecimento sobre os clientes que estão em risco. Acredito que abordagens em nível individual podem contribuir com elementos importantes. É desejável, ainda, que as empresas tenham conhecimento sobre o Valor Vitalício (CLV) de todos os seus clientes.

Em relação a metodologia desenvolvida, podem ser buscadas outras configurações de variáveis explicativas, e notadamente outras proposições para variável dependente, o que pode gerar diferentes resultados. Amostras maiores devem ser testadas sobre sua capacidade de agregar melhor resultados, pois não vislumbramos um nível de saturação, pelo contrário. A replicação do modelo em segmentos/nichos/clusters mais homogêneos pode gerar análises mais consistentes, com maior capacidades explicativas, e maior sensibilidade. A replicação deste modelo em outros contextos também pode oferecer resultados distintos. Também é desejável realizar análises nos grupamentos de clientes mais rentáveis, retomando a perspectiva de Pareto.

Em relação ao desenvolvimento de estratégias de retenção, aderente a Neslin et al. (2006), cabe avaliar a performance e retorno observado nas campanhas de retenção, através de indicadores como CLV, custo de retenção e recursos disponíveis (humanos e financeiros). Para alcançar maior efetividade nas ações de retenção, ressalta-se a importância de melhor conhecer os atributos valorizados pelos clientes, o que gera valor a cada perfil. Portanto, é desejável colher diretamente dos clientes suas percepções sobre benefícios/incentivos a serem ofertados nas ações de retenção, que não precisam estar pautadas somente por perspectivas financeiras, conforme o exposto por Burez e Van Den Poel (2007), que evidenciou a melhor performance de retenção, dentre 4 grupos, foi com o grupo que foi convidado a responder uma pesquisa de satisfação, superando 2 grupos que receberam benefícios financeiros.

REFERÊNCIAS

- ABDOLVAND, N; BARADARAN, V; ALBADVI, A. Activity-level as a link between customer retention and consumer lifetime value. **Iranian Journal of Management Studies**, v. 8, n. 4, p. 567-587, Oct. 2015.
- ADEBIYI, S. O.; OYATOYE, E. O.; AMOLE, B. B. Improved Customer Churn and Retention Decision Management Using Operations Research Approach. **EMAJ: Emerging Markets Journal**, [s. l.], v. 6, n. 2, p. 12–21, 2016.
- AGRESTI, A.; FINLAY, B; **Métodos estatísticos para as ciências sociais**. 4. ed. Porto Alegre: Penso, 2017.
- BACEN; **Estatísticas sobre o sistema financeiro**. Disponível em: <https://www.bcb.gov.br/#itemestatisticassistemafinanceiro>
- BANCO DO BRASIL. **Sumário do Resultado Anual 2018**. Brasília, 2019.
- BARNEY, Jay. Firm Resources and Sustained Competitive Advantage. **Journal of Management**. v. 17, n. 1, p. 99 - 120.
- BARRETO, I.F.; CRESCITELLI, E.; FIGUEIREDO, J.B. Resultados de marketing de relacionamento: proposição de modelo por meio de mapeamento cognitivo. **Revista Brasileira de Gestão de Negócios**, v. 17, n. 58, p. 1371-1389, Oct. 2015.
- BERGER, P; NASR, N. Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing*. v. 12, n. 1, p. 17-30, Jan. 1998.
- BERRY, M.; LINOFF, G. **Data mining techniques**: for marketing, sales, and customer relationship management. 3. ed. Indianapolis: Wiley Publishing Inc, 2011.
- BRAUN, M.; SCHWEIDEL, D. A. Modeling Customer Lifetimes with Multiple Causes of Churn. **Marketing Science**, [s. l.], v. 30, n. 5, p. 881–902, 2011.
- BOTELHO, Delane; TOSTES, Frederico Damian. Modelagem de probabilidade de churn. **Revista de Administração de Empresas**, São Paulo, v. 4, n. 396, 2010.
- BUCKINX, W.; VAN DEN POEL, D. Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. **European Journal Of Operational Research**, v. 164, n. 1, p. 252–268, 2005.
- BUREZ, J.; VAN DEN POEL, D. **CRM at a pay-TV company**: using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 2007.
- _____.; _____. Separating financial from commercial customer churn: a modeling step towards resolving the conflict between the sales and credit department. **Expert Systems with Applications**, v. 35, p. 497-514, 2008.

_____.; _____. Handling class imbalance in customer churn prediction. **Expert Systems with Applications**, v. 36, n. 3, p. 4626-4636, 2009.

CASTRO, C.L. de; BRAGA, A. P; Supervised learning with imbalanced data sets: an overview. **Sba: Controle & Automação Sociedade Brasileira de Automatica**, Campinas, SP, n. 5, p. 441, 2011.

ÇAVDAR, A. B.; FERHATOSMANOĞLU, N. Airline customer lifetime value estimation using data analytics supported by social network information. **Journal of Air Transport Management**, [s. l.], v. 67, p. 19–33, 2018.

CHENEY, C. **Understanding Churn: Why Do Customers Unsubscribe?** Promotional Marketing, [s. l.], p. 1, 2016. Disponível em: <<https://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=118268709&lang=pt-br&site=eds-live>>. Acesso em: 30 jun. 2018.

CIRILLO, A. **RStudio for R Statistical Computing Cookbook**. Birmingham, UK: Packt Publishing, 2016.

COUSSEMENT, K.; VAN DEN POEL, D. Churn prediction in subscription services: an application of support vector machines while comparing two parameter-selection techniques. **Expert Systems with Applications**, n. 34, p. 313–327, 2008.

DATTA, P.; MASAND, B.; MANI, D.R.; LI, B. Automated cellular modeling and prediction on a large scale. *Issues on the Application of Data Mining 2001*; p. 485–502.

DEBNATH, R.; DATTA, B.; MUKHOPADHYAY, S. Customer Relationship Management Theory and Research in the New Millennium: Directions for Future Research. **Journal of Relationship Marketing**, v. 15, n. 4, p. 299-325, Oct. 2016.

DEPREN, Serpil Kilic. Churn propensity model for customers who made a complaint in retail banking. **European Journal of Business and Social Sciences**, 2018.

DWYER, F.R. Customer Lifetime Valuation to Support Marketing Decision Making. **Journal of Direct Marketing**, v. 3, n. 4, p. 8-15, Sept. 1989.

FADER, P.; HARDIE, B. Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. **Marketing Science**, v. 29, n. 1, 85-93, Jan. 1, 2010.

FADER, P.; HARDIE, B.; LEE, K. RFM and CLV: Using iso-value curves for customer base analysis. **Journal of Marketing Research**, v. 42, n. 4, p. 415-430, Nov. 1, 2005.

FARQUAD, M.; RAVI, V.; RAJU, S.B. Churn prediction using comprehensible support vector machine: An analytical CRM application. **Applied Soft Computing Journal**, n. 19, p. 31-40, June 1, 2014.

FINTECHLAB. **Radar Fintechlab 2019**. Disponível em: <<https://fintechlab.com.br/index.php/2019/06/12/8a-edicao-do-radar-fintechlab-registra-mais-de-600-iniciativas/>>. Acesso em: 30 jun. 2019.

GHORBANI, A; TAGHIYAREH, F. CMF: A framework to improve the management of customer churn. **2009 IEEE Asia-Pacific Services Computing Conference (APSCC)**, p. 457, 2009.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GLADY, N.; BAESENS, B.; CROUX, C. Modeling churn using customer lifetime value. **European Journal of Operational Research**, v. 197, n. 1, p.402–411, Aug. 2009.

GRÖNROOS, C. Relationship marketing: the nordic school perspective. In: **Handbook of Relationship Marketing**. Sage Publications. Thousand Oaks, 2000. p. 95-117.

GUIMARÃES, Cezar T.; MILANI, Gabriela N.F.; CRESCITELLI, Edson. Aplicação de ações de marketing de relacionamento em lojas de departamento. **VIII SEMEAD**, Anais, São Paulo, 2005.

GUPTA, S.; LEHMANN, D.R. **Managing Customers as Investments: The Strategic Value of Customers in the Long Run**. NJ, Upple Saddle River: Wharton School Publishing, Pearson Education, Inc.. 2005.

GUPTA, S.; LEHMANN, D.R. Customer Lifetime Value and Firm Valuation. **Journal of Relationship Marketing**. v. 5, n. 2/3, p. 87-110, June 2006.

GÜR ALI, Ö.; ARITÜRK, U. Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. **Expert Systems With Applications**, n. 41, p. 7889-7903, Dec. 1, 2014.

HADDEN, J. et al. Computer assisted customer churn management: State-of-the-art and future trends. **Computers and Operations Research**, n. 34, p. 2902-2917, Jan. 1, 2007.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. New York: Springer, 2009.

HUGHES, A. **Database marketing estrategico**. São Paulo: Makron, 1998.

HUNG, S., YEN, D.; WANG, H. Applying data mining to telecom churn management. **Expert Systems with Applications**, n. 31, p. 515–524, 2006.

HUNT, Shelby D.; MORGAN, Robert M. The Comparative Advantage Theory of Competition. **Journal of Marketing**, v. 59, n. 2, p. 1–15, 1995.

HWANG, H., JUNG, T.; SUH, E. An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. **Expert Systems with Applications**, n. 26, p. 181–188, 2004.

JAHANZEB, S.; JABEEN, S. Churn management in the telecom industry of Pakistan: A comparative study of Ufone and Telenor. **Journal of Database Marketing & Customer Strategy Management**, v. 14, n. 2, p. 120-129, Jan. 2007.

