

INSPER – INSTITUTO DE ENSINO E PESQUISA
MESTRADO PROFISSIONAL EM ADMINISTRAÇÃO DE EMPRESAS

DOUGLAS GALLO MARTINELLI

**EFETIVIDADE DA IMPLEMENTAÇÃO DO INDICADOR DE *CHURN* NA
RETENÇÃO DE CLIENTES EM UM BANCO COMERCIAL**

São Paulo

2020

DOUGLAS GALLO MARTINELLI

**EFETIVIDADE DA IMPLEMENTAÇÃO DO INDICADOR DE *CHURN* NA
RETENÇÃO DE CLIENTES EM UM BANCO COMERCIAL**

Dissertação apresentada ao Programa de Mestrado Profissional em Administração de Empresas com ênfase em estratégia do Insper Instituto de Ensino e Pesquisa, como parte dos requisitos para a obtenção do título de Mestrado Profissional em Administração de Empresas.

Área de concentração: Estratégia Competitiva

Orientador: Prof. Dr. Silvio Abrahão Laban Neto

São Paulo

2020

MARTINELLI, DOUGLAS GALO

Efetividade da implantação do Indicador de *Churn* na retenção de clientes em um Banco Comercial 64 f.

Douglas Gallo Martinelli – São Paulo, 2020.

Dissertação (Mestrado) – Insper.

Orientador: Prof. Dr. Silvio Abrahão Laban Neto

1. Análise de *Churn*.
2. Relacionamento com clientes.
3. Bancos.
4. Marketing de relacionamento.

DOUGLAS GALLO MARTINELLI

**EFETIVIDADE DA IMPLEMENTAÇÃO DO INDICADOR DE *CHURN* NA
RETENÇÃO DE CLIENTES EM UM BANCO COMERCIAL**

Dissertação apresentada ao Programa de Mestrado Profissional em Administração de Empresas com ênfase em Estratégia, como parte dos requisitos para a obtenção do título de Mestre Profissional em Administração de Empresas; Insper Instituto de Ensino e Pesquisa; Estratégia Competitiva.

Data de aprovação: 14/12/2020

Banca examinadora:

Orientador: Prof. Dr. Silvio Abrahão Laban Neto
Insper

Prof. Dr. Danny Pimentel Claro
Insper

Prof. Dr. Nelson Barth
FGV EAESP

Insper

A minha família, minha esposa, amigos e médicos, que jamais mediram esforços para me apoiar em meio às dificuldades e estiveram sempre presentes, compartilhando as alegrias.

Agradecimentos

Agradeço, em especial, ao professor Prof. Dr. Silvio Abrahão Laban Neto pela orientação e dedicação que possibilitaram essa experiência inesquecível de crescimento pessoal, acadêmico e profissional. Ao Prof. Dr. Danny P. Claro por compartilhar seu conhecimento e apoiar esse projeto. A todos os professores do programa de Mestrado Profissional em Administração do Insper, que me acolheram e prestaram um ensino de excelência ao longo destes últimos dois anos. Ao Insper, pela grande oportunidade de crescimento profissional, que me possibilitou ser um profissional focado e preparado para desafios.

Aos queridos colegas da turma MPA12, por todo o conhecimento compartilhado e que transformaram o processo de aprendizado em ambiente divertido e prazeroso. Ao Dr. Paulo Hoff e sua equipe, por cuidar da minha saúde me possibilitando vivenciar esse sonho.

E por fim, o mais importante, agradeço à Lucia, minha esposa Lucia, que sem ela, nada dessa jornada seria possível. A minha querida família e amigos especiais, em quem confio e sei que posso contar.

Resumo

MARTINELLI, Douglas Gallo. **Efetividade da Implementação do indicador de Churn na retenção de clientes em um Banco Comercial no Brasil, 2020.**

Dissertação (Mestrado) – Insper Instituto de Ensino e Pesquisa, São Paulo, 2020.

O ambiente competitivo, proveniente de novas tecnologias, a desregulamentação e a preocupação com os comportamentos complexos dos consumidores estão transformando a maneira como produtos e serviços financeiros são ofertados e acessados pelos clientes. Nesse cenário, manter a vantagem competitiva sustentável no longo prazo tornou-se um grande desafio para os bancos, assim como a dificuldade em reter e conquistar novos clientes. Em geral, a aquisição de um novo cliente custa muito mais caro do que a retenção do mesmo (Sharma e Panigrahi, 2011; Keramati *et al.*, 2016; Ganesh *et al.* 2000; Liou *et al.*, 2009).

Essa dissertação tem como objetivo analisar, utilizando a técnica da estatística de dados longitudinais em painel, qual o efeito da implementação do indicador da taxa de *churn* (*churn rate*) na carteira de clientes de cada gerente comercial do banco, como instrumento efetivo para reter clientes, bem como identificar fatores que afetam a probabilidade de retenção dos mesmos. Com base em uma amostra de clientes de um banco com atuação no Brasil, colhidas em um período de dois anos, sendo um ano antecessor e outro sucessor à implementação do indicador da taxa de *churn*.

Os resultados obtidos a partir das análises, como será comprovado nos capítulos subsequentes desta dissertação, sugerem que a implementação do indicador da taxa *churn* apresentou resultado negativo para a retenção dos clientes. Dessa forma, o estudo contribui com a discussão em torno do valor da utilização da análise de *churn* como estratégia de retenção de clientes, assim como a utilização de ferramentas analíticas para tomada de decisão.

PALAVRAS-CHAVE:

Análise de *Churn*; Retenção de Clientes; Relacionamento com Clientes.

Abstract

MARTINELLI, Douglas Gallo. Effectiveness of the implementation of the Churn indicator in customer retention at a commercial bank in Brazil, 2020.

Dissertation – Insper Instituto de Ensino e Pesquisa, São Paulo, 2020.

The competitive environment, stemming from new technologies, deregulation and also from the concern for the complex consumer behavior, has been changing the way financial products and services are offered and accessed by customers. In this scenario, maintaining a sustainable competitive advantage in the long term, retaining exiting customers and attracting new ones have become a major challenge to banks. In general, acquiring new customers can cost much more than retaining an existing customer. (Sharma and Panigrahi, 2011; Keramati et al., 2016; Ganesh et al. 2000; Liou et al., 2009).

This dissertation aims to analyze, using the panel longitudinal data statistic technique, the effect of implementing the churn rate indicator in the client portfolio of commercial managers of the bank, as an effective way to retain customers. This study also intends to investigate and identify factors that can affect customers retention. The data was collect from a customer base of a comercial bank operating in Brazil, in a period of two years, being one year before and the other after the implementation of the churn rate indicator.

The findings of this study, which will be presented in the next chapters, suggest that the implementation of the churn rate indicator, had a negative result effect in customer retention. Thus, this study contributes to the discussion around the effectiveness of using churn analysis as a customer retention strategy, as well as the use of the analytical tools for decision making.

KEYWORDS:

Churn analysis; Customer rentention; Relationship with customers.

Sumário

1. Introdução.....	13
2. Revisão da Literatura.....	16
2.1 <i>Churn</i>	16
2.2 Análise da Preditiva de <i>Churn</i>	20
2.3 Marketing de Relacionamento.....	23
3. Questões de Pesquisa e Desenvolvimento das Hipóteses.....	30
3.1 Definições das questões de pesquisa.....	30
3.2 Desenvolvimento das Hipóteses (Modelo Teórico).....	31
4. Metodologia e Design de Pesquisa.....	34
4.1 Objeto do Estudo.....	34
4.2 Amostragem e Base de Dados.....	34
4.3 Descrição das Variáveis.....	35
4.3.1 Variável Dependente.....	36
4.3.2 Variáveis Independentes.....	37
4.3.3 Variáveis de Controle.....	39
4.4 Técnica Estatística de Análise de Dados.....	44
4.5 Regressão Logística.....	45
4.6 Métricas de Avaliação.....	46
5. Análise e discussão de Resultados.....	49
5.1 Modelo.....	49
5.2 Resumo dos resultados.....	55
6. Conclusão, limitações e sugestões para pesquisas futuras.....	56
6.1 Principais conclusões.....	56
6.2 Limitações e sugestões para estudos futuros.....	57
Referências Bibliográficas.....	59

Lista de Figuras

Figura 1: Segmentação de clientes.....	15
Figura 2: Fórmula do cálculo da Taxa de <i>Chur</i>	16
Figura 3: Modelo de <i>framework</i> preditivo de <i>churn</i>	20
Figura 4: Modelo proposto para as hipóteses desse estudo.....	33
Figura 5: Curva ROC Regressões Logísticas.....	50
Figura 6: Curva ROC.....	54

Lista de Quadros

Quadro 1: Principais estudos relacionados à análise de <i>churn</i>	22
Quadro 2: 3 níveis do Marketing de Relacionamento.....	24
Quadro 3: Principais constructos do Marketing de Relacionamento com impacto em Fidelidade.....	28
Quadro 4: Variáveis dependentes, independentes e de controle.....	36
Quadro 5: Matriz de Confusão para um classificador binário.....	47
Quadro 6: Resumo da Matriz de Confusão e métricas para avaliação de modelos.....	47
Quadro 7: Matriz de Confusão dos resultados da Regressão Logística.....	54
Quadro 8: Resumo dos resultados das hipóteses.....	55

Lista de Tabelas

Tabela 1: Estatísticas descritivas da amostra.....	41
Tabela 2: Estatísticas descritivas das variáveis explicativas.....	42
Tabela 3: Estatísticas descritivas das variáveis de controle.....	42
Tabela 4: Estatísticas descritivas das variáveis binárias.....	43
Tabela 5: Estatísticas descritivas da variável dependente.....	43
Tabela 6: Estatísticas descritivas da variável dependente, comparativamente, antes e depois...44	
Tabela 7: Modelo – Hipóteses H1, H2, H3 e H4.....	51
Tabela 8: Resultado curva ROC.....	54

Sumário Executivo

O ambiente competitivo, uma consequência das novas tecnologias, desregulamentação e preocupação com os comportamentos complexos de consumidores, alterou a forma como produtos e serviços financeiros são ofertados e acessados pelos clientes. Nesse cenário, manter a vantagem competitiva sustentável ao longo prazo tornou-se um grande desafio, assim como a dificuldade em reter e conquistar novos clientes.

Com avanço da tecnologia, utilização de grandes bases de dados pelas empresas tornou-se acessível e permitiu a avaliação da efetividade da implementação de estratégias de retenção de clientes. Este estudo busca contribuir por meio da análise da efetividade da implementação do indicador da taxa *churn* na carteira de clientes de um banco comercial de médio porte, com atuação no Brasil.

O banco, que é o objeto de análise desta pesquisa, atua em todo território brasileiro, atendendo empresas com faturamento anual acima de R\$100 milhões por ano e oferecendo diversos produtos e serviços específicos para Pessoas Jurídicas (PJ).

Os principais motivadores deste estudo estão centrados na observação da baixa aplicação de métodos quantitativos com rigor estatístico para a tomada de decisão nos bancos de médio porte no Brasil e, sobretudo, pela escassez de estudos específicos sobre o tema.