KHAKABI, S.; GHOLAMIAN, M.R.; NAMVAR, M.; Data Mining Applications in Customer Churn Management. 2010. **International Conference on Intelligent Systems, Modelling and Simulation**. Intelligent Systems, Modelling and Simulation (ISMS), 220, 2010.

KING, G.; NIELSEN, R. Why Propensity Scores Should Not Be Used for Matching. **Political Analysis**, v. 1, n. 20, May 2019.

KOTLER, P. Philip Kotler explores the new marketing paradigm. **Marketing Science Institute Review**, n. 1, p. 4-5, 1991.

KOTLER, P; Armstrong, G. Principles of Marketing. **Journal of Marketing Management**, v. 7, n. 4, p. 428–429, 1991.

KOTLER, P; KELLER, K. **Administração de Marketing**. 14. ed. São Paulo: Pearson, 2013.

KUMAR, V.; VENKATESAN, R., BOHLING, T; BECKMANN, D. The Power of CLV: Managing Customer Lifetime Value at IBM. **Marketing Science**, v. 27, n. 4, p. 585-599, July 2008.

KUMAR, D.; RAVI, V. Predicting credit card customer churn in banks using data mining. **International Journal of Data Analysis Techniques and Strategies**, v. 1, n. 1, p. 4–28, 2008.

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

KUMAR, V.; REINARTZ, W. **Customer Relationship Management: Concept, Strategy, and Tools**. Berlin: Springer-Verlag, 2012.

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

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.; POZZA, I.; GANESH, J. Revisiting the satisfaction-loyalty relationship: Empirical generalizations and directions for future research. **Journal of Retailing**, v. 89, n. 3, p. 246-262, Sept. 1, 2013.

KURTZ, D.L.; CLOW, K.E. **Services marketing**. New York: John Wiley & Sons, 1998., 1998.

LARIVIÈRE, B.; VAN DEN POEL, D. Predicting customer retention and profitability by using random forest and regression forest techniques. **Expert Systems with Applications**, v. 29, n. 2, p. 472–484, 2005.

LEE, Y; IDRIS, A; KHAN, A; Genetic Programming and Adaboosting based churn prediction for Telecom. **2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Systems, Man, and Cybernetics (SMC)**, 2012.

LEJEUNE, Miguel A.P.M. Measuring the impact of data mining on churn management. **Internet Research**, v. 11, n. 5, p. 375 – 387, 2011.

LEMMENS, A.; CROUX, C. Bagging and boosting classification trees to predict churn. **Journal of Marketing Research**, 2006.

LIÉBANA-CABANILLAS, F. et al. The moderating effect of user experience on satisfaction with electronic banking: empirical evidence from the Spanish case. **Information Systems & e-Business Management**, v. 14, n. 1, p. 141-165, Feb. 2016.

LIMA, E., MUES, C.; BAESENS, B. Domain knowledge integration in data mining using decision tables: Case studies in churn prediction. **Journal of the Operational Research Society**, n. 60, p. 1096–1106, 2009.

LIN, N. et al. Making customer relationship management work: evidence from the banking industry in Taiwan. **Service Industries Journal**, v. 29, n. 9, p. 1183-1197, Oct. 2009.

LUQUE, A. et al. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. **Pattern Recognition**, [s. l.], v. 91, p. 216–231, 2019.

MARROQUÍN, Juan Carlos. Novo consumidor: mais informado, mais exigente, mais consciente. **Folha de São Paulo**, São Paulo, 27 out. 2014. Disponível em: <<http://www1.folha.uol.com.br/topofmind/2014/10/1528719-novo-consumidor-mais-informado-mais-exigente-mais-consciente.shtml>>. Acesso em: 30 jun. 2018.

MCAULEY, Daniel. **What is FinTech?**. Wharton FinTech, 2015. Disponível em: <<https://medium.com/wharton-fintech/what-is-fintech-77d3d5a3e677>>. Acesso em: 17 de maio de 2018.

MELLO, S.B. de; LEÃO, A.M. de Souza. Uma Revisão Crítica sobre Uso de Diferentes Conceitos de Valor no Marketing: A Critical Review of Different Value Concepts Use in Market Marketing. **RAC – Eletrônica**, v. 2, n. 1, p. 37-53, Jan. 2008.

MUTANEN, T.; NOUSIAINEN, S.; AHOLA, J. Customer churn prediction: a case study in retail banking. 218. **IOS Press**, v. 77, n. 83, p. 218, Jan. 1, 2010.

NESLIN, S.; GUPTA, S.; KAMAKURA, W.; JUNXIANG, L.; MASON, C. Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. **Journal Of Marketing Research (JMR)**, v. 43, n. 2, p. 204-211, May 2006.

NITZAN, I; LIBAI, B. Social Effects on Customer Retention. **Journal of Marketing**, v. 75, n. 6, p. 24-38, Nov. 2011.

OLIVER, R. L. **Satisfaction: a behavioral perspective on the consumer** 2. ed. New York: ME Sharp, 2010.

OLSEN, T.L.; PARKER, R.P. Inventory management under Market size dynamics. **Management Sci**, v. 54, n. 10, p. 1805–1821, 2008.

PARETO, Vilfredo. The Pareto Principle (the 80/20 principle). **Guide to Management Ideas**, [s. l.], p. 168–169, 2003.

PAYNE, A. **Handbook of CRM: achieving excellence in customer management**. Oxford: Elsevier Butterworth-Heinemann, 2006.

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

PEPPERS, D.; ROGERS, M. Making the transition to one-to-one. **Marketing Inc.** v. 19, n. 1, p. 63, Jan. 1997.

RAFIQUE, M.; EVANS, R. D.; NAWAZ, M. T. Absorptive capacity: A hub of blue ocean and red ocean strategies and capability transformation in innovative business environments. 2nd **International Conference on Knowledge-Based Engineering and Innovation (KBEI)**, 2015.

R CORE TEAM (2019). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<https://www.R-project.org/>>. Acesso em: 15 jul. 2019.

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

REICHHELD, F.F.; SASSER JR., W.E. Zero defections: Quality comes to services. **Harvard Business Review**, v. 68, n. 5, p. 105-111, Sept. 1990.

ROGERS, David L. **The digital transformation playbook**. Columbia: Columbia Business School, 2016.

RStudio Team (2015). **RStudio: Integrated Development for R**. RStudio. Boston, MA. Disponível em: <<http://www.rstudio.com/>>. Acesso em: 15 jul. 2019.

RUST, R.T.; ZAHORIK, A.J. Customer satisfaction, customer retention, and market share. **Journal of Retailing**, v. 69, n. 2, p.193-215, 1993.

RUST, R; LEMON, K; ZEITHAML, V; Return on Marketing: Using Customer Equity to Focus Marketing Strategy. **Journal of Marketing**, v. 68, n. 1, p. 109-127, 2004.

RUST, R.T.; ZEITHAML, V.A.; LEMON, K.N. **Driving Customer Equity**: how customer lifetime value is reshaping corporate strategy. Porto Alegre: Bookman, 2001.

SAMPIEIRI, Hernandez Roberto; COLLADO, Carlos Fernández; BAPTISTA LUCIO, María Pilar. **Metodologia de Pesquisa**, 5. ed. Porto Alegre: Penso, 2013.

SANTOS, P.M. Finazzi; PORTO, R. Barreiros. Responsabilidade ambiental e lealdade de clientes em banco de varejo. : Environmental responsibility and customers loyalty in retail bank. **RAE: Revista de Administração de Empresas**, v. 54, n. 6, p. 606-618, Nov. 2014.

SHAH, Denish; KUMAR, V.; KIM, Kihyun Hannah; CHOI, Jeewon Brianna. Linking Customer Behaviors to Cash Flow Level and Volatility: Implications for Marketing Practices. **Journal of Marketing Research**, v. 54, n. 1, p. 27-43, 2017.

SONG, M; ZHAO, X; HAJHONG, E; Ou, Z. Statistic-Based CRM Approach via Time Series Segmenting RFM on Large Scale Data. **2016 IEEE/ACM 9th International Conference on Utility and Cloud Computing (UCC), Utility and Cloud Computing (UCC)**, 2016.

STEINMANN, Dan; MURPHY, Lincoln; MEHTA, Nick. Customer Success: how innovative companies are reducing churn and growing recurring revenue. Indianapolis: John Wiley & Sonc Inc, 2016.

SUBRAMANYA, K; SOMANI, A; Enhanced feature mining and classifier models to predict customer churn for an E-retailer. **7th International Conference on Cloud Computing, Data Science & Engineering**, 2017.

THOMAS, L.; TANG, L.; FLETCHER, M.; MARSHALL, A.; THOMAS, L.; PAN, J. Assessing the impact of derived behavior information on customer attrition in the financial service industry. **European Journal Of Operational Research**, v. 236, n. 2, p. 624-633, 2014.

TOLEDO, Geraldo Luciano; MORETTI, Sérgio Luiz do Amaral. Valor para o cliente e valor do cliente: conceitos e implicações para o processo de marketing. **Desenvolvimento em Questão**, v. 14, n. 35, p. 400-419, jul./set. 2016.

VAN DEN POEL, D; LARIVIÈRE, B. Customer attrition analysis for financial services using proportional hazard models. **European Journal of Operational Research**, v. 157, n. 1, p. 196, Aug. 16, 2004.

VANDECRUYS, O. et al. Mining software repositories for comprehensible software fault prediction models. **The Journal of Systems & Software**, n. 81, p. 823-839, Jan. 1, 2008.

VELOSO, F. J. M. **Um modelo para previsão de churn na área do retalho**. Universidade do Minho: Braga, 2013.

VERBEKE, W; et al. Building comprehensible customer churn prediction models with advanced rule induction techniques. **Expert Systems With Applications**, n. 38, p. 2354-2364, Jan. 1, 2011.

VERGARA, Sylvia Constant. **Métodos de pesquisa em administração**. 6. ed. São Paulo: Atlas, 2015.

WAZLAWICK, R.S. **Metodologia de pesquisa para ciência da computação**. Rio de Janeiro: Elsevier/Campus, 2014.

ZEITHAML, V. A. Consumer perceptions of price, quality and value: a means-end model and synthesis of evidence. **Journal of Marketing**, v. 52, n. 3, p. 2-22, 1988.

ZIMMERMAN, N. et al. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. **Atmospheric Measurement Techniques**, [s. l.], v. 11, n. 1, p. 291–313, 2018.

ZHU, B.; BAESENS, B.; BROUCKE, S.K.L.M. **An empirical comparison of techniques for the class imbalance problem in churn prediction**. Information Sciences, v. 408, p. 84-99, Oct. 2017.

APENDICE A – VARIÁVEIS EXPLICATIVAS FORNECIDAS PELO BANCO DO BRASIL

VARIÁVEIS EXPLICATIVAS	DESCRIÇÃO	Padrão
CLIENTE	Identificação do cliente em padrão anonimizado	Catagórico
ANO / MES	Ano e mês de referência da captura dos demais dados	Catagórico
RESTRICAO CADASTRAL	Ocorrência de restrições cadastrais (SPC, SERASA, CADIN ou Internas)	Catagórico
FXA_VLR_PGTO_FATURA_CARTAO	Faixa de valor pagamento da fatura de cartão (Abaixo do mínimo a pagou integral)	Catagórico
ESTADO CADASTRO	Situação cadastral (Normal, pendente, inativo, bloqueado)	Catagórico
RECEBE_BENEFICIO	Recebeu benefício	Catagórico
RISCO CLIENTE	Classificação de risco de crédito do cliente	Catagórico
UTLZA_CXQ_ESPE_ATE_50P	Utilizou cheque especial até 50%.	Catagórico
UTLZA_CXQ_ESPE_MAISQ_50P	Utilizou cheque especial mais do que 50%.	Catagórico
ESTADO_CTA CORRENTE	Estado da conta corrente (Ativa, sem movimentação, encerrada)	Catagórico
RECEBIMENTO SALARIO	Indica se cliente recebe salário no BB	Catagórico
MODALIDADE_PCTE_SERVICOS	Modalidade de Pacote de Serviços	Catagórico
CAPACIDADE_CIVIL	Capacidade Civil (Capaz, Relativamente Incapaz, Interditado...)	Catagórico
ESTADO_CIVIL	Estado Civil	Catagórico
ESTAGIO_VIDA	Estágio de Vida (Critério interno, desconsiderado)	Catagórico
FUNCAO REMUNERADA	Principal atividade profissional	Catagórico
NATUREZA_OCUPACAO	18 distintas natureza da ocupação profissional (perfil de vínculo profissional)	Catagórico
NICHO_MERCADO	22 distintas classificações de nichos/perfil	Catagórico
NIVEL_INSTRUCAO	Nível de escolaridade	Catagórico
NRO_CONVENIO_13SAL	Convênio de Crédito Antecipação 13º Salário	Catagórico
NRO_CONVENIO_CONSIG	Convênio de Crédito Consignado	Catagórico
NRO_CONVENIO_CRED_SAL	Convênio com acesso a linha Crédito Salário	Catagórico
PERFIL_DIGITAL	9 distintas classificações internas sobre perfil de utilização de soluções digitais	Catagórico
PRD_OFERTADO_UM	Primeira sugestão de oferta - Propensão	Catagórico
PRD_OFERTADO_DOIS	Segunda sugestão de oferta - Propensão	Catagórico
PRD_OFERTADO_TRES	Terceira sugestão de oferta - Propensão	Catagórico
SEGMENTO_CLIENTE	Segmento do cliente (5 diferentes perfis, ponderado renda e/ou volume de investimento)	Catagórico
SEGMENTO_CLIENTE	Segmento do cliente (5 diferentes perfis, ponderado renda e/ou volume de investimento)	Catagórico
SEXO	Sexo	Catagórico
TIPO_CADASTRAMENTO	Tipo do Cadastro (Completo, intermediário, simplificado, indicativo)	Catagórico
ESTADO LIMITE CREDITO	Estado Limite de Crédito (Vigente ou Não Vigente)	Dummy (0/1)
INDICADOR_POSSE_ROTATIVO	Se possui produtos de crédito rotativo (cartão ou cheque especial)	Dummy (0/1)
INDICADOR_LOB_VAI	Indicador Livre Opção Bancária - Transferir proventos a outro banco	Dummy (0/1)
INDICADOR_LOB_VEM	Indicador Livre Opção Bancária - Trazer proventos de outro banco	Dummy (0/1)
POSSE_CONSORCIO_MOTO	Posse consórcio de moto	Dummy (0/1)
POSSE_CONSORCIO_VEICULO	Posse consórcio de veículo	Dummy (0/1)
POSSE CRED VEICULO	Posse crédito veículo	Dummy (0/1)
POSSE_CRED_VEICULO_PROPRIO	Posse CDC Veículo Próprio	Dummy (0/1)
POSSE_CXQ_ESP	Posse Cheque Especial	Dummy (0/1)
POSSE_DEPOSITO_A_PRAZO	Posse Depósito a Prazo/CDB	Dummy (0/1)
POSSE_ENVIO_SMS	Posse Serviço SMS	Dummy (0/1)
POSSE_FUNDO_INVESTIMENTO	Posse Fundos de Investimento	Dummy (0/1)

POSSE_LCA	Posse Letra de Crédito de Agronegócio	Dummy (0/1)
POSSE_LCI	Posse Letra de Crédito Imobiliária	Dummy (0/1)
POSSE_POUPANCA	Posse Poupança	Dummy (0/1)
POSSE_PREVIDENCIA_ABERTA	Posse Previdência	Dummy (0/1)
POSSE_SEG_PATRIMONIO	Posse Seguro Patrimonial	Dummy (0/1)
POSSE_SEGURO_PESSOAL	Posse Seguro Pessoal	Dummy (0/1)
POSSE_SEGURO_VEICULO	Posse Seguro Auto/Veículo	Dummy (0/1)
POSSE_CAP_PGTO_MENSAL	Posse Capitalização Pagamento Mensal	Dummy (0/1)
POSSE_CAP_PGTO_UNICO	Posse Capitalização Pagamento Único	Dummy (0/1)
POSSE_CARTAO_CREDITO	Posse Cartão de Crédito	Dummy (0/1)
POSSE_CONSORCIO_ELETRO	Posse Consórcio Eletro	Dummy (0/1)
QTDE_SIMULA_CDC	Histórico de simulação de CDC Crédito Novo por parte do cliente	Dummy (0/1)
QTDE_SIMULA_CDC_RENOV	Histórico de simulação de CDC Renovação por parte do cliente	Dummy (0/1)
QTDE_SIMULA_FIN_IMOB	Histórico de simulação de Financiamento Imobiliário por parte do cliente	Dummy (0/1)
QTDE_SIMULA_FIN_VEICULO	Histórico de simulação de Financiamento de Veículo por parte do cliente	Dummy (0/1)
USO_AG	Interação canal agência	Dummy (0/1)
USO_ATEND_INTERNET	Interação ambiente atendimento internet	Dummy (0/1)
USO_CALL_CENTER	Interação ativa central de atendimento	Dummy (0/1)
UTLZ_CNL_INTERNET	Utilização canal internet	Dummy (0/1)
UTZ_CART_AEREO	Utilização cartão em estabelecimentos segmento aéreo	Dummy (0/1)
UTZ_CART_ALUG_CAR	Utilização cartão em estabelecimentos segmento aluguel de veículos	Dummy (0/1)
UTZ_CART_AUTOPECAS	Utilização cartão em estabelecimentos segmento autopeças	Dummy (0/1)
UTZ_CART_COMBUSTIVEL	Utilização cartão em estabelecimentos segmento combustíveis	Dummy (0/1)
UTZ_CART_CONSTRUCAO	Utilização cartão em estabelecimentos segmento construção	Dummy (0/1)
UTZ_CART_DECORACAO	Utilização cartão em estabelecimentos segmento decoração	Dummy (0/1)
UTZ_CART_DROGARIA	Utilização cartão em estabelecimentos segmento drogaria	Dummy (0/1)
UTZ_CART_ENTRETENIMENTO	Utilização cartão em estabelecimentos segmento entretenimento	Dummy (0/1)
UTZ_CART_ESTACIONAMENTO	Utilização cartão em estabelecimentos segmento estacionamento	Dummy (0/1)
UTZ_CART_HOSPITAL	Utilização cartão em estabelecimentos segmento hospital	Dummy (0/1)
UTZ_CART_HOTEL	Utilização cartão em estabelecimentos segmento hotel	Dummy (0/1)
UTZ_CART_JOALHERIA	Utilização cartão em estabelecimentos segmento joalheria	Dummy (0/1)
UTZ_CART_LJ_DEPARTAMENTO	Utilização cartão em estabelecimentos segmento lojas de departamento	Dummy (0/1)
UTZ_CART_PEDAGIO	Utilização cartão em estabelecimentos segmento pedágio	Dummy (0/1)
UTZ_CART_RESTAURANTE	Utilização cartão em estabelecimentos segmento restaurante	Dummy (0/1)
UTZ_CART_SAQUES	Utilização cartão função saque em espécie	Dummy (0/1)
UTZ_CART_SERVICOS	Utilização cartão em estabelecimentos segmento serviços	Dummy (0/1)
UTZ_CART_SUPERMERCADO	Utilização cartão em estabelecimentos segmento supermercado	Dummy (0/1)
UTZ_CART_TELEMARKETING	Utilização cartão em estabelecimentos segmento telemarketing	Dummy (0/1)
UTZ_CART_TURISMO	Utilização cartão em estabelecimentos segmento turismo	Dummy (0/1)
UTZ_CART_VAREJO	Utilização cartão em estabelecimentos segmento varejo	Dummy (0/1)
UTZ_CART_VEICULOS	Utilização cartão em estabelecimentos segmento veículos	Dummy (0/1)
UTZ_CART_VESTUARIO	Utilização cartão em estabelecimentos segmento vestuário	Dummy (0/1)
UTZ_ELETRO	Utilização cartão de débito eletrônico	Dummy (0/1)
MARGEM DE CONTRIBUICAO	Margem de contribuição proporcionada pelas operações do cliente	Quantitativo
MARGEM CARTÃO DE CRÉDITO	Margem de contribuição arrecadada através de cartão de crédito	Quantitativo
MARGEM CXQ_ESP	Margem de contribuição arrecadada através de cheque especial	Quantitativo
MARGEM EMPRESTIMO	Margem de contribuição arrecadada através de empréstimos/financiamentos	Quantitativo
Valor PACOTE	Valor pago referente a Pacote de Serviços	Quantitativo
VLR RENDIMENTO	Valor dos rendimentos do cliente	Quantitativo
VLR_TOT_CHQ_ESP	Valor total utilizado de cheque especial	Quantitativo
ENDIVIDAMTO SFN	Saldo de endividamento no SFN (exceto BB)	Quantitativo