Encontra-se na literatura diversas referências sobre análise de *churn* mas, na maioria das vezes, focada na comparação de modelagens preditivas de *churn*. Este estudo, em contrapartida, busca colocar luz sobre esta discussão por meio da avaliação de duas questões:

1. *A implementação do indicador da taxa de churn afeta a probabilidade de reter o cliente?*
2. *Quais fatores afetam a probabilidade de reter o cliente?*

O estudo foi estruturado a partir de uma base de dados que reúne dados internos da base de clientes do banco ao longo de dois anos, um ano antes e um ano após a implementação do indicador de *churn*. Os dados contêm informações sobre segmentos, gerentes, produtos e totalizam 35.256 observações.

A base de dados foi analisada utilizando o método de regressão de dados em painel, no qual foi definido como indivíduo do painel o cliente. De acordo com os resultados do modelo, a implementação do indicador de *churn* apresentou resultados negativos para a retenção dos clientes do banco.

Isto significa que, de acordo com os resultados obtidos no desenvolvimento desta pesquisa, tal implementação não cumpriu seu papel como instrumento para aumentar retenção de clientes. Contudo, em relação aos fatores que afetam essa probabilidade de retê-los, tanto a experiência do gerente, quanto a quantidade de clientes distribuídos por profissional, na presença das demais variáveis, não foram relevantes para explicar a probabilidade deste cliente tornar-se inativo.

A variável tempo de casa do gerente apresentou aumento do *churn* de clientes. Possíveis hipóteses para este resultado seriam a falta de treinamento, a falta de alinhamento do indicador com metas e os incentivos para os colaboradores do banco. A baixa utilização de análises quantitativas no planejamento das ações estratégicas e a falta de um modelo preditivo que possibilitasse ações comerciais e de marketing prévias sobre os clientes com maior probabilidade de encerrar o relacionamento com o banco é, também, considerada uma hipótese para o aumento do *churn* de clientes. O período no qual este estudo foi realizado demonstrou ter um impacto relevante no resultado apurado.

Com aplicação gerencial direta, este estudo contribui com a discussão do valor da utilização da análise de *churn* como estratégia de retenção de clientes, assim como a utilização de ferramentas analíticas para a tomada de decisão estratégica.

1. Introdução

Nos últimos anos, o setor de serviços bancários, até então dominado pelos bancos, passou por uma grande transformação. Em termos de organização do Sistema Financeiro Nacional (SFN), em 2018 havia 1.677 instituições autorizadas a funcionar pelo Banco Central do Brasil (BCB), 172 delas pertencentes ao segmento bancário, sendo 131 bancos comerciais ou múltiplos com carteira comercial. Em 2016, novos segmentos de instituições financeiras foram criados com a regulamentação das chamadas *fintechs* e o surgimento dos bancos digitais.

As novas tecnologias associadas à internet têm transformado o relacionamento entre clientes e empresas. Os investimentos no setor bancário em tecnologia somaram R\$ 118,7 bilhões entre 2012 e 2017, com média anual de R\$ 19,8 bilhões, segundo estudo publicado pela Federação Brasileira de Bancos (FEBRABAN). Assim, os clientes se tornaram mais exigentes e passaram a dar maior valor à experiência, ao atendimento personalizado, à agilidade e à diversidade. Por outro lado, as organizações concentraram-se no estudo e análise do comportamento destes clientes ao longo de suas jornadas, aumentando significativamente a importância das áreas de relacionamento.

Diante disso, a concorrência no setor tornou-se mais complexa e passou a exigir modelos de marketing cada vez mais sofisticados, com a finalidade de atrair e reter clientes de diferentes formas, ainda que os serviços e os produtos oferecidos fossem similares.

Com o objetivo de melhorar os processos de retenção, aquisição, fidelidade e rentabilidade dos clientes, diversas ferramentas de gestão de relacionamento são utilizadas. Um exemplo é o CRM (*Customer Relationship Management*), que consiste em um conjunto de processos de suporte ao negócio, incluindo técnicas de mineração de dados (*Data Mining*) para a descoberta de conhecimento a partir de uma enorme quantidade de dados (Nie *et al.* 2011) e *Machine Learning*, que, ao integrar as informações e suas interações em uma base de dados, é capaz de descobrir padrões para personalizar o atendimento, além das ofertas de produtos e serviços.

Essas técnicas são eficazes na identificação de clientes mais propensos a deixar empresas que atuam em setores de grande competitividade e baixo custo de troca (*switching*

cost). No setor bancário americano, por exemplo, estima-se que 30% da base de clientes seja vulnerável à migração (Accenture, 2009).

Quando aplicadas, possuem a capacidade de aumentar a percepção de qualidade do cliente, reduzindo sua probabilidade de “*churn*” e, com isso, aumentar o desempenho da organização. Contudo, são poucas as pesquisas que têm como foco o aumento de desempenho de bancos no mercado brasileiro por meio da implementação do indicador da taxa de *churn* (taxa de cancelamento ou de abandono registrada na base de clientes) e da análise preditiva de *churn*.

Diante do aumento da concorrência no setor bancário brasileiro, implementar a análise de *churn* ou gerenciamento de abandono e identificar as principais causas de evasão de clientes torna-se fundamental para a sobrevivência dos bancos brasileiros, uma vez que tanto as *fintechs*, como os bancos digitais, possuem modelos de negócios baseados em soluções tecnológicas e estruturas enxutas de menor custo.

Esta dissertação tem, portanto, o objetivo de responder às seguintes questões:

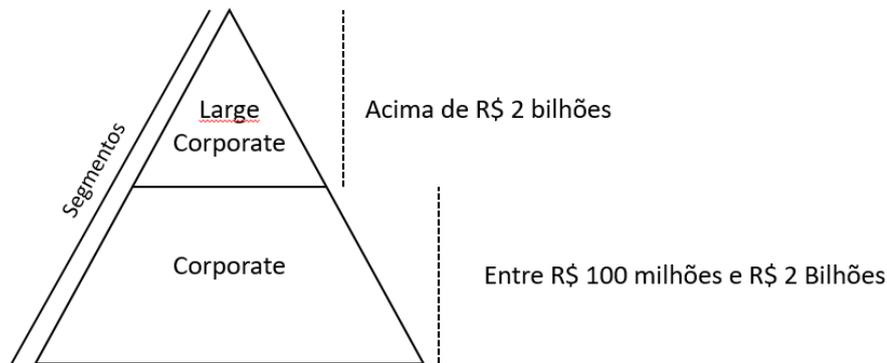
1. *A implementação do indicador da taxa de churn afeta a retenção dos clientes?*
2. *Quais são os fatores que afetam a probabilidade de reter o cliente?*

A partir da análise de dados disponibilizados antes e após a implementação da ferramenta, este estudo de caso apresenta como objeto central de pesquisa um banco comercial que atua há mais de 30 anos no Brasil, obtendo lucros consistentes neste período.

Como é possível observar na **Figura 1**, que apresenta a esquematização da segmentação dos clientes do banco comercial em questão, o objeto desta pesquisa possui uma carteira de crédito expandida, que engloba a carteira de empréstimos, garantias prestadas e títulos privados, com saldo de R\$ 28.7 bilhões em maio de 2020, operando com clientes corporativos, faixa de faturamento entre R\$ 100 milhões e R\$ 2 bilhões no segmento *Corporate* e acima de R\$ 2 bilhões no segmento *Large Corporate*.

Figura 1: Segmentação de Clientes

Empresas com Faturamento Anual:



Fonte: Formulada pelo autor (2020)

Diante das transformações de serviços bancários no Brasil e da necessidade de se utilizar ferramentas de relacionamento com o cliente e *Data Science*, abrangendo as perspectivas de *Data Analytics*, esse trabalho pretende contextualizar os pontos sensíveis à implementação da técnica da análise de *churn* em um banco comercial brasileiro.

O conteúdo está organizado em sete seções:

1. Introdução, na qual apresentar-se-á a motivação, as etapas e os objetivos do trabalho;
2. Revisão da literatura, na qual serão abordados de forma detalhada os referenciais teóricos necessários para o desenvolvimento deste estudo;
3. Questões de pesquisa e desenvolvimento das hipóteses testadas;
4. Metodologia e Design de Pesquisa, na qual explicar-se-á como e quais caminhos metodológicos foram utilizados no desenvolvimento desta pesquisa;
5. Análise e discussão de resultado, na qual foram reunidos os resultados quantitativos, em números, tabelas e gráficos, e as interpretações qualitativas de cada um deles;
6. Conclusões da pesquisa, suas possíveis limitações e sugestões para estudos futuros;
7. Referências bibliográficas.

2. Revisão da literatura

Este capítulo aborda os principais conceitos teóricos e a literatura relacionada com a questão que se pretende abordar e investigar neste trabalho. Além disso, demonstrar-se-á como estes conceitos relacionam-se entre si, criando o embasamento teórico necessário para a compreensão da metodologia elaborada e resultados obtidos, conforme será apresentado nos capítulos seguintes.

2.1 Churn

O termo “*churn*” refere-se ao resultado de uma estratégia de retenção sem sucesso. Segundo Sharma e Panigrahi (2011), *churning* é o cliente que deixa uma empresa, migrando para outra. Assim, *churn* define o cessar das trocas comerciais entre um cliente e uma empresa (Ascarza, *et al.*, 2017).

De acordo com o estudo realizado por Katelaris e Themistocleous (2017), que corrobora o estudo de Ascarza *et al.*, a dissolução (*churn*) ocorre justamente quando um consumidor cessa a relação com uma empresa para iniciar um relacionamento com uma instituição concorrente, caracterizando o processo de *churn*, que pode ser medido pela taxa de *churn* (*churn rate*).

Nesta taxa de *churn*, considera-se um KPI (*Key Performance Indicator*) que indica, em termos percentuais, a taxa de consumidores que terminaram a relação com a empresa. Este indicador é calculado pela divisão do número de consumidores que uma empresa perdeu ao longo de um determinado período de tempo pelo número total de consumidores nesse mesmo período (Lopes, Brito e Alves, 2012; Katelaris e Themistocleous, 2017).

Figura 2: Fórmula do cálculo da Taxa de *Churn*

$$\text{Taxa de Churn} = \frac{\text{Nº de consumidores que terminaram a relação}}{\text{Nº total de consumidores}} \times 100$$

Equação 1 – Fórmula do cálculo da Taxa de Churn

Fonte: Formulada pelo autor (2020)

Este método de medição de *churn* depende do conhecimento da organização sobre os clientes atuais e ativos, bem como as suas capacidades para calcular e saber concretamente o número de clientes propensos a abandonarem a relação (Axelsson e Notstam, 2017). A medida pode, ainda, indicar o nível de insatisfação, quais os melhores preços e condições praticados por competidores e outras razões relacionadas ao ciclo de vida do cliente, estejam elas ligadas a fatores externos (ofertas concorrentes) ou a fatores internos (empresa ou serviço apresentado).

Grande parte destas razões faz referência ao serviço apresentado, à experiência do consumidor (Keaveney, 1995) e à relação desenvolvida entre a organização e os seus clientes (Katelaris e Themistocleous, 2017). Se, por exemplo, o consumidor tiver uma experiência negativa ao se relacionar com a empresa ou se a empresa apresentar um mau serviço ao cliente, este fica insatisfeito, aumentando a probabilidade de abandonar a relação. Dessa forma, é possível concluir que as empresas devem estar atentas às necessidades dos seus clientes e lhes proporcionar experiências positivas para evitar o cessar desta relação.

Para Capraro, Broniarczyk e Srivastava (2003), além da satisfação dos consumidores ser uma das principais métricas para detectar e gerir a probabilidade de abandono dos clientes, o reconhecimento de marcas ou produtos concorrentes também tem um peso importante na previsão de *churn*.