ENDIVIDAMTO BB	Saldo de endividamento BB	Quantitativo
VLR_TOT_EMP	Saldo devedor de empréstimos e consignados	Quantitativo
VLR_CDC_TOTAL	Saldo devedor de operações de crédito (empréstimos, consignados + financiamentos)	Quantitativo
ULTIMA MOV CC	Data da última movimentação em Conta Corrente	Quantitativo
DT_ATUALIZAO_CADASTRAL	Data da última atualização cadastral	Quantitativo
DT_ULT_UTLZ_INTERNET	Data da última utilização do canal internet	Quantitativo
DT_ULT_UTLZ_MOBILE	Data da última utilização do canal Mobile	Quantitativo
VLR_CDC_EMP	Saldo devedor de empréstimos	Quantitativo
VLR_INVESTIMENTO	Saldo de investimentos (Poupança, Fundos, CDB, LCA, LCI...)	Quantitativo
VLR_TOT_POUPANCA	Saldo em poupança	Quantitativo
QTDE_INTERACOES_AG	Quantidade de interações - canal agência	Quantitativo
QTDE_INTERACOES	Quantidade interações - Todos canais	Quantitativo
QTDE_INTERACOES_MOBILE	Quantidade de interações - canal mobile	Quantitativo
QTDE_PGTOS_INTERNET	Quantidade de pagamentos - internet	Quantitativo
QTDE_PRODUTOS_OBSERVADOS	Quantidade de Produtos Observados	Quantitativo
DATA_INICIO_CORRENTISTA	Data que cliente se tornou correntista (Tempo de conta)	Quantitativo
DT_REVISAO_CADASTRO	Data última revisão cadastral	Quantitativo
DT_ULT_UTLZ_ATEND	Data última interação em atendimento presencial agência	Quantitativo
DT_ULT_UTLZ_BPOSTAL	Data última utilização de banco postal	Quantitativo
DT_ULT_UTLZ_INTRANET	Data última interação ativa - canal agência (contato telefônico)	Quantitativo
DT_ULT_UTLZ_TAA	Data última utilização Terminal de Auto-atendimento	Quantitativo
DT_ULT_UTLZ_URA	Data última interação Central de Atendimento	Quantitativo
IDADE	Idade do cliente	Quantitativo
RENDA_PRESUMIDA	Renda presumida (critério interno, baseado na movimentação financeira)	Quantitativo
VLR_SAQUE	Montante de saques em espécie realizado da conta	Quantitativo

APENDICE B – PADRÃO DOS DADOS RECEBIDOS

CLIENTE	ANO / MES	MARGEM DE CONTRIBUICAO	RESTRICAO CADASTRAL	Valor PACOTE	ESTADO LIMITE CREDITO	RISCO CLIENTE	VLR RENDIMENTO	ENDIVIDAMTO SFN	ENDIVIDAMTO BB	VLR INVESTIMENTO	POSSE CRED VEICULO	POSSE PREVIDENCIA ABERTA	FUNCAO REMUNERADA	IDADE	SEXO
AMDEBRARJ	201710	R\$ 382,26	Sem Restrição / Ir	R\$ 20,36	VIGENTE	B	R\$ 16.451	R\$ 164.747	R\$ 164.747	R\$ -	S	N	ENGENHEIRO	40	M
AMDEBRARJ	201711	R\$ 379,69	Sem Restrição / Ir	R\$ 20,36	VIGENTE	B	R\$ 16.451	R\$ 163.384	R\$ 163.384	R\$ -	S	N	ENGENHEIRO	40	M
AMDEBRARJ	201712	R\$ 417,61	Sem Restrição / Ir	R\$ 20,36	VIGENTE	B	R\$ 16.451	R\$ 160.698	R\$ 160.698	R\$ -	S	N	ENGENHEIRO	40	M
AMDEBRARJ	201801	R\$ 482,44	Sem Restrição / Ir	R\$ 20,36	VIGENTE	B	R\$ 16.451	R\$ 195.217	R\$ 159.377	R\$ -	S	N	ENGENHEIRO	40	M
AMDEBRARJ	201802	R\$ 455,14	Sem Restrição / Ir	R\$ 20,36	VIGENTE	B	R\$ 16.451	R\$ 195.217	R\$ 159.377	R\$ -	S	N	ENGENHEIRO	40	M
AMDEBRARJ	201803	R\$ 473,13	Sem Restrição / Ir	R\$ 27,48	VIGENTE	B	R\$ 16.451	R\$ 156.534	R\$ 156.534	R\$ -	S	N	ENGENHEIRO	40	M
AMDEBRARJ	201804	R\$ 549,63	Sem Restrição / Ir	R\$ 27,48	VIGENTE	B	R\$ 16.451	R\$ 155.254	R\$ 155.254	R\$ -	S	N	ENGENHEIRO	40	M
AMDEBRARJ	201805	R\$ 512,85	Sem Restrição / Ir	R\$ 27,48	VIGENTE	B	R\$ 16.451	R\$ 153.827	R\$ 153.827	R\$ -	S	N	ENGENHEIRO	40	M
AMDEBRARJ	201807	R\$ 477,21	Sem Restrição / Ir	R\$ 27,48	VIGENTE	B	R\$ 16.451	R\$ 152.458	R\$ 152.458	R\$ -	S	N	ENGENHEIRO	40	M
AMDEBRARJ	201808	R\$ 483,60	Sem Restrição / Ir	R\$ 27,48	VIGENTE	B	R\$ 16.451	R\$ 127.021	R\$ 127.021	R\$ -	S	N	ENGENHEIRO	40	M
AMDEBRARJ	201809	R\$ 634,28	Sem Restrição / Ir	R\$ 27,48	VIGENTE	B	R\$ 16.451	R\$ 125.566	R\$ 125.566	R\$ -	S	N	ENGENHEIRO	41	M
AMDEBRARJ	201810	R\$ 897,80	Sem Restrição / Ir	R\$ 27,48	VIGENTE	B	R\$ 16.451	R\$ 160.188	R\$ 124.303	R\$ -	S	N	ENGENHEIRO	41	M
AMDEBRARJ	201811	R\$ 501,51	Sem Restrição / Ir	R\$ 27,48	VIGENTE	B	R\$ 16.451	R\$ 158.602	R\$ 122.674	R\$ -	S	N	ENGENHEIRO	41	M
AMDEBRARJ	201812	R\$ 372,68	Sem Restrição / Ir	R\$ 27,48	VIGENTE	B	R\$ 16.451	R\$ 119.655	R\$ 119.655	R\$ -	S	N	ENGENHEIRO	41	M
AMDEBRARJ	201901	R\$ 399,39	Fraca	R\$ 27,48	VIGENTE	B	R\$ 16.451	R\$ 118.208	R\$ 118.208	R\$ -	S	N	ENGENHEIRO	41	M
AMDEBRARJ	201902	R\$ 164,10	Fraca	R\$ 27,48	VIGENTE	B	R\$ 16.451	R\$ 118.208	R\$ 118.208	R\$ -	S	N	ENGENHEIRO	41	M
AMDEBYFDT	201710	R\$ 153,21	Sem Restrição / Ir	R\$ 55,89	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 1.400	R\$ 255.786	N	S	ODONTOLOGO	55	F
AMDEBYFDT	201711	R\$ 223,72	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 1.826	R\$ 289.536	N	S	ODONTOLOGO	55	F
AMDEBYFDT	201712	R\$ 261,83	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 1.400	R\$ 300.916	N	S	ODONTOLOGO	55	F
AMDEBYFDT	201801	R\$ 202,73	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 1.400	R\$ 316.918	N	S	ODONTOLOGO	55	F
AMDEBYFDT	201802	R\$ 179,10	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 1.400	R\$ 332.518	N	S	ODONTOLOGO	56	F
AMDEBYFDT	201803	R\$ 81,59	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 1.400	R\$ 349.339	N	S	ODONTOLOGO	56	F
AMDEBYFDT	201804	R\$ 173,89	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 1.400	R\$ 349.339	N	S	ODONTOLOGO	56	F
AMDEBYFDT	201805	R\$ 205,41	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 2.000	R\$ 351.497	N	S	ODONTOLOGO	56	F
AMDEBYFDT	201807	R\$ 171,47	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 2.000	R\$ 352.259	N	S	ODONTOLOGO	56	F
AMDEBYFDT	201808	R\$ 170,63	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 2.000	R\$ 353.340	N	S	ODONTOLOGO	56	F
AMDEBYFDT	201809	R\$ 206,93	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 2.000	R\$ 355.645	N	S	ODONTOLOGO	56	F
AMDEBYFDT	201810	R\$ 306,49	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 2.000	R\$ 356.459	N	S	ODONTOLOGO	56	F
AMDEBYFDT	201811	R\$ 192,37	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 7.342	R\$ -	R\$ 2.000	R\$ 378.328	N	S	ODONTOLOGO	56	F
AMDEBYFDT	201812	R\$ 1.229,40	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 9.653	R\$ -	R\$ 2.000	R\$ 380.738	N	S	ODONTOLOGO	56	F
AMDEBYFDT	201901	R\$ 435,65	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 9.653	R\$ -	R\$ 2.000	R\$ 531.358	N	S	ODONTOLOGO	56	F
AMDEBYFDT	201902	R\$ 86,73	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 9.653	R\$ -	R\$ 2.000	R\$ 56.671	N	S	ODONTOLOGO	57	F
AMDEQBKCK	201710	R\$ 68,94	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 1.817	R\$ 36.843	R\$ 36.843	R\$ -	N	N	APOSENTADO OU PE	72	F
AMDEQBKCK	201711	R\$ 82,84	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 2.108	R\$ 36.844	R\$ 36.844	R\$ -	N	N	APOSENTADO OU PE	72	F
AMDEQBKCK	201712	R\$ 100,91	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 2.108	R\$ 36.843	R\$ 36.843	R\$ -	N	N	APOSENTADO OU PE	72	F
AMDEQBKCK	201801	R\$ 77,44	Sem Restrição / Ir	R\$ 58,28	VIGENTE	B	R\$ 2.108	R\$ 36.843	R\$ 36.843	R\$ -	N	N	APOSENTADO OU PE	72	F

APENDICE C – REGRESSÃO LOGÍSTICA

Output RStudio – Regressão logística – A1

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No  Yes
##           No 1565 372
##           Yes  41 101
##
##           Accuracy : 0.8013
##           95% CI : (0.7835, 0.8183)
##           No Information Rate : 0.7725
##           P-Value [Acc > NIR] : 0.0007972
##           Kappa : 0.2496
##           McNemar's Test P-Value : < 2.2e-16
##           Sensitivity : 0.21353
##           Specificity : 0.97447
##           Pos Pred Value : 0.71127
##           Neg Pred Value : 0.80795
##           Prevalence : 0.22751
##           Detection Rate : 0.04858
##           Detection Prevalence : 0.06830
##           Balanced Accuracy : 0.59400
##           'Positive' Class : Yes
## AUC_Logit_2018-11-30 [1] 0.7331163
```

Output RStudio – Regressão logística – A2

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No  Yes
##           No 1608 352
##           Yes  38  79
##
##           Accuracy : 0.8122
##           95% CI : (0.7948, 0.8288)
##           No Information Rate : 0.7925
##           P-Value [Acc > NIR] : 0.01347
##           Kappa : 0.2191
##           McNemar's Test P-Value : < 2e-16
##           Sensitivity : 0.18329
##           Specificity : 0.97691
##           Pos Pred Value : 0.67521
##           Neg Pred Value : 0.82041
##           Prevalence : 0.20751
##           Detection Rate : 0.03804
##           Detection Prevalence : 0.05633
##           Balanced Accuracy : 0.58010
##           'Positive' Class : Yes
## AUC_Logit_2018-12-31 [1]: 0.7422325
```

Análise A3 - Regressão Logística

Output RStudio – Regressão logística – A3

```
## Confusion Matrix and Statistics
##
##               Reference
## Prediction   No  Yes
##      No  1608  335
##      Yes   46   91
##
##               Accuracy : 0.8168
##               95% CI : (0.7995, 0.8332)
##      No Information Rate : 0.7952
##      P-Value [Acc > NIR] : 0.007239
##               Kappa : 0.2483
##      McNemar's Test P-Value : < 2.2e-16
##               Sensitivity : 0.21362
##               Specificity : 0.97219
##               Pos Pred Value : 0.66423
##               Neg Pred Value : 0.82759
##               Prevalence : 0.20481
##               Detection Rate : 0.04375
##      Detection Prevalence : 0.06587
##               Balanced Accuracy : 0.59290
##               'Positive' Class : Yes
## AUC_Logit_2019-01-31 [1]: 0.756774
```

APENDICE D – REGRESSÃO LOGÍSTICA STEPWISE

Output RStudio – Regressão Stepwise – A1

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No  Yes
##           No 1573 394
##           Yes  33  79
##
##           Accuracy : 0.7946
##           95% CI : (0.7766, 0.8118)
##           No Information Rate : 0.7725
##           P-Value [Acc > NIR] : 0.008115
##           Kappa : 0.2004
##           McNemar's Test P-Value : < 2.2e-16
##           Sensitivity : 0.16702
##           Specificity : 0.97945
##           Pos Pred Value : 0.70536
##           Neg Pred Value : 0.79969
##           Prevalence : 0.22751
##           Detection Rate : 0.03800
##           Detection Prevalence : 0.05387
##           Balanced Accuracy : 0.57324
##           'Positive' Class : Yes
##           AUC_Logit_2018-11-30 [1] 0. 7157699
```

Output RStudio – Regressão Stepwise – A2

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No  Yes
##           No 1625 365
##           Yes  21  66
##
##           Accuracy : 0.8142
##           95% CI : (0.7967, 0.8307)
##           No Information Rate : 0.7925
##           P-Value [Acc > NIR] : 0.007454
##           Kappa : 0.199
##           McNemar's Test P-Value : < 2.2e-16
##           Sensitivity : 0.15313
##           Specificity : 0.98724
##           Pos Pred Value : 0.75862
##           Neg Pred Value : 0.81658
##           Prevalence : 0.20751
##           Detection Rate : 0.03178
##           Detection Prevalence : 0.04189
##           Balanced Accuracy : 0.57019
##           'Positive' Class : Yes
##           AUC_Logit_2018-12-31` [1]: 0.7009653
```