Para chegar a essa conclusão, os pesquisadores supracitados analisaram a ligação entre o nível de reconhecimento de alternativas concorrentes por parte dos consumidores e o abandono da relação cliente-organização, deduzindo que a análise deste reconhecimento de concorrência permite perceber as razões pelas quais os consumidores insatisfeitos nem sempre abandonam a relação. Ou seja, apesar da insatisfação originar uma avaliação negativa da empresa, se estes consumidores não reconhecerem outras marcas ou produtos como concorrentes e possíveis substitutos, acabam por não cessar a relação, mesmo que ofertas rivais existam.

Além disso, Oliver (1999) afirmou que os consumidores fidelizados a uma marca são menos vulneráveis a ofertas concorrentes e detêm menor propensão ao reconhecimento das mesmas.

Em outro estudo que focaliza a temática em questão, Axelsson e Notstam (2017) defendem que, para prever e prevenir *churn*, as empresas devem considerar o tipo de lealdade apresentada pelos seus consumidores. De acordo com os autores, é possível reconhecer 18

diferenças entre clientes que, apesar de terem adquirido produtos da marca, ainda não se converteram em clientes envolvidos com a mesma, sendo leiais durante um longo período, mas, diminuindo a sua frequência de compra (Traynor, 2017).

Consequentemente, diante destas variáveis, o valor de retenção desses clientes também difere. Logo, esses dois segmentos devem ser abordados distintamente, com o objetivo de mantê-los ativos na relação com a organização.

Para as empresas que confiam nos seus clientes e os consideram um importante ativo, o término de relação e dissolução, ou seja, o processo de *churn*, é uma preocupação central e constante. Desde a década passada, à medida que os clientes se tornaram mais conscientes de outras ofertas e/ou oportunidades de mudança, as empresas têm demonstrado maior apreensão ao *churn*. Somada a esse fator, a transparência do mercado também contribuiu para que o processo de *churn* acontecesse mais fácil e constantemente (Holtrop, Wieringa, Gijzenberg e Verhoef, 2017).

Há, portanto, uma preocupação crescente e emergente que concerne às companhias, quando o assunto é a perda de clientes, que caracteriza o *churn* (Veloso, 2013).

Sendo os estudos de *churn* de clientes uma das áreas que tratam da retenção dos mesmos, a predição de *churn*, ou gerenciamento de *churn*, tem recebido bastante notoriedade em diversos âmbitos dos negócios, principalmente, em ambientes competitivos muito intensos, como as telecomunicações (Amin *et al.*, 2018; Zhao *et al.*, 2017; Tsai e Lu, 2009), as instituições financeiras (Van den Poel and Larivieri, 2004), o varejo (Yu *et al.*, 2011; Buckinx e Van den Poel, 2005), os seguros, os jornais, entre outros, devido ao fato óbvio de que a perda de clientes representa, diretamente, a perda de receita das empresas, sem contar o impacto na relação de custo já mencionada entre a aquisição e a retenção de clientes.

Diante deste cenário, metodologias e técnicas como as de *Data Mining* e *Machine Learning*, tornaram-se essenciais para a Inteligência de Negócios das empresas e, por isso, vêm sendo utilizadas com o objetivo de diminuir as taxas de *churn*.

Atualmente, a retenção de clientes apresenta um desafio significativo (Ascarza, *et al.*, 2017). Do ponto de vista dos consumidores, 85% dos clientes afirmam que as empresas deveriam dedicar mais esforços à retenção dos mesmos (Handley, 2013). No entanto, a partir da perspectiva empresarial, o foco das estratégias de retenção deve centrar-se na previsão de *customer churn* (Ascarza, *et al.*, 2017).

Ou seja, para as organizações, atrair novos consumidores é um fator positivo, mas, devido aos altos custos de aquisição destes novos consumidores, muitas empresas apostam, cada vez mais, no desenvolvimento de planos para reter os consumidores atuais, que já têm relação com as mesmas. Isso se deve ao fato de que o abandono da relação dos consumidores com as instituições (*churn*) as impacta e afeta os seus desempenhos, diminuindo a receita e as vendas, podendo, inclusive, construir uma imagem negativa das mesmas (Katelaris e Themistocleous, 2017).

Dessa forma, por meio da implementação da análise de *churn*, é pretendido diminuir a taxa de evasão dos clientes e, conseqüentemente, aumentar a retenção dos mesmos.

Halinen e Tähtinen (2002) defendem que, devido aos custos envolvidos, as organizações, de fato, não devem construir e desenvolver relações profundas com todos os seus consumidores. Na verdade, os recursos que são alocados para certos consumidores podem servir melhor quando considerados para outros fins, como para a dissolução da relação com clientes não rentáveis para a empresa.

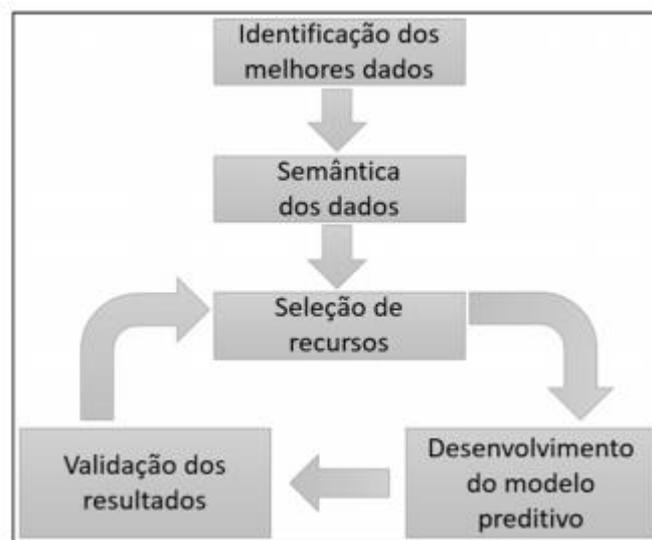
No entanto, de acordo com Dwyer, Schurr e Oh (1987), a empresa, ao desenvolver o relacionamento com o cliente, obtém informação útil sobre esse, ambos ganham confiança e um sentimento mútuo de compromisso. Isto permite que a organização possa agir de forma ponderada ao interagir com os seus clientes, mas, ainda assim, a possibilidade de abandono do relacionamento está implicitamente presente durante todo o ciclo de vida dessa relação.

Na revisão de literatura sobre *churn*, ficou evidente que as estratégias de retenção de clientes e de programas de fidelização apresentam um profundo impacto no negócio das organizações e são capazes de gerar elevado lucro (Singh e Khan, 2012). Por isso, pode-se concluir que a retenção de clientes tem como principal objetivo manter uma alta proporção de consumidores valiosos, reduzindo a taxa dos chamados consumidores “defeituosos”. Para cumprir essa finalidade, é necessário conhecer os clientes e focar nas atividades de marketing, uma vez que nem todos devem ser abrangidos, ou seja, retidos (Buttle, 2004).

2.2 Análise de Preditiva de Churn

Datta *et al.* (2001) apresentam um modelo composto por cinco estágios para a elaboração de um *framework* preditivo do *churn*, conforme demonstrado na **Figura 3**:

Figura 3: Modelo de *framework* preditivo de *churn*



Fonte: Datta *et al.* (2001)

Os estágios são assim definidos:

- a) **Identificação dos melhores dados:** inicialmente, é fundamental identificar os dados que melhor se adaptam ao tipo de análise que será executada e à finalidade pretendida.
- b) **Semântica dos dados:** neste estágio, entender-se-á o contexto dos dados inseridos em um banco de dados. É comum que os dados sejam difíceis de interpretar, por isso, essa é uma etapa importante do processo.
- c) **Seleção de recursos:** processo de identificação dos melhores campos para a previsão.
- d) **Desenvolvimento do modelo preditivo:** aqui, são desenvolvidas técnicas de modelagem preditivas, como Regressão Logística, Árvore de Decisão, Redes

Neurais, *Support Vector Machine* e *Random Forests*, cada uma com suas peculiaridades, limitações, restrições, entregas e vantagens.

- e) **Validação dos resultados:** neste estágio, vários métodos documentados podem ser utilizados para validar um modelo preditivo de *churn*.

A análise preditiva de *churn* usa variáveis de bancos de dados para identificar valores desconhecidos, enquanto a análise descritiva serve para identificar padrões que possam descrever os dados a serem interpretados pelos analistas.

Os estudos apresentados no **Quadro 1** têm em comum o propósito de entregar importantes informações gerenciais. Diante do caráter preditivo dos modelos testados, estes podem revelar os clientes cuja evasão pode ser considerada uma tendência.

Regressão Logística, Árvores de Decisão, Redes Neurais, SVM (*support vector machine*) e Análise de Sobrevivência são os métodos mais populares para essa finalidade (Buckinx e Van den Poel, 2005; Coussement 167 e Van den Poel, 2008; Karahoca e Karahoca, 2011; Neslin, Gupta, 168 Kamakura, Junxiang, e Mason, 2006).

Quadro 1: Principais estudos relacionados à análise de churn

Autor	Ano	Artigo	Setor	Metodologia de Pesquisa
Eiben <i>et al.</i>	1998	<i>Genetic modelling of customer retention</i>	Instituição Financeira	Regressão Logística, programação genética, análise de dados aproximados, CHAID
Larivière e Van den Poel	2004	<i>Investigating the role of product features in preventing customer Churn, by using survival analysis and choice modeling: The case of financial services</i>	Serviços Financeiros	<i>Hazard model (multinomial probit model)</i>
Larivière e Van den Poel	2005	<i>Expert Systems With Applications</i>	Instituição Financeira	Regressão Logística, Regressão Linear, Florestas Aleatórias e Florestas de Regressão.
Burez J., Van den Poel D.	2008	<i>CRM at a pay-TV company</i>	Banco, telecom, jornal, TV por assinatura e supermercado	Regressão Logística e Florestas Aleatórias
Kumar D.A., Ravi V.	2008	<i>Predicting credit card customer churn in banks using data mining</i>	Instituição Financeira	Regressão Logística e Árvores de Decisão, Redes Neurais, Máquina de Vetores de Suporte (SVM)
Botelho, D., Tostes, F. D.	2010	Modelagem de probabilidade de churn	Serviços Financeiro	Regressão Logística
Nie <i>et al.</i>	2011	<i>Credit card churn forecasting by logistic regression and decision tree</i>	Serviços Financeiros	Regressão Logística, Árvore de Decisão
Gur Ali <i>et al.</i>	2014	<i>Dynamic Churn prediction framework with more effective use of rare event data: The case of private banking</i>	Private Bank	Regressão Logística e análise de sobrevivência

Fonte: Elaborado pelo autor (2020)

A regressão logística é uma técnica que fornece as vantagens do poder preditivo e da compreensibilidade (Verbeke *et al.*, 2011; Risselada *et al.*, 2010; Martens *et al.*, 2011), mesmo quando contraposta a técnicas mais sofisticadas de mineração de dados (Coussement *et al.*, 2017). Além disso, é capaz de modelar relações lineares de forma eficaz (De Caigny *et al.*, 2018).