Output RStudio – Regressão Stepwise – A3

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No  Yes
##           No 1617 359
##           Yes  37  67
##
##           Accuracy : 0.8096
##           95% CI : (0.7921, 0.8263)
##           No Information Rate : 0.7952
##           P-Value [Acc > NIR] : 0.05353
##           Kappa : 0.1875
##           Mcnemar's Test P-Value : < 2e-16
##           Sensitivity : 0.15728
##           Specificity : 0.97763
##           Pos Pred Value : 0.64423
##           Neg Pred Value : 0.81832
##           Prevalence : 0.20481
##           Detection Rate : 0.03221
##           Detection Prevalence : 0.05000
##           Balanced Accuracy : 0.56745
##           'Positive' Class : Yes
## AUC_Logit_2019-01-31 [1]: 0.7209525
```

APENDICE E – COEFICIENTES DA REGRESSÃO LOGÍSTICA E REGRESSÃO STEPWISE

Coeficientes da Regressão Logística e Regressão Logística Stepwise						
Variável dependente: Desidratado	Logistic Regression			Stepwise Logistic Regression		
Variáveis independentes	Análise / Referência			Análise / Referência		
	1	2	3	1	2	3
RISCO_CLIENTEB	-0.52** (0.24)	-0.50** (0.24)	-0.38 (0.25)	-0.50** (0.23)	-0.44** (0.22)	-0.36 (0.23)
RISCO_CLIENTEC	-0.56** (0.28)	-0.69** (0.28)	-0.62** (0.29)	-0.55** (0.26)	-0.75*** (0.25)	-0.76*** (0.26)
RISCO_CLIENTED	-1.58*** (0.54)	-1.37** (0.54)	-1.78*** (0.59)	-1.43*** (0.52)	-1.21** (0.50)	-1.64*** (0.54)
RISCO_CLIENTEE	-3.01** (1.29)	-2.54** (1.12)	-3.67*** (1.18)	-2.42** (1.16)	-1.78* (0.96)	-2.23** (0.88)
RESTRICAO_CADASTRALFraca	-0.71 (0.81)	-0.79 (0.89)	-1.72* (0.88)	-0.81 (0.75)	-0.55 (0.73)	-1.40* (0.76)
RESTRICAO_CADASTRALRelativa	-1.03 (0.72)	-1.23 (0.77)	-1.35* (0.76)	-0.96 (0.71)	-0.87 (0.68)	-0.82 (0.69)
RESTRICAO_CADASTRALSem Restrição / Informativa	-1.53** (0.76)	-1.53* (0.85)	-2.18*** (0.83)	-1.63** (0.71)	-1.25* (0.68)	-1.77** (0.71)
ESTADO.LIMITE.CREDITOVIGENTE	-0.46 (0.43)	-0.45 (0.44)	-0.20 (0.43)			
ESTADO_CIVILDIVORCIADO	0.41* (0.23)	0.57** (0.23)	0.49** (0.23)	0 (0.22)	0.57*** (0.22)	0.53** (0.22)
ESTADO_CIVILSEPARADO	0.25 (0.29)	0.81*** (0.28)	0.67** (0.29)	0 (0.27)	0.78*** (0.27)	0.60** (0.28)
ESTADO_CIVILSOLTEIRO	0.09 (0.17)	0.28 (0.17)	0.22 (0.18)	0 (0.16)	0.25 (0.16)	0.20 (0.17)
ESTADO_CIVILVIUVO	-0.03 (0.30)	0.08 (0.31)	-0.43 (0.32)	0 (0.27)	-0.04 (0.27)	-0.36 (0.28)
NATUREZA_OCUPACAOATIVIDADE SEM REMUNERACAO	-1.08* (0.57)	-1.64** (0.71)	-1.69** (0.72)			
NATUREZA_OCUPACAOEMPREGADO DE PESSOA FÍSICA	0.31 (1.27)	-15.61 (1,005.09)	-14.62 (622.22)			
NATUREZA_OCUPACAOEMPREGADO EM EMPRESA DO SETOR PRIVADO	0.16 (0.25)	0.23 (0.25)	0.24 (0.25)			
NATUREZA_OCUPACAOEMPREGADO EVENTUAL/TEMPORARIO	-12.05 (882.74)	-13.68 (1,455.40)	-13.30 (882.74)			
NATUREZA_OCUPACAOEMPREGADOR- TITULAR/PROPRIET DE EMPRESA	0.09 (0.26)	0.11 (0.27)	0.21 (0.27)			
NATUREZA_OCUPACAOFUNC DE EMPRESA PUBL OU DE ECON MISTA	-0.01 (0.42)	-0.42 (0.47)	-0.50 (0.48)			
NATUREZA_OCUPACAOFUNC EMP PUBL OU ECON MISTA NAO CONCURS	-14.37 (882.74)	-14.83 (1,455.40)	-13.33 (882.74)			
NATUREZA_OCUPACAOPOLITICO	-0.18 (0.57)	-0.10 (0.55)	-0.60 (0.60)			
NATUREZA_OCUPACAOPROFIS LIBERAL/TRAB SEM VINC DE EMPREGO	0.01 (0.28)	-0.25 (0.30)	-0.45 (0.31)			
NATUREZA_OCUPACAOSERVIDOR NAO CONCURSADO/REGIME ESPECIAL	-0.38 (0.63)	-0.23 (0.63)	-1.50* (0.86)			
NATUREZA_OCUPACAOSERVIDOR PUBLICO DA ADMINISTRACAO DIRETA	-0.20 (0.21)	-0.12 (0.22)	-0.24 (0.22)			
NATUREZA_OCUPACAOSERVIDOR PUBLICO DE AUTARQUIA E FUNDACAO	0.27 (0.32)	0.20 (0.33)	-0.09 (0.35)			
NATUREZA_OCUPACAOSOCIO DE MICRO EMPRESA	0.09 (0.45)	-0.41 (0.49)	0.35 (0.46)			
NATUREZA_OCUPACAOTRABALHADOR SEM VINCULO DE EMPREGO	-0.06 (0.25)	-0.30 (0.26)	-0.44 (0.27)			
NICHO_MERCADOAGENTE PÚBLICO ESTRATÉGICO	16.99 (598.85)	30.41 (1,064.45)	16.75 (622.32)	16.79 (591.22)	0	0
NICHO_MERCADOBENEFICIARIO DO INSS	1.47 (1.54)	13.89 (350.18)	2.20 (1.64)	1.34 (1.52)	0	0
NICHO_MERCADODIRIGENTE DE EMPRESA	1.09	14.21	1.42	1.23	0	0

	(1.11)	(350.17)	(1.21)	(1.10)	0	0
NICHO_MERCADOEMPREGADO PÚBLICO	1.41	14.49	1.73	1.37	0	0
	(1.11)	(350.17)	(1.21)	(1.08)	0	0
NICHO_MERCADOEstilo UNV	2.33**	14.93	1.90	2.09*	0	0
	(1.18)	(350.18)	(1.31)	(1.15)	0	0
NICHO_MERCADOFUNIONÁRIO BB	2.64**	15.30	2.31*	2.48**	0	0
	(1.16)	(350.17)	(1.26)	(1.13)	0	0
NICHO_MERCADOINVESTIDORES	1.80*	14.77	2.09*	1.78*	0	0
	(1.08)	(350.17)	(1.18)	(1.07)	0	0
NICHO_MERCADOINVESTIDORES QUALIFICADOS	2.37**	14.79	2.21*	2.34**	0	0
	(1.11)	(350.17)	(1.22)	(1.09)	0	0
NICHO_MERCADOMATURIDADE DIGITAL	1.74	14.71	1.74	1.64	0	0
	(1.09)	(350.17)	(1.20)	(1.08)	0	0
NICHO_MERCADOMÉDIO PRODUTOR RURAL	2.28	15.43	3.91***	2.20	0	0
	(1.42)	(350.18)	(1.48)	(1.39)	0	0
NICHO_MERCADOPEQUENO PRODUTOR RURAL	2.91**	15.78	3.39***	2.77**	0	0
	(1.20)	(350.18)	(1.30)	(1.17)	0	0
NICHO_MERCADOPRODUTOR RURAL NÃO NEGOCIAL	17.24	30.54	18.39	16.97	0	0
	(594.48)	(1,496.93)	(882.74)	(573.10)	0	0
NICHO_MERCADOPROFISSIONAIS ESTRATÉGICOS	1.25	14.25	1.77	1.27	0	0
	(1.09)	(350.17)	(1.20)	(1.07)	0	0
NICHO_MERCADOPROVENTISTA	1.50	14.38	2.21*	1.49	0	0
	(1.11)	(350.17)	(1.21)	(1.09)	0	0
NICHO_MERCADOSEM INFORMACAO	1.48	14.65	2.20*	1.49	0	0
	(1.16)	(350.18)	(1.26)	(1.14)	0	0
NICHO_MERCADOSERVIDOR PUBLICO	1.34	14.20	1.73	1.18	0	0
	(1.10)	(350.17)	(1.20)	(1.08)	0	0
NIVEL_INSTRUCAOFUNDAMENTAL	-0.69	-0.92*	-0.65	-0.88*	0	0
	(0.54)	(0.55)	(0.53)	(0.52)	0	0
NIVEL_INSTRUCAOMÉDIO	-0.18	-0.21	-0.40	-0.25	0	0
	(0.43)	(0.42)	(0.42)	(0.41)	0	0
NIVEL_INSTRUCAOMESTRADO	0.62	0.48	0.46	0.59	0	0
	(0.54)	(0.54)	(0.54)	(0.51)	0	0
NIVEL_INSTRUCAOPOS-GRADUACAO	0.36	0.03	-0.26	0.30	0	0
	(0.47)	(0.46)	(0.47)	(0.45)	0	0
NIVEL_INSTRUCAOSEM INFORMACAO	-12.50	-13.18	2.65	-13.67	0	0
	(620.10)	(1,006.94)	(1.68)	(616.92)	0	0
NIVEL_INSTRUCAOSUPERIOR COMPLETO	0.19	-0.02	-0.20	0.11	0	0
	(0.41)	(0.40)	(0.39)	(0.39)	0	0
NIVEL_INSTRUCAOSUPERIOR EM ANDAMENTO	-0.01	-0.35	-0.61	-0.11	0	0
	(0.67)	(0.70)	(0.77)	(0.65)	0	0
SEXOM	-0.12	-0.03	-0.06			
	(0.13)	(0.13)	(0.14)			
MARGEM.DEMAIS.EXCETO.CARTÃO	-0.0001**			-0.0002**	0	0
	(0.0001)			(0.0001)	0	0
MARGEM.DE.CONTRIBUICAO		-0.0001	-0.0000	0	-0.0002*	0
		(0.0001)	(0.0000)	0	(0.0001)	0
MARGEM.CARTÃO.DE.CRÉDITO	-0.001**	-0.0003	-0.0003	-0.001**	-0.0004**	-0.0002
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0001)
MARGEM_CXQ_ESP	0.0004	-0.001	-0.002***	0	-0.0005	-0.001***
	(0.0005)	(0.0004)	(0.0004)	0	(0.0003)	(0.0004)
MARGEM_EMPRESTIMO	-0.0002	0.0002	0.0000	0	0.0003	0
	(0.0002)	(0.0002)	(0.0001)	0	(0.0002)	0
VLR_TOT.MC_CARTAO	0.001	-0.0002	0.0003	0.001	0	0
	(0.0005)	(0.0004)	(0.0004)	(0.0004)	0	0
Valor.PACOTE	-0.002	0.002	-0.01	0	0	-0.02*
	(0.01)	(0.01)	(0.01)	0	0	(0.01)
INDICADOR_POSSE_ROTATIVO	-1.06***	-0.58**	-0.70**	-0.96***	-0.50*	-0.59**
	(0.29)	(0.27)	(0.29)	(0.27)	(0.26)	(0.27)
VLR_RENDIMENTO	-0.0000	-0.0000	-0.0000*	0	0	-0.0000*
	(0.0000)	(0.0000)	(0.0000)	0	0	(0.0000)
VLR_TOT_CHQ_ESP	0.0001	-0.0000	-0.0001***	0	0	-0.0001**
	(0.0000)	(0.0000)	(0.0000)	0	0	(0.0000)
ENDIVIDAMENTO.SFN	-0.0000	0.0000**	0.0000*	0	0.0000*	0.0000*
	(0.0000)	(0.0000)	(0.0000)	0	(0.0000)	(0.0000)
ENDIVIDAMENTO.BB	0.0000	-0.0000				
	(0.0000)	(0.0000)				
VLR_TOT_EMP			-0.0000	0	0	-0.0000

			(0.0000)	0	0	(0.0000)
VLR_CDC_EMP	0.0000*	0.0000	0.0000	0.0000**	0	0
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	0	0
VLR_INVESTIMENTO	-0.0000**	-0.0000***	-0.0000***	-0.0000**	-0.0000***	-0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
VLR_TOT_POUPANCA	0.0000	0.0000	0.0000			
	(0.0000)	(0.0000)	(0.0000)			
QTDE_INTERACOES_AG	0.004	0.001	-0.005			
	(0.01)	(0.01)	(0.01)			
QTDE_INTERACOES	0.002	-0.001	0.003			
	(0.003)	(0.003)	(0.003)			
QTDE_INTERACOES_MOBILE	0.0003	0.003	-0.001			
	(0.004)	(0.004)	(0.004)			
QTDE_PGTOS_INTERNET	-0.04	-0.09**	-0.04	0	-0.08**	0
	(0.03)	(0.04)	(0.04)	0	(0.04)	0
QTDE_PRODUTOS_OBSERVADOS	-0.58***	-0.64***	-0.68***	-0.57***	-0.61***	-0.61***
	(0.06)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)
IDADE	0.002	-0.07	0.16			
	(0.21)	(0.24)	(0.22)			
NRO_CONVENIO_13SAL	-0.0000	-0.0000	-0.0000			
	(0.0000)	(0.0000)	(0.0000)			
NRO_CONVENIO_CONSIG	-0.0000	0.0000	0.0000			
	(0.0000)	(0.0000)	(0.0000)			
NRO_CONVENIO_CRED_SAL	0.0000	-0.0000	0.0000			
	(0.0000)	(0.0000)	(0.0000)			
QTDE_SIMULA_CDC	0.12	0.07	0.04			
	(0.14)	(0.14)	(0.13)			
QTDE_SIMULA_CDC_RENOV	-0.36	-0.002	-0.09			
	(0.28)	(0.27)	(0.28)			
QTDE_SIMULA_FIN_IMOB	0.21	-0.18	0.06			
	(0.22)	(0.22)	(0.22)			
QTDE_SIMULA_FIN_VEICULO	0.82	0.67	-0.56	0.87	0	0
	(0.56)	(0.59)	(0.58)	(0.54)	0	0
RENDA_PRESUMIDA	0.0000	0.0000	0.0000			
	(0.0000)	(0.0000)	(0.0000)			
SEGMENTO_CLIENTE	0.0001	0.0002	0.0004			
	(0.0004)	(0.0004)	(0.0004)			
USO_AG	-0.13	0.25	-0.50**	0	0	-0.48***
	(0.20)	(0.19)	(0.20)	0	0	(0.18)
USO_ATEND_INTERNET	-0.04	-0.04	-0.01			
	(0.06)	(0.06)	(0.07)			
USO_CALL_CENTER	-0.03	0.19	0.20			
	(0.21)	(0.21)	(0.21)			
UTLZ_CNL_INTERNET	0.11	0.22	0.19			
	(0.21)	(0.21)	(0.21)			
UTZ_CART_AEREO	0.25	-0.19	-0.03	0.27*	0	0
	(0.16)	(0.17)	(0.17)	(0.15)	0	0
UTZ_CART_ALUG_CAR	0.11	-0.15	0.09			
	(0.16)	(0.17)	(0.17)			
UTZ_CART_AUTOPECAS	-0.04	-0.02	-0.18			
	(0.15)	(0.16)	(0.17)			
UTZ_CART_COMBUSTIVEL	-0.11	0.29*	0.04	0	0.25*	0
	(0.15)	(0.16)	(0.15)	0	(0.14)	0
UTZ_CART_CONSTRUCAO	-0.16	-0.06	-0.08			
	(0.15)	(0.14)	(0.15)			
UTZ_CART_DECORACAO	-0.03	0.20	-0.08			
	(0.18)	(0.18)	(0.19)			
UTZ_CART_DROGARIA	-0.01	-0.07	-0.02			
	(0.14)	(0.14)	(0.15)			
UTZ_CART_ENTRETENIMENTO	-0.27	-0.29	0.33**	-0.29*	-0.29*	0.30*
	(0.18)	(0.18)	(0.17)	(0.17)	(0.17)	(0.16)
UTZ_CART_ESTACIONAMENTO	0.28	0.14	-0.24			
	(0.20)	(0.19)	(0.21)			
UTZ_CART_HOSPITAL	0.01	0.22	-0.15			
	(0.16)	(0.16)	(0.16)			
UTZ_CART_HOTEL	-0.44***	-0.08	-0.16	-0.41***	0	0
	(0.14)	(0.15)	(0.15)	(0.13)	0	0
UTZ_CART_JOALHERIA	-0.15	0.09	0.15			
	(0.23)	(0.20)	(0.20)			
UTZ_CART_PEDAGIO	0.75*	-0.51	-0.23	0.71*	0	0
	(0.42)	(0.47)	(0.41)	(0.41)	0	0

UTZ_CART_RESTAURANTE	0.26*	0.11	0.36**	0.23*	0	0.35**
	(0.15)	(0.15)	(0.16)	0	0	0
UTZ_CART_SERVICOS	0.06	-0.06	-0.29**	0	0	-0.27**
	(0.12)	(0.13)	(0.13)	0	0	(0.12)
UTZ_CART_SUPERMERCADO	0.09	-0.10	0.07			
UTZ_CART_TELEMARKETING	-0.28**	-0.18	-0.08	-0.23*	-0.20	0
	(0.12)	(0.14)	(0.13)	(0.12)	(0.13)	0
UTZ_CART_TURISMO	0.17	0.02	0.05			
	(0.18)	(0.20)	(0.22)			
UTZ_CART_VAREJO	-0.11	-0.05	0.05			
	(0.12)	(0.14)	(0.14)			
UTZ_CART_VEICULOS	0.28	0.36*	-0.24	0	0.35*	-0.27
	(0.19)	(0.20)	(0.18)	0	(0.19)	(0.18)
UTZ_CART_VESTUARIO	-0.15	-0.01	0.09			
	(0.12)	(0.13)	(0.13)			
UTZ_ELETRO	0.06	-0.07	-0.17	0	0	-0.18
	(0.12)	(0.13)	(0.13)	0	0	(0.13)
VLR_SAQUE	0.0000	0.0000*	0.0000	0	0.0000*	0
	(0.0000)	(0.0000)	(0.0000)	0	(0.0000)	0
Constant	-0.47	-13.27	-0.42	-0.75	0.28	0.71
	(1.38)	(350.18)	(1.41)	(1.30)	(0.71)	(0.74)
Observations	2,079	2,077	2,08	2,079	2,077	2,08
Log Likelihood	-967.81	-911.82	-890.33	-988.73	-952.99	-931.77
Akaike Inf. Crit.	2,149.62	2,037.63	1,994.67	2,067.45	1,957.97	1,919.53
Note:			*p<0.1; **p<0.05; ***p<0.01			*p<0.1; **p<0.05; ***p<0.01

APENDICE F – RANDOM FORESTS

Output Rstudio – Random Forests – A1