De acordo com Owczarczuk (2010), essa técnica é de fácil interpretabilidade, no entanto, como limitações, pode-se incluir as premissas rígidas de normalidade (Delen, 2010), desafios ao modelar efeitos interativos (De Caigny *et al.*, 2018) e relações não lineares entre o preditor e a variável de resultado (Zhang *et al.*, 2007; De Caigny *et al.*, 2018). Apesar de suas limitações, a regressão logística é uma técnica amplamente utilizada para modelar o *churn* em uma variedade de indústrias (Delen, 2010; Nie *et al.*, 2011, De Caigny *et al.* 2018). Por isso,

justifica-se a escolha da regressão logística como técnica de classificação selecionada para compor a metodologia de análise deste estudo.

2.3 Marketing de Relacionamento

O Marketing de Relacionamento surgiu com Berry, em 1983, no âmbito de Marketing de Serviços, quando o autor reconheceu a fidelização de clientes como forma de alcançar maior competitividade das empresas e maior satisfação dos clientes. Berry (1983) definiu o Marketing de Relacionamento como a atração, a manutenção e, em organizações multisserviços, o aumento de relacionamento com clientes.

O autor enfatizou que a atração de novos clientes deveria ser vista apenas como um passo intermediário no processo de marketing. Em contrapartida, solidificar relacionamentos, transformar clientes indiferentes em leais e servi-los é o que deveria ser considerado o cerne do Marketing de Relacionamento (Berry, 1995).

Há três níveis de Marketing de Relacionamento, estabelecidos em suas ações táticas (Berry, 1995). Esses níveis dependem da relevância dos benefícios do relacionamento para o cliente, que conduzem à fidelização, conforme demonstrado no **Quadro 2**.

No nível de relacionamento 1, são oferecidos **incentivos financeiros** para que o cliente permaneça fiel. É comum, nesse nível, a prática de programas de desconto para as compras repetidas.

No nível 2, o relacionamento se sustenta em **aspectos sociais** que dificultam o rompimento do relacionamento das partes envolvidas, como a personalização e a customização do relacionamento. Alguns exemplos são: comunicações regulares com os clientes, chamando-os pelos nomes, continuidade do serviço pelo mesmo representante de vendas e aumento do serviço básico por meio de atividades educacionais ou de entretenimento, como treinamentos, seminários, festas etc.

No nível 3, o Marketing de Relacionamento tem como base as **soluções estruturais** para importantes problemas dos clientes. Quando o Marketing de relacionamento oferece valores agregados, que são difíceis ou caros para o cliente, e que não são facilmente disponíveis em outro local, cria-se um forte fundamento para a manutenção e o incremento do relacionamento. Portanto, este nível pode ser entendido como situações de parcerias.

Quadro 2: 3 Níveis de Marketing de Relacionamento

Nível	Grau de customização	Potencial para vantagem competitiva sustentável	Exemplos
Financeiro	Baixo	Baixo	<i>American Airlines Advantage Program</i>
Social	Médio	Médio	<i>Harley Davidson's Harley Owners Group</i>
Estrutural	Médio para alto	Alto	<i>Federal Express Powership Program</i>

Fonte: Berry (1995)

O relacionamento com o cliente é uma ferramenta importante para o crescimento dos negócios, especialmente para o setor de serviços (Berry, 1995). Payne (2006) sugere três características distintas do Marketing de Relacionamento em relação ao Marketing Transacional; a primeira delas é a ênfase na retenção de clientes e a ampliação do “valor vitalício” dos clientes por meio de estratégias voltadas a retenção, a segunda é o reconhecimento de que as empresas precisam desenvolver relacionamento se quiserem obter sucesso no longo prazo, e a terceira característica é que o marketing precisa ser visto como uma responsabilidade de toda a empresa ou multifuncional.

Segundo Gaurav (2008), as empresas têm adotado o Marketing de Relacionamento a fim de assegurar que, por meio da lealdade, os clientes optem pelos mesmos produtos e serviços, ou seja, sejam fiéis à empresa. Já de acordo com Domazet (2010), o Marketing de Relacionamento foca em ganhar e reter a lealdade de clientes a longo prazo, na valorização do cliente, na qualidade do serviço oferecido, no comprometimento em atender as expectativas dos clientes e na preocupação de todos os colaboradores da empresa em promover o relacionamento com o cliente.

Ainda em pesquisa bibliográfica, de Barreto, Crescitelli e Figueiredo (2015), 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.

Portanto, capturar informações sobre as preferências, satisfações, hábitos, medos, entre outros aspectos relacionados ao cliente, torna-se fundamental para o desenvolvimento de um

relacionamento pautado na confiança, lealdade, familiaridade e, conseqüentemente, na fidelização.

Para Claro (2006), existe, ainda, no setor bancário, a preocupação em se diferenciar da concorrência, visto a facilidade de se replicar ofertas, produtos e serviços, o que torna o relacionamento com o cliente fundamental para alcançar a vantagem competitiva. O relacionamento é, portanto, o fator-chave de sucesso para a diferenciação da oferta no mercado, uma vez que a oferta de valor só é possível por meio do conhecimento adquirido a partir das informações citadas no parágrafo anterior.

A estratégia de Marketing de Relacionamento, dessa forma, possibilita a oferta contínua de valor superior, com melhores e mais adequados produtos e serviços, causando, nos clientes, a percepção de que estão recebendo mais do que estão, efetivamente, comprando, o que traz um grande potencial de benefícios tanto para o fornecedor, quanto para o cliente em questão.

Assim, o complexo gerenciamento do relacionamento com os clientes deixou de ser uma tarefa simples e passou a exigir das empresas ferramentas robustas de gestão.

Segundo Ribeiro (1997), as ações táticas da prática do Marketing de Relacionamento sustentada em bancos de dados devem permitir:

a) O diálogo com o cliente pelos meios utilizados para o *feedback*, como a análise dos padrões de *feedback* e dos históricos de compras, sistemas de atendimento ao consumidor, análise das cartas dos clientes; pesquisas formais e informais.

b) O aprofundamento na base de clientes por meio de programas de continuidade, de afinidade e de aumento de volume de vendas, como: *cross selling* (venda cruzada) ou *upgrade* (intensificação do volume, valor ou frequência da compra).

c) O trabalho com nichos de mercado com eficiência e lucro por meio das informações no banco de dados, que permitem encontrar nichos de mercado e viabilizar a atuação nesses nichos, vencendo a barreira do tamanho do mercado e da perda de economia de escala.

d) Maior controle sobre o canal de distribuição, oferecendo melhores informações sobre o público-alvo e, assim, permitindo melhor adequação da oferta.

e) Melhor **contabilização do marketing**, que se trata da capacidade de mensurabilidade do sistema e da capacidade de controle das variáveis que influenciam as respostas.

Em resposta às novas tecnologias, às demandas de mercado e à extensa base de dados, as empresas passaram a utilizar como ferramenta de gerenciamento de clientes o CRM (*Customer Relationship Management*). Para Payne (2005), o CRM é uma abordagem estratégica que se preocupa com a criação de melhor valor para o acionista por meio do desenvolvimento de relações adequadas com os principais clientes e segmentos de clientes.

O CRM une o potencial de estratégias de marketing de relacionamento e de TI (Tecnologia da Informação) para criar relacionamentos rentáveis de longo prazo com os clientes e outras partes interessadas. Essa ferramenta de gerenciamento de clientes fornece oportunidades aprimoradas para utilizar dados e informações que permitem entender os clientes e cocriar valores com eles, o que exige interação cross-funcional de processos, pessoas, operações e recursos de marketing.

Dentro das organizações, o CRM tem como principal objetivo satisfazer e criar relações de longo prazo com os clientes (Vasiliu, 2012). É, portanto, considerado, a “ponte estratégica entre a tecnologia da informação e as estratégias de marketing destinadas a construir relacionamentos de longo prazo e lucratividade” (Ryals e Payne, 2001 p.3).

Ainda de acordo com Payne (2006), trata-se da sistematização de ações interativas relacionais e transacionais para fortalecer o vínculo com clientes e, naturalmente, rentabilizar a relação pela percepção de maior entrega de valor, o que contribui para a elevação da margem de contribuição do cliente e a extensão do ciclo de relacionamento.

Portanto, dado o ambiente competitivo e a concorrência agressiva em diversos setores, como no setor bancário, o CRM é fundamental, uma vez que os bancos precisam manter seus clientes existentes e, simultaneamente, atrair novos clientes potenciais. Vale lembrar que uma das virtudes do CRM é sua competência preditora, que propicia a implementação de estratégias de retenção com o objetivo de blindar a evasão (*churn*) de clientes rentáveis.

A gestão do relacionamento com o cliente, tema em que se enquadra este estudo, é responsável, entre outras atividades, por ações que previnam atrito ou o abandono do cliente.

Dessa forma, uma das questões centrais que incitaram o desenvolvimento deste trabalho é: *Quais os fatores que afetam significativamente a probabilidade de reter o cliente?*

Sob a perspectiva do Marketing de Relacionamento, conceito definido neste item da revisão bibliográfica, um dos fatores-chaves relativos à ocorrência do *churn* é a lealdade e fidelidade do cliente. 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, uma vez que o Marketing de Relacionamento não possui relação direta com o CLV.

O Marketing de Relacionamento contribuiu, portanto, para a elevação dos níveis de fidelidade, sendo a fidelidade em si diretamente relacionada ao CLV.

O **Quadro 3**, apresentado a seguir, consolida o conceito dos constructos abordados por Barreto, Crescitelli e Figueiredo (2015), que segmentam os elementos iniciais do Marketing de Relacionamento com impacto direto sobre a fidelidade em: Satisfação, Qualidade, Comprometimento, Confiança, Personalização e Custos de Mudança.

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).

Com base na conexão entre Marketing de Relacionamento, CRM e CLV para auxiliar as empresas na tomada de decisões que envolvem alocação de recursos e estratégia, é importante ter uma tipologia de valor do cliente.

Segundo Kumar *et al.* (2008), é essencial que as ferramentas de CRM, por meio do CLV, amparem decisões consistentes ao longo do tempo de acordo com as seguintes ponderações:

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

O CLV contribui com informações consistentes a partir da perspectiva de clientes a serem considerados ativos e de longo prazo. Segundo Payne (2006), é importante que a empresa tenha uma visão clara da lucratividade gerada por seus principais clientes e segmentos dos mesmos, considerando, principalmente, o potencial de cada cliente em relação à rentabilidade futura.

3. Questões de Pesquisa e Desenvolvimento das Hipóteses

O presente capítulo tem como objetivo apresentar as questões de pesquisa e o desenvolvimento das hipóteses. Primeiramente, na **seção 3.1**, serão expostas as questões de pesquisa e, em seguida, apresentadas as hipóteses.

3.1 Definição das Questões de Pesquisa

Diante da importância do gerenciamento de *churn* para as empresas, o presente estudo tem como objetivo principal investigar a efetividade da implantação do indicador da taxa de *churn* como instrumento de gerenciamento da carteira de clientes do banco e os principais fatores que afetam a probabilidade de reter este cliente. Levando-se em conta as oportunidades encontradas na literatura, e com base na revisão bibliográfica apresentada no capítulo anterior, definimos as seguintes questões de pesquisa:

Questão de Pesquisa Principal: *A implementação do indicador da taxa de churn, afeta a retenção dos clientes?*

Sub-questão de Pesquisa: *Quais fatores afetam a probabilidade de reter o cliente?*

O presente estudo visa, portanto, identificar o impacto da implementação do indicador da taxa de *churn* e o possível efeito de algumas características dos gerentes dos bancos que possam influenciar a probabilidade de retenção dos clientes.

3.2 Desenvolvimento das Hipóteses (Modelo Teórico)

Para responder à principal questão de pesquisa desse estudo e com a finalidade de verificar o impacto da implementação do indicador da taxa de *churn*, foi avaliado o desempenho do modelo proposto a seguir.

Com base nas informações obtidas na revisão bibliográfica, apresentada no capítulo anterior, a implementação do gerenciamento do *churn*, possibilita a melhoria do conhecimento sobre a probabilidade de um cliente abandonar a sua relação com o banco (Van den Poel, D., e Larivière, B, (2004), sendo que a utilização apropriada deste indicador pode ser capaz de aumentar seus lucros em até 85% devido à melhoria de 5% na taxa de retenção (Reichheld e Sasser,1990).

A retenção bem-sucedida reduz a necessidade de buscar clientes novos e potencialmente de maior risco, além de permitir que as organizações se concentrem com mais precisão nas necessidades de seus clientes já existentes, construindo e potencializando os relacionamentos (Dawes e Swailes, 1999; Engel *et al.*, 1995).

Em outras palavras, a implementação e, conseqüente monitoramento da taxa de churn deveria aumentar a atenção dedicada pelos gerentes para atuar de forma mais proativa e evitar a saída e, conseqüentemente, a perda de clientes.

Portanto, diante dos argumentos encontrados na literatura sobre a importância da implementação de estratégias bem-sucedidas de retenção, é proposta a seguinte hipótese de pesquisa:

H1: *A implementação do indicador da taxa de churn apresenta impacto negativo no churn de clientes.*

Todavia, características individuais e específicas dos gerentes e de sua atuação podem afetar o relacionamento com os clientes e, conseqüentemente, sua retenção. Referente às variáveis relacionadas aos intermediários, ainda há, até o presente momento, uma compreensão insuficiente do impacto de vendedores, ou intermediários, sobre o comportamento dos clientes (Guenzi e Pelloni, 2004). No entanto, esse é um aspecto importante para investigar o papel do

gerente de relacionamento, uma vez que ele atua como o jogador principal e crucial na interação com os clientes do banco.

Neste estudo, investigamos três variáveis relacionadas a esses gerentes de relacionamento: a experiência, o tempo de casa de cada gerente e o número de clientes atendidos por cada um deles.

Quanto maior a experiência do gerente, o esperado é que o mesmo contribua com melhores resultados para a instituição financeira. Gerentes mais experientes, possuem, portanto, carteiras de clientes mais relevantes para os bancos, apresentando, também, maior conhecimento técnico de produtos e serviços bancários. Diante do exposto, é proposta a segunda hipótese desta pesquisa:

H2: *Quanto maior a experiência do gerente de atendimento, menor o churn de clientes.*

O banco, que é objeto de pesquisa do presente estudo, sofreu com a evasão de cliente nos últimos anos. Em estudo realizado em 2018, por uma empresa de consultoria externa, somente 28% da base de clientes fez negócios recorrentes nos últimos 4 anos. Portanto, é preciso investigar se gerentes com mais tempo de casa, de fato, possuem maior propensão a propiciar a retenção do cliente, o que nos leva à terceira hipótese dessa pesquisa:

H3: *Quanto maior o tempo de casa do gerente, menor o churn clientes.*

Muitos dos mecanismos subjacentes relacionados à interação entre vendedores e clientes ainda não são totalmente claros (Van Dolen, Lemmink, Ruyter e Jong, 2002). No cenário de serviços financeiros, é plausível acreditar que alguns clientes experimentam menos atenção, quando um gerente de relacionamento atende uma grande base e, como consequência, não consegue conhecer cada cliente detalhadamente.

Durante o desenvolvimento deste estudo, foram consideradas essas informações com a intenção de se investigar o seu impacto no comportamento na evasão dos clientes. A quantidade de clientes por gerente, indica, *a priori*, um esforço do banco em distribuir uma carteira

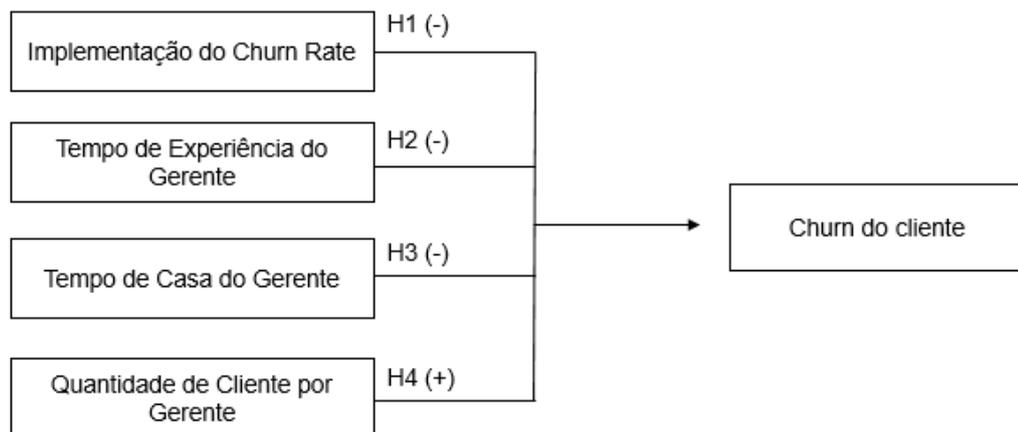
balanceada, que seja gerenciável e permita que os profissionais possam desenvolver um relacionamento mais ajustado com os seus clientes.

É esperado que, quanto menor a quantidade de clientes por gerente, mais tempo este profissional terá para se aprofundar no relacionamento com cada cliente, reduzindo, assim, o *churn*, o que nos leva à quarta hipótese desta pesquisa:

H4: *Quanto maior a quantidade de clientes por gerente, maior o churn de clientes.*

Diante do exposto neste item de desenvolvimento das hipóteses que guiaram o desenvolvimento metodológico desta pesquisa, tem-se o modelo proposto, conforme apresentado na **Figura 4**:

Figura 4: Modelo proposto para as hipóteses deste estudo



Fonte: Formulada pelo autor (2020)

4. Metodologia e Design de Pesquisa

Neste capítulo, será apresentada a base de dados utilizados para o desenvolvimento deste estudo, além da metodologia de investigação. Primeiramente na seção 4.1, será apresentado o objeto do presente estudo e as hipóteses que serão testadas. No item 4.2 tem-se acesso a toda informação relativa aos dados e suas fontes e, por fim, no item 4.3 deste capítulo, é apresentada a descrição detalhada do conjunto de variáveis que compõe o modelo a estimar.

Serão, ainda, apresentadas as especificidades relativas à metodologia de estimação, os pressupostos e implicações subjacentes, assim como a forma do modelo a estimar.

4.1 Objeto do Estudo

O objeto do estudo será o *churn* de clientes da carteira de um banco comercial localizado em São Paulo, a partir da análise do impacto da implementação de uma ferramenta de gestão que mede a taxa de *churn* da carteira de clientes em dois momentos distintos: (a) antes, (b) depois da implementação e da investigação das características dos gerentes comerciais do banco.

4.2 Amostragem e Base de Dados

A base de dados do presente estudo de caso é derivada de um banco comercial que, há mais de 30 anos, atua no Brasil, concedendo crédito e ofertando operações de derivativos, câmbio, *trade finance*, garantias nacionais e internacionais e mercado de capitais e investimento para empresas de grande porte, com faturamento anual acima de R\$ 100 milhões ao ano, podendo ultrapassar R\$ 2 bilhões e localizadas na cidade de São Paulo.

O universo da pesquisa constitui-se, portanto, da base agregados a nível de cliente. Cada cliente possui um código específico na base de dados do banco e seus dados contemplam uma posição mensal histórica para cada uma das variáveis coletadas ao longo de 24 meses, no período de abril de 2018 a abril de 2020, totalizando 36.725 observações/cliente.

O período mensal da coleta de dados foi definido de acordo com o tempo mínimo a ser considerado no ciclo de apuração de metas e acompanhamento de resultados do banco.

Foram retirados da análise os clientes que permaneceram com quantidade de produtos e operações constantes ao longo de todo período considerado neste estudo. Dessa forma, evitou-se o enviesamento da amostra, que poderia ocorrer porque alguns produtos possuem apuração de resultado *accrual* (ao longo da operação), não sendo possível detectar o *churn* do cliente. Ao todo, foram excluídos 32 clientes.

Após a seleção da amostra coletada para análise, que contabilizou 1469 clientes, foram identificados 27.358 (77,6%) eventos nos quais os clientes deixaram de operar com o banco em algum momento no intervalo de dois anos. Os demais 7.898 eventos (22,4%) representam, por meio de transações com o banco, a classificações de clientes que permaneceram ou passaram a ser clientes ativos em algum momento neste período, o que permite realizar a comparação da taxa de *churn* e demais análises pertinentes a este estudo.

O acompanhamento da taxa de *churn* foi realizado ao longo do período supracitado, totalizando um ano antes e um ano depois da ocorrência do evento da implementação do indicador da taxa de *churn* na carteira de clientes. Este procedimento permite avaliar o impacto desta implementação da taxa de *churn* na amostra.

Por fim, afirma-se que os dados foram coletados a partir da utilização do sistema SAS *Enterprise Guide*, um software para a coleta e o armazenamento de dados. Todas as informações obtidas a partir das operações realizadas entre o banco, objeto de análise do presente estudo, e seus clientes foram disponibilizadas em planilhas eletrônicas.

4.3 Descrição das variáveis

No **Quadro 4**, apresentado abaixo, estão contidas as variáveis utilizadas no estudo:

Quadro 4: Variáveis dependentes, independentes e de controle

Tipo de Variável	Variável	Descrição	Formato/Unidade
Dependente	ProxyChurn	Identifica o churn do cliente no mês	Dummy (0 = Não Churner; 1 = Churner)
Independente	IChurnRate	Identifica a Implementação da ferramenta (Churn Rate)	Dummy (0 = Antes; 1 = Depois)
Independente	Experiência.G	Tempo de experiência do gerente em meses	Número positivo
Independente	TempoCasa.G	Tempos de casa do gerente em meses	Número positivo
Independente	Qtde.Cliente.G	Quantidade de cliente por gerente por mês	Número positivo
Controle	Gênero.G	Identifica o gênero do gerente	Dummy (0 = Homem, 1 = Mulher)
Controle	Qtde.Total.G	Quantidade total de gerentes no banco por mês	Número positivo
Controle	Segmento.C	Identifica o segmento do cliente	Dummy (0 = Corporate; 1 = Large Corporate)
Controle	Qtde.Operações	Quantidade de operações do cliente por mês	Número positivo
Controle	Qtde.Produutos	Quantidade de produtos do cliente por mês	Número positivo
Controle	Rec.Judicial	Pedidos de Recuperação Judicial no mês - SERASA	Número positivo
Controle	Falência	Pedidos de Falência Requeridos no mês - SERASA	Número positivo
Controle	USDVar	Variação do dolar real médio por mês	Dummy (0 = Corporate; 1 = Large Corporate)
Controle	A_2018	Identifica o ano de 2018	Dummy (1 = é 2018; 0 = não é 2018)
Controle	A_2019	Identifica o ano de 2019	Dummy (1 = é 2019; 0 = não é 2019)
Controle	A_2020	Identifica o ano de 2020	Dummy (1 = é 2020; 0 = não é 2020)

Fonte: Elaborado pelo autor (2020)

Foram utilizadas três variáveis independentes, sendo duas contemplando características individuais dos gerentes (**Experiência.G** e **TempoCasa.G**) e uma que caracteriza o esforço do banco em atribuir uma carteira adequada para cada gerente (**Qtde.Cliente.G**). As variáveis de controle são formadas por *dummies* de ano, variáveis com características dos gerentes (**Gênero.G**), clientes (**Segmento.C**, **Qtde.Operações** e **Qtde.Produutos**), mercado (**USDVar**) e conjuntura econômica (**Rec.Judicial** e **Falência**).