```
## Random Forest
##
## 1560 samples
##   64 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1403, 1405, 1405, 1404, 1404, ...
## Resampling results across tuning parameters:
##
##   mtry Accuracy   Kappa
##    2   0.7737266 0.01478279
##    5   0.7820518 0.10870755
##   10   0.7872047 0.19088980
##   20   0.7827133 0.21435826
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 10.
##
## Trained on 1,092 samples, validated on 468 samples (batch_size=30,
epochs=50)
## Final epoch (plot to see history):
## val_loss: 0.361
## val_acc: 0.8953
##   loss: 0.3949
##   acc: 0.8718
##
## Confusion Matrix and Statistics
##
##           Reference
## Prediction No Yes
##           No 391 105
##           Yes 10 13
##
##           Accuracy : 0.7784
##           95% CI : (0.7402, 0.8134)
##           No Information Rate : 0.7726
##           P-Value [Acc > NIR] : 0.4002
##           Kappa : 0.1191
##           McNemar's Test P-Value : <2e-16
##           Sensitivity : 0.11017
##           Specificity : 0.97506
##           Pos Pred Value : 0.56522
##           Neg Pred Value : 0.78831
##           Prevalence : 0.22736
##           Detection Rate : 0.02505
##           Detection Prevalence : 0.04432
##           Balanced Accuracy : 0.54262
##           'Positive' Class : Yes
## `AUC_RF_2018-11-30` [1]: 0.5426159
```

Output RStudio – Random Forests – A2