A seguir, apresentar-se-á uma descrição mais detalhada da variável dependente, de cada variável independente e de controle.

4.3.1 Variável Dependente

A partir das informações obtidas na etapa metodológica da revisão bibliográfica, apresentada no primeiro capítulo dessa dissertação, recorrer-se-á à regressão logística, uma

técnica estatística clássica utilizada para classificação, sendo especificamente adaptada à situação em que a variável dependente é dicotômica (ou binária), para que sejam feitas as estimativas que interessam a esta pesquisa.

De acordo com Gujaratti (2006), a variável dependente é expressa como uma função de uma ou mais variáveis, chamadas de variáveis explicativas. Em sua conceituação, essa variável objetiva medir a desidratação de clientes.

Para alcançar maior aderência ao contexto bancário, no qual situa-se o presente estudo de caso, foram realizados alguns ajustes nesta variável dependente, uma vez que o cliente que possui relacionamento bancário, normalmente possui, além da conta corrente, outros produtos.

Mesmo que o objetivo central deste estudo seja identificar os casos de clientes que realizaram o *churn* total, fez-se, também, um esforço para mapear as ocorrências de *churn* parcial, pois, além de mais comuns, já sinalizam a perda de negócios, de acordo com os conceitos de Hadden *et al.* (2007).

Dessa forma, construímos como marcador a variável dependente (**Y**), que expressa a desidratação a ser analisada, definida pela variável **ChurnProxy**, representada pela quantidade de produtos e quantidade de operações, sem distinção de peso.

Foi definido como *churner* o cliente que reduziu ou manteve a quantidade de produtos e/ou operações contratadas em relação ao mês anterior, durante o período de observação proposto, e não *churner* o cliente que aumentou a quantidade de produtos ou operações em relação ao mês anterior.

A variável dependente (**Y**), de forma categórica, foi expressa conforme abaixo:

$$Y_t = \begin{cases} 0, & \text{não churn} \\ 1, & \text{churn} \end{cases}$$

4.3.2 Variáveis Independentes

As variáveis independentes procuram contribuir para explicar o comportamento de *churn* dos clientes do banco, durante o período analisado. Para mitigar ocorrências de multicolinearidade entre as variáveis independentes, fez-se a verificação de correção entre as variáveis, eliminando da análise as variáveis que apresentavam mais de 81% de correlação com alguma outra variável. A seguir, apresentar-se-á as variáveis independentes.

(a) Implementação da Taxa de *Churn*

Trata-se da ferramenta de gestão que possibilita aos gerentes o acesso à informação, incentivando à ação de retenção de seus clientes. Foi utilizada uma variável *dummy*, indicando o momento de implementação do indicador da taxa de *churn*. Desta forma, adotou-se o valor (0) para os meses de abril 2018 a abril 2019 e o valor (1) para os meses de maio 2019 a abril 2020.

(b) Experiência do Gerente

Esta variável corresponde à idade, em meses, dos gerentes. Para que não fosse classificada como uma variável fixa, a mesma foi coletada no início do período do estudo, sendo recalculada a cada mês.

Foram utilizadas as respectivas idades dos gerentes como *proxy* de experiência, uma vez que não foi possível o acesso ao currículo de cada gerente, sendo esperado que os gerentes, ao adquirirem mais experiência, com o passar do tempo, tenham relacionamentos mais estreitos com seus clientes.

(c) Tempo de Casa do Gerente

Esta variável representa o tempo de casa, ou seja, a quantidade de meses trabalhados por cada gerente no banco, que é o objeto do presente estudo, durante o período analisado.

É esperado que, quanto maior o tempo de casa, mais adaptados à cultura, às metas e aos incentivos do banco os gerentes estejam, bem como mais aptos a utilizar esta estrutura de modo a produzir melhores resultados.

(d) Quantidade de Clientes por Gerente

Trata-se da quantidade de clientes da carteira gerenciada por cada profissional do banco. Essa variável indica o esforço do banco em manter uma carteira de clientes gerenciável,

permitindo que seja desenvolvido um relacionamento mais ajustado e, assim, apresentando menos propensão ao *churn*.

4.3.3 Variáveis de controle

Como forma de compreender outros fatores que influenciam o comportamento de *churn* dos clientes, foram utilizadas as variáveis de controle, que podem afetar diretamente tal ação e que não estão, necessariamente, associadas à implementação do indicador da taxa de *churn* ou às características dos gerentes. As variáveis de controle consideradas foram:

(a) Gênero do Gerente

Variável dicotômica que indica se sexo do gerente do banco é masculino (0) ou feminino (1).

(b) Quantidade total de Gerentes

Essa variável representa o número total de gerentes do banco analisado e indica uma possível preocupação da instituição em oferecer recursos para atender os clientes atuais. Simultaneamente, a análise dessa variável de controle permite-nos avaliar que o aumento da base de gerentes representa a preocupação em capturar novos clientes no mercado.

(c) Segmento do Cliente

Variável dicotômica indicando a classificação dos clientes por segmento, podendo ser *Corporate* (0), empresas com faturamento entre R\$100 milhões a R\$2 bilhões ao ano, ou *Large Corporate* (1), empresas com faturamento acima de R\$ 2 bilhões ao ano.

(d) Quantidade de Operações

Essa variável de controle indica o número de operações realizadas pelo cliente ao mês, sendo considerada uma *proxy* de relacionamento. Mais operações mensais, ou a variação positiva da quantidade de operações, indica maior relacionamento com o banco.

A variável foi construída utilizando-se como referência a quantidade de operações do mês anterior por sua influência no comportamento de *churn* ou não *churn* do cliente no mês seguinte.

(e) Quantidade de Produtos

Essa variável de controle indica o número de produtos consumidos pelo cliente no mês, sendo considerada uma *proxy* de relacionamento. Mais produtos consumidos mensalmente, ou a variação positiva da quantidade de produtos, indica maior relacionamento com o banco.

A variável foi construída utilizando-se como referência a quantidade de produtos consumidos no mês anterior, por sua influência no comportamento de *churn* ou não *churn* do cliente no mês seguinte.

(f) Pedidos de Recuperação Judicial e Falência

Essas variáveis de controle tentam individualizar ao máximo o *churn* involuntário, abrangendo todos os tamanhos de empresas. É válido considerá-la, uma vez que o aumento de empresas em recuperação judicial e falência impacta a demanda por produtos bancários.

(g) Variação Cambial

A variação cambial do real em relação ao dólar americano foi escolhida como variável de controle pois impacta a demanda por produtos bancários.

(h) Controle de Ano

Essa variável pretende controlar efeitos macroeconômicos e outras mudanças que possam afetar todas as variáveis devido a fatores temporais em cada um dos anos, 2018, 2019 e 2020, período estipulado para as análises desta pesquisa.

Na **Tabela 1**, apresentada abaixo, podem ser observadas as estatísticas descritivas da amostra utilizada para o desenvolvimento deste estudo de caso.

Tabela 1: Estatísticas descritivas da amostra

Variável	Observações	Média	Desvio Padrão	Mínimo	Máximo
Experiência.G	36725	511,82	81,89	324	732
TempoCasa.G	36725	74,83	44,65	0	324
Qtde.Cliente.G	36725	35,81	11,02	3	57
Qtde.Total.G	36725	88,12	7,85	78	103
Qtde.Operações	36725	2,51	4,23	0	216
Qtde.Produutos	36725	1,22	0,97	0	10
Rec.Judicial	36725	11,48	23,88	73	176
Falência	36725	28,72	28,72	60	171
USDVar	36725	2,16%	5,36%	-10,05%	15,56%

Fonte: Elaborada pelo autor a partir do *software* Stata e banco de dados da empresa estudada (2020)

Na **Tabela 2**, apresentada abaixo, é possível observar que a experiência média dos gerentes aumentou apenas de acordo com a evolução natural do tempo, treze meses, quando comparados os períodos pré e pós implementação do indicador da taxa de *churn*, assim como o tempo de casa destes profissionais,. A quantidade de clientes por gerente, no entanto, permaneceu constante.

Tabela 2: Estatísticas descritivas das variáveis explicativas

Variáveis	Antes		Depois	
	Média	Desvio Padrão	Média	Desvio Padrão
Experiência.G	505	0,64	518	0,59
TempoCasa.G	68	0,35	81	0,32
QtdeCliente.G	36	0,09	36	0,08

Fonte: Elaborada pelo autor a partir do *software* Stata e banco de dados da empresa estudada (2020)

Já na **Tabela 3** é possível observar que o banco aumentou a quantidade de gerentes após a implementação do indicador da taxa de *churn*, o que sugere uma estratégia para o crescimento da base de clientes novos ou a preocupação em prestar um melhor atendimento aos clientes ativos.

Na média, a quantidade de operações mensais dos clientes sofreu redução na comparação entre os dois períodos. Já quantidade de produtos consumidos permaneceu estável, demonstrando uma baixa penetração de *cross-sell* entre os clientes.

Tabela 3: Estatísticas descritivas das variáveis de controle

Variáveis	Antes		Depois	
	Média	Desvio Padrão	Média	Desvio Padrão
QtdeTotal.G	82	0,20	94	0,00
QtdeOperações	3	0,03	2	0,33
QtdeProdutos	1	0,01	1	0,01

Fonte: Elaborada pelo autor a partir do *software* Stata e banco de dados da empresa estudada (2020)

Na **Tabela 4**, apresentada abaixo, podem ser observadas as estatísticas descritivas categóricas.

Tabela 4: Estatísticas descritivas das variáveis binárias

Variável	Total	Frequência (0)	Frequência (1)	Percentual (0)	Percentual (1)
ProxyChurn	35256	7898	27358	22,40%	77,60%
IChurnRate	35256	17628	17628	50,00%	50,00%
Gênero.G	35256	32064	3192	90,95%	9,05%
Segmento.C	35256	30504	4752	86,52%	13,48%

Fonte: Elaborada pelo autor a partir do *software* Stata e banco de dados da empresa estudada (2020)

De acordo com as análises das variáveis dependentes, concluiu-se que a maior quantidade de clientes é do segmento *corporate* (0), que contempla 86,52% da base de clientes. A área comercial do banco é composta, predominantemente, por gerentes comerciais homens, que representam 90,95% do total de gerentes.

O período pré e pós implementação do indicador da taxa de *churn*, representado pela variável IChurnRate, é composto pela mesma quantidade de observações. Por fim, a *proxy* de relacionamento, ProxyChurn, apresentou uma quantidade relevante de clientes desidratados (77,60%). A seguir, na **Tabela 5** e na **Tabela 6**, é possível analisar as características da variável dependente detalhadamente.