```

## Random Forest
##
## 1559 samples
##   64 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1402, 1403, 1404, 1403, 1404, 1403, ...
## Resampling results across tuning parameters:
##
##   mtry Accuracy   Kappa
##    2  0.7966551  0.03592762
##    5  0.8037024  0.12563098
##   10  0.8107539  0.22901208
##   20  0.8113785  0.26989359
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 20.
##
## Trained on 1,090 samples, validated on 468 samples (batch_size=30,
epochs=50)
## Final epoch (plot to see history):
## val_loss: 0.3556
## val_acc: 0.8953
##   loss: 0.3598
##   acc: 0.8881
##
## Confusion Matrix and Statistics
##
##           Reference
## Prediction No Yes
##      No  390  87
##      Yes   21  20
##
##           Accuracy : 0.7915
##           95% CI : (0.7539, 0.8257)
##   No Information Rate : 0.7934
##   P-Value [Acc > NIR] : 0.5687
##           Kappa : 0.176
## Mcnemar's Test P-Value : 3.985e-10
##           Sensitivity : 0.18692
##           Specificity : 0.94891
##   Pos Pred Value : 0.48780
##   Neg Pred Value : 0.81761
##           Prevalence : 0.20656
##   Detection Rate : 0.03861
##   Detection Prevalence : 0.07915
##   Balanced Accuracy : 0.56791
##   'Positive' Class : Yes
## `AUC_RF_2018-12-31` [1]: 0.5679105

```

Output RStudio – Random Forests – A3

```

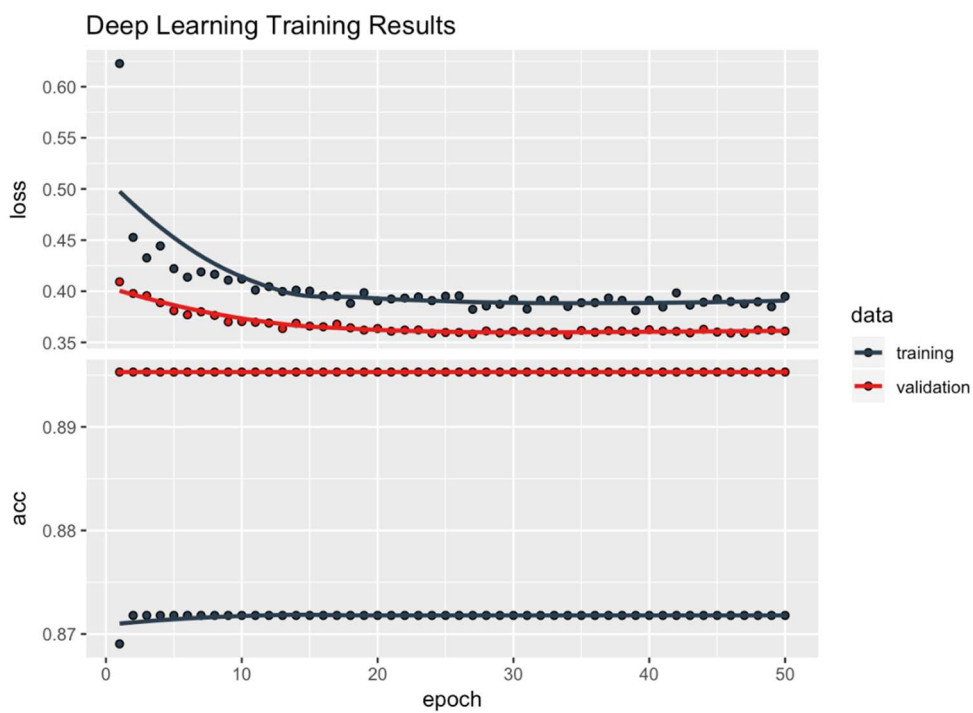
## Random Forest
##
## 1561 samples
##   64 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1405, 1405, 1405, 1404, 1405, 1405, ...
## Resampling results across tuning parameters:
##
##   mtry Accuracy   Kappa
##    2  0.7988486  0.03283835
##    5  0.8058917  0.12775278
##   10  0.8167810  0.25079288
##   20  0.8148661  0.27463130
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 10.
##
## Trained on 1,092 samples, validated on 468 samples (batch_size=30, epochs=50)
## Final epoch (plot to see history):
## val_loss: 0.3395
## val_acc: 0.906
##   loss: 0.3541
##   acc: 0.8864
##
## Confusion Matrix and Statistics
##
##           Reference
## Prediction No Yes
##           No 401 87
##           Yes 12 19
##
##           Accuracy : 0.8092
##           95% CI : (0.7728, 0.8422)
##           No Information Rate : 0.7958
##           P-Value [Acc > NIR] : 0.2412
##           Kappa : 0.2038
##           McNemar's Test P-Value : 1.028e-13
##           Sensitivity : 0.17925
##           Specificity : 0.97094
##           Pos Pred Value : 0.61290
##           Neg Pred Value : 0.82172
##           Prevalence : 0.20424
##           Detection Rate : 0.03661
##           Detection Prevalence : 0.05973
##           Balanced Accuracy : 0.57509
##           'Positive' Class : Yes
## `AUC_RF_2019-01-31` [1]: 0.5750948

```

APENDICE G – REDES NEURAIS

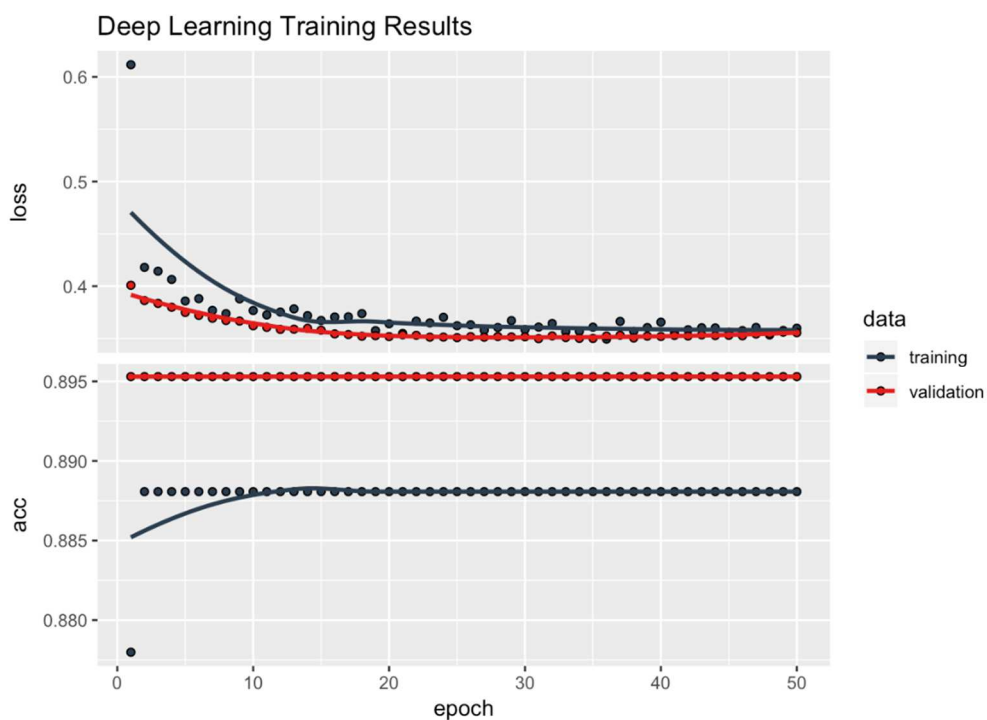
Output RStudio – Redes Neurais – A1

```
## Trained on 1,092 samples, validated on 468 samples
## (batch_size=30, epochs=50)
## Final epoch (plot to see history):
## val_loss: 0.361
## val_acc: 0.8953
## loss: 0.3949
## acc: 0.8718
##
##           Truth
## Prediction no yes
##      no 458 61
##      yes  0  0
##
## `Accuracy_keras_2018-11-30`
## # A tibble: 2 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>       <dbl>
## 1 accuracy binary      0.882
## 2 kap     binary          0
## AUC_keras_2018-11-30` [1] 0.5007517
```



Output RStudio – Redes Neurais – A2

```
## Trained on 1,090 samples, validated on 468 samples
(batch_size=30, epochs=50)
## Final epoch (plot to see history):
## val_loss: 0.3556
## val_acc: 0.8953
## loss: 0.3598
## acc: 0.8881
##
##          Truth
## Prediction no yes
##          no 466 53
##          yes  0  0
##
## `Accuracy_keras_2018-12-31`
## # A tibble: 2 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>       <dbl>
## 1 accuracy binary       0.898
## 2 kap     binary         0
## AUC_keras_2018-12-31[1] 0.5097174
```



Output RStudio – Redes Neurais – A3

```

## Trained on 1,092 samples, validated on 468 samples
## (batch_size=30, epochs=50)
## Final epoch (plot to see history):
## val_loss: 0.3395
##   val_acc: 0.906
##     loss: 0.3541
##     acc: 0.8864
##
##           Truth
## Prediction no yes
##           no 470 50
##           yes  0  0
##
## `Accuracy_keras_2019-01-31`
## # A tibble: 2 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>       <dbl>
## 1 accuracy binary       0.904
## 2 kap     binary         0
## AUC_keras_2018-01-31[1] 0.5313617

```