Tabela 5: Estatísticas descritivas da variável dependente

Proxy.Churn	Frequência	%
0	7.898	22,4
1	27.358	77,6
Total	35.256	100

Fonte: Elaborada pelo autor a partir do *software* Stata e banco de dados da empresa estudada (2020)

Tabela 6: Estatísticas descritivas da variável dependente, comparativamente, antes e depois

Antes/Depois	Proxy.Churn		Total
	0	1	
0	4.293	13.335	17.628
1	3.605	14.023	17.628
Total	7.898	27358	

Fonte: Elaborada pelo autor a partir do *software* Stata e banco de dados da empresa estudada (2020)

É possível observar na **Tabela 6** que, no período após a implementação do indicador da taxa de *churn* (1), houve um aumento na quantidade de clientes inativos e uma diminuição na quantidade de clientes ativos (0), indicando que a implementação do indicador da taxa de *churn* não resultou efetivamente no aumento da retenção dos clientes.

4.4 Técnica Estatística de Análise de Dados

Este estudo trata-se de um caso longitudinal com dados transversais coletados em dois períodos, antes e após a implementação do indicador de *churn* na carteira de 1469 clientes de um banco comercial com atuação no Brasil.

Para a análise dos dados, utilizou-se a técnica econométrica em painel, pela estrutura de coleta de dados, que consiste em uma série temporal para cada registro de corte transversal do conjunto de dados. Este modo de estruturar os dados longitudinais mostra-se necessário, diante da ordenação cronológica das observações, uma vez que busca analisar o comportamento mensal dos clientes ativos, ao longo de 24 meses, antes e depois da implementação do indicador da taxa de *churn*.

A escolha do método painel permite a avaliação dos efeitos das variáveis independentes na variável dependente, acompanhando o cliente ao longo do tempo, com o objetivo de identificar as suas características únicas.

4.5 Regressão Logística

Como apresentado anteriormente, nesta dissertação, pretende-se investigar o impacto da implementação do indicador da taxa de *churn* na evasão de clientes (**H1**) e quais fatores afetam significativamente a probabilidade de reter o cliente (**H2**, **H3** e **H4**).

Para a análise de **H1**, **H2**, **H3** e **H4** utilizou-se, como caminho metodológico, a Regressão Logística Multinível Longitudinal. A Regressão Logística é um recurso que permite estimar a probabilidade associada à ocorrência de determinado evento, em face a um conjunto de variáveis explanatórias (Levine *et al.*, 2008). A Regressão Logística auxilia, também, na determinação de quais características levam a uma determinada escolha.

Berry e Linoff (2011) ressaltam a importância de modelar o *churn* com resultado binário, uma vez que, modelos preditivos de *churn* com desfecho binário, como o que se apresentará nos capítulos seguintes deste estudo, tem um horizonte de tempo definido, gerando melhores resultados da pesquisa. Dessa forma, os dados do histórico do cliente são combinados a um marcador, a variável dependente, que o define como ativo ou não ($Y=0$, se o cliente for ativo; $Y=1$, se o cliente for inativo).

Outra vantagem da Regressão Logística é sua abordagem probabilística, já que essa regressão estima a chance de ocorrer certo evento, a partir de uma série de variáveis independentes ou explanatórias. Segundo Hosmer e Lemeshow (2000), o principal objetivo da Regressão Logística é achar o melhor relacionamento entre a variável resposta (variável dependente) e um conjunto de variáveis explicativas ou preditivas, sendo o modelo final aquele que apresentar o melhor ajuste e for, naturalmente, razoável de se explicar.

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

- a) A modelagem é bem conhecida, conceitualmente simples e frequentemente usada em marketing, especialmente no nível do consumidor final.
- b) Facilidade de interpretação do logit.
- c) A modelagem logit mostrou bons e robustos resultados em estudos gerais preditivos, o que é corroborado pelo referencial teórico.

- d) Mais especificamente, nas abordagens de métricas de 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.

No modelo de Regressão Logística com efeitos mistos, aplicado no presente estudo, os dados estão aninhados em dois níveis distintos. Inicialmente temos 35.256 observações que compõem o nível 1 (mudanças relacionadas aos clientes ocorridas durante o período). Neste nível, a dependência entre as observações ocorre em função do mesmo cliente ser medido “n” vezes ao longo do tempo. Já no nível 2, 1.469 grupos (clientes), é verificada quais mudanças individuais ao longo do tempo diferem entre clientes em função de suas características fixas, o que caracteriza a estrutura de dados agrupados em dois níveis.

No nível 1, a trajetória temporal do evento foi acompanhada pela inclusão da variável Tempo associada à ocorrência do *churn* entre os meses do estudo. Com essa inclusão é possível fazer o diagnóstico do padrão temporal do *churn* de clientes.

De acordo com West, Welch e Gałecki (2015), a denominação modelos lineares mistos vem do fato de que esses modelos apresentam especificação linear e as variáveis explicativas envolvem um misto de efeitos fixos e aleatórios, ou seja, podem ser inseridas tanto em componentes de efeitos fixos, quanto em componentes de efeitos aleatórios. Enquanto os parâmetros estimados de efeitos fixos indicam a relação entre as variáveis explicativas e a variável dependente métrica, os componentes de efeitos aleatórios podem ser representados pela combinação de variáveis explicativas e termos aleatórios não observados.

4.6 Métricas de Avaliação

A avaliação da qualidade preditiva do modelo pode ser realizada por meio de medidas como acurácia, sensibilidade e especificidade. Uma maneira eficaz de se avaliar os erros e acertos cometidos pelo modelo em questão é por meio da matriz de confusão, que consiste em

uma matriz na qual os dados reais são interpostos aos valores preditos, conforme posso ser observado no **Quadro 5**:

Quadro 5: Matriz de Confusão para um classificador binário

Matriz de confusão		Previsto pelo Modelo	
		Não churner (N)	Churn
Situação Real	Não churner (N)	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	Churner (P)	Falso Negativo (FN)	Verdadeiro Positivo (VP)

Fonte: Elaborado pelo autor (2020)

Ao longo da diagonal em azul, estão representadas as predições corretas do modelo: verdadeiros positivos (**VP**) e verdadeiros negativos (**VN**). Os elementos fora dessa diagonal representam os erros cometidos pelo modelo: falsos negativos (**FN**) e falsos positivos (**FP**).

Existem importantes métricas que podem ser extraídas da matriz de confusão, conforme destacado por Luque *et al.* (2019), sendo elas; Sensibilidade ou Taxa de Verdadeiros Positivos, Especificidade ou Taxa de Verdadeiros Negativos, Acurácia ou Taxa de Casos Corretamente Classificados e Precisão ou Taxa de Positivos Previstos, conforme demonstrado no **Quadro 6**, apresentado abaixo:

Quadro 6: Resumo da Matriz de Confusão e métricas para avaliação de modelos

Matriz de confusão		Previsto pelo Modelo		Sensibilidade	Especificidade	Precisão	Acurácia
		Não churner (N)	Churn				
Situação Real	Não churner (N)	Verdadeiro Negativo (VN)	Falso Positivo (FP)	$= \frac{VP}{VP + FN}$	$= \frac{VN}{VN + FP}$	$= \frac{VP}{VP + FP}$	$= \frac{VP + VN}{VP + VN + FP + FN}$
	Churner (P)	Falso Negativo (FN)	Verdadeiro Positivo (VP)				

Fonte: Elaborado pelo autor (2020)

Derivada da Matriz de Confusão, o *trade off* da relação entre os indicadores de sensibilidade e especificidade, em um plano gráfico, gera o que se denomina Curva ROC (*Receiver Operating Characteristic*). A curva ROC fornece uma estimativa da capacidade discriminativa do classificador em termos das probabilidades de erro, refletindo os erros de classificação (Castro e Braga, 2011).

Fawcett (2003) define a curva ROC como uma métrica para avaliação de modelos que permite estudar a variação relacionada às medidas de sensibilidade e à especificidade do modelo, para diferentes valores de ponto de corte.

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).

Para finalizar este capítulo sobre a metodologia e o *design* desta pesquisa, vale a pena ressaltar que todos os dados foram analisados no *software* Stata16.

5. Análise e Discussão de Resultados

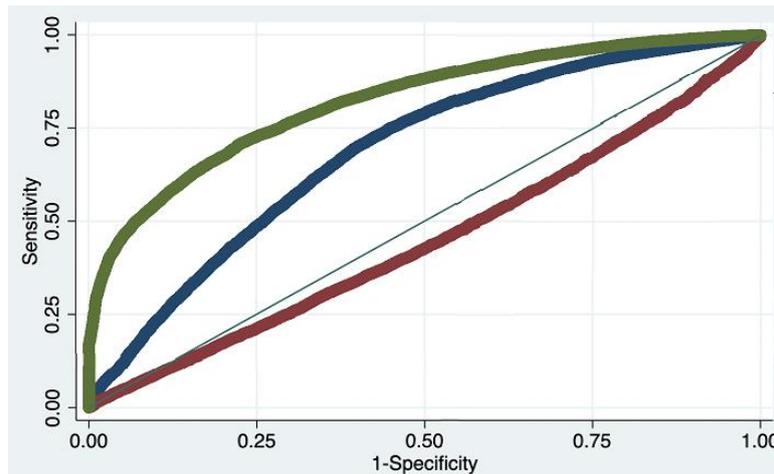
Neste capítulo, apresentar-se-á todo o processo de estimação e resultados. Para tal, encontra-se organizado da seguinte forma: na seção **5.1**, são apresentadas algumas notas prévias relativas ao tratamento e manipulação dos dados e os resultados do modelo proposto para o estudo das hipóteses 1, 2, 3 e 4. Na seção **5.2**, são expostos todos os resultados do processo de estimação, sendo apresentada uma análise e interpretação dos resultados obtidos.

5.1 Modelo

A variável dependente (**Y**), ProxyChurn, foi construída a partir de informações obtidas da base de dados referentes às quantidades de operações e produtos que os clientes transacionaram com o banco ao longo de dois anos, de abril 2018 a abril de 2020. Aderente às premissas descritas anteriormente, a técnica preditiva desenvolvida foi a Regressão Logística Multinível.

A técnica estatística base desse estudo é a Regressão Logística em Painel, porém, para selecionar o melhor modelo entre efeitos fixos, randômicos e mistos utilizou-se a métrica Área sob a Curva ROC. A regressão logística de efeitos mistos apresentou o resultado de 0,8212, seguida pela regressão de efeitos fixos com 0,6879 e por fim, a regressão de efeitos aleatórios 0.4334, sendo o melhor poder explicativo apresentado pela Regressão Logística Multinível de efeitos mistos (0.8212). A Figura 5 apresenta as respectivas curvas ROC de regressão logística.

Figura 5: Curva ROC Regressões Logísticas



Fonte: Elaborada pelo autor a partir do *software* Stata e banco de dados da empresa estudada (2020)

Na **Tabela 7**, são apresentados os resultados do modelo completo proposto para o estudo das hipóteses **H1**, **H2**, **H3** e **H4**. A primeira coluna apresenta apenas as variáveis controle e, na segunda, estão dispostas as variáveis controle, com a inclusão da implementação do indicador de *churn* (**IchurnRate**). A terceira coluna mostra a variável **Experiência.G** com as variáveis de controle. A quarta mostra a variável **TempoCasa.G** com as variáveis de controle e a quinta coluna, a variável **Qtde.Cliente.G** com as variáveis de controle. A sexta coluna refere-se ao modelo completo, enquanto a última coluna, só às variáveis independentes.

O modelo de regressão logística multinível longitudinal foi estimado por meio do *software* STATA16. O valor de $p\text{-valor} < 0,05$ (5%) será considerado estatisticamente significativo para todas as variáveis analisadas.

Tabela 7: Modelo – Hipóteses H1, H2, H3 e H4

Variável Resposta: ProxyChurn (odds ratio)	H	Variáveis Controle	IChurnRate + Controle	Experiência.G + Controle	TempoCasa.G + Controle	Qtde.Cliente. G + Controle	Modelo Completo	Modelo Completo (sem variáveis controle)
IChurnRate	H1		1,332*** (0,104)				1,308*** (0,101)	1,660*** (0,067)
Experiência.G	H2			1,001* (0,001)			1,001 (0,001)	1,001* (0,003)
TempoCasa.G	H3				1,003*** (0,001)		1,003*** (0,001)	1,001*** (0,000)
Qtde.Cliente.G	H4					0,997 (0,004)	0,993 (0,004)	0,997 (0,002)
Gênero.G		0,872 (0,121)	0,872 (0,121)	0,918 (0,129)	0,966 (0,136)	0,847 (0,119)	0,964 (0,139)	
Qtde.Total.G		1,006* (0,003)	0,993 (0,005)	1,005 (0,003)	1,005 (0,003)	1,006* (0,003)	0,993 (0,005)	
Segmento.C		0,968 (0,114)	0,969 (0,114)	0,944 (0,112)	0,868 (0,105)	0,929 (0,112)	0,824 (1,028)	
Qtde.Operações		1,051*** (0,007)	1,050*** (0,007)	1,050*** (0,007)	1,050*** (0,007)	1,050*** (0,007)	1,050*** (0,007)	
Qtde.Produutos		1,644*** (0,047)	1,649*** (0,047)	1,642*** (0,046)	1,644*** (0,046)	1,644*** (0,046)	1,647*** (0,047)	
Rec.Judicial		1,004*** (0,001)	1,003*** (0,001)	1,004*** (0,001)	1,004*** (0,001)	1,004*** (0,001)	1,003*** (0,001)	
Falência		0,997*** (0,001)	0,996*** (0,001)	0,997*** (0,001)	0,997*** (0,001)	0,997*** (0,001)	0,996*** (0,001)	
USDVar		0,653 (0,194)	0,661 (0,195)	0,659 (0,195)	0,669 (0,198)	0,653 (0,194)	0,680 (0,202)	
Ano								
2019		1,443*** (0,055)	1,272*** (0,065)	1,429*** (0,549)	1,404*** (0,055)	1,443*** (0,055)	1,242*** (0,064)	
2020		2,120*** (0,171)	1,920*** (0,163)	2,089*** (0,169)	2,039*** (0,166)	2,121*** (0,172)	1,844*** (0,158)	
Cons		0,846 (0,228)	2,996* (1,307)	0,507 (0,180)	0,749 (0,204)	1,036 (0,315)	2,155 (1,117)	1,358* (0,199)

Nota: *** p<0.01, **p<0.05, * p<0.1; (erro padrão)

Fonte: Formulada pelo autor (2020)

Para o modelo rodado, a variável resposta é o *churn* do cliente, na qual os resultados dos coeficientes da **Tabela 7** indicam se cada variável aumenta ou reduz a probabilidade da ocorrência de *churn*. O *odds ratio* é a razão de probabilidade, na qual os resultados acima de 1 indicam que a variável aumenta a probabilidade da ocorrência do *churn*, enquanto os resultados abaixo de 1 indicam redução desta probabilidade.

Na variável de controle **ano**, significativa no modelo da primeira coluna, observa-se a crescente probabilidade do cliente se tornar *churn* ao longo do tempo. As maiores quantidades de operações e produtos oferecidos para consumo também aumentam a probabilidade de *churn*.

As variáveis de controle significantes permaneceram as mesmas em todos os modelos. Além disso, observou-se significância para a implementação do indicador da taxa de *churn*

(**IChurnRate**) no modelo da segunda coluna, sendo que, nesta interpretação, após a implementação do indicador da taxa de *churn*, a probabilidade de *churn* dos clientes aumentou em 33,2% ($\beta = 1,332$, $p < 0.01$). O mesmo efeito positivo foi observado no modelo da quarta coluna, com a inclusão da variável **tempo de casa do gerente** ($\beta = 1,003$, $p < 0.01$).

Observa-se que com a inclusão das variáveis **experiência do gerente** ($\beta = 1,001$, $p > .05$), na terceira coluna, e a **quantidade de clientes por gerente** ($\beta = 0,997$, $p > .05$), na quinta coluna, não foram significantes para explicar o comportamento de *churn* dos clientes.

A partir dos questionamentos propostos em **H1**, **H2**, **H3** e **H4**, os resultados do modelo completo mostram um efeito positivo e significativo na probabilidade de *churn* dos clientes após a implementação do indicador da taxa de *churn*. Não foram observadas alterações nas variáveis de controle, sendo assim, ao contrário do esperado, após a implementação do indicador da taxa de *churn* na carteira de clientes do banco, houve um aumento de 30,8% na probabilidade do cliente deixar o banco ($\beta = 1,308$, $p < 0.01$). Portanto, **H1** não foi suportada.

Além disso, constatou-se que a variável **tempo de casa dos gerentes** apresenta um efeito positivo e significativo sobre o *churn* de clientes. Com o aumento do **tempo de casa dos gerentes**, mensurado pelos meses trabalhados no banco, a probabilidade do cliente se tornar *churn* aumenta em 0,3% ($\beta = 1,003$, $p < 0.01$), não suportando **H3**.

Pelos resultados do modelo completo, a **experiência do gerente** aumenta a probabilidade do cliente se tornar *churn* ($\beta = 1,001$, $p > 0.05$), assim como a **quantidade de clientes por gerente** ($\beta = 0,993$, $p > 0.05$), mas, não têm significância estatística, mesmo na ausência das demais variáveis.

Portanto, ao contrário da expectativa da instituição financeira, objeto de análise desta pesquisa, os resultados não foram satisfatórios com a implementação do indicador da taxa de *Churn*. Ao contrário, **H1** foi rejeitada e, com isso, houve uma relação positiva entre a implementação do **IChurnRate** e a evasão de clientes do banco.

Analisando-se os fatores de sucesso da implementação dessas ferramentas em grades bancos, pode-se direcionar a reflexão para uma tentativa de entender o porquê do fracasso da implementação da estratégia de retenção de clientes através do indicador da taxa de *churn*.

Dessa maneira, conclui-se que apenas a implementação do indicador da taxa de *churn* como estratégia de retenção de clientes não garante a redução da taxa de *churn* e, com isso, melhores resultados. Fica evidente, portanto, a necessidade de uma combinação de tecnologia,

organização de processos e de recursos humanos, além de um período de tempo maior de análise.

Um dos fatores de sucesso observados durante o desenvolvimento metodológico desse estudo é a performance tecnológica da ferramenta. Devido à falta de orçamento destinado à estratégia de retenção, decidiu-se realizar o desenvolvimento do sistema de apuração da taxa de *churn* e modelo da análise preditiva de *churn* internamente, no banco. Vale ressaltar que não foi possível a contratação de uma consultoria especializada, a qual poderia otimizar os processos internos e atenuar todos os pontos de resistência à mudança.

Assim, fica evidente a importância de um plano estratégico para retenção de clientes com base no indicador da taxa de *churn*, que envolva e integre a tecnologia aos processos e aos recursos humanos, com ênfase para gerentes comerciais, que são os agentes principais de relacionamento entre o banco e os clientes.

Considerando a dificuldade do plano estratégico em bancos de médio porte, o envolvimento do CEO torna-se condição fundamental para o sucesso da implementação da estratégia de retenção de clientes. Os bancos de médio porte possuem, normalmente, processos e estruturas hierárquicas muito mais simples e diretos, quando comparados aos grandes bancos, o que os tornam muito mais flexíveis e rápidos às mudanças, e são altamente centralizados nos CEOs, que faz com que toda a organização esteja inclinada a responder às decisões que estes apontam como de alta significância de valor.

A Matriz de Confusão apresentada a seguir, no **Quadro 7**, mostrou que a taxa de acerto geral do modelo de Regressão Logística Multinível é de 77,22%. Este estudo teve como objetivo identificar quais clientes tendem ao *churn*, por isso, temos a Sensibilidade como principal métrica de avaliação do modelo, pois expressa a quantidade de positivos corretamente classificados, que alcançou 97,85%. Esses números, corroboram com a utilização dos modelos de Regressão Logística Multinível para estimar a probabilidade de um cliente se tornar inativo.

O **Quadro 7** apresenta a Matriz de Confusão da análise realizada:

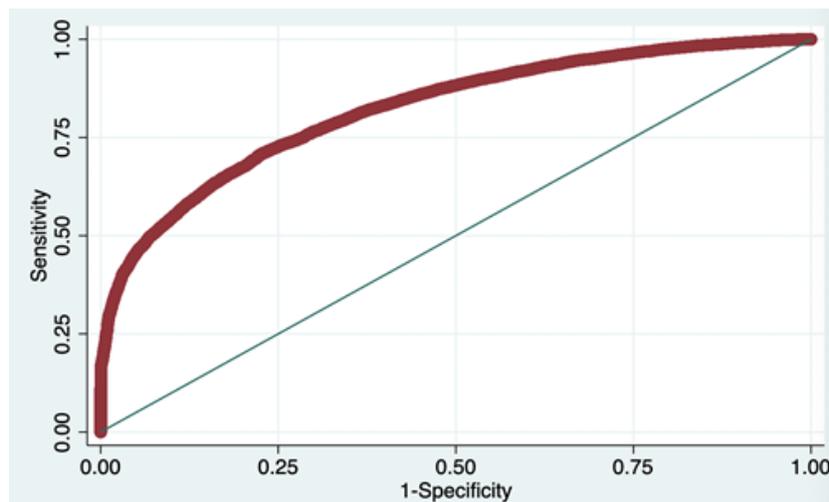
Quadro 7: Matriz de Confusão dos resultados da Regressão Logística

Matriz de confusão		Previsto pelo Modelo		Sensibilidade	Especificidade	Precisão	Acurácia	AUC
		Não churmer (N)	Churn					
Situação Real	Não churmer (N)	1470	7897	97,85%	15,693%	77,22%	76,89%	0,82
	Churmer (P)	589	26769					

Fonte: Elaborado pelo autor (2020)

A capacidade de previsão do modelo completo, curva ROC, que demonstra como o modelo se comporta relacionando os acertos e erros das classes previstas (*churn* e não *churn*) às observações reais, pode ser observada na **Figura 6** e na **Tabela 8**, apresentadas abaixo:

Figura 6: Curva ROC



Fonte: Elaborada pelo autor a partir do *software* Stata e banco de dados da empresa estudada (2020)

Tabela 8: Resultado Curva ROC

	Observações	Área ROC	Desvio Padrão	Intervalo de Confiança - 95%	
ROC	35.256	0.8212	0.0024	0.81647	0.82603

Fonte: Elaborada pelo autor a partir do *software* Stata e banco de dados da empresa estudada (2020)

A métrica **Área sob a Curva ROC** apresentou o resultado de 0,8212, o que significa que há 82,12% de probabilidade de o modelo completo classificar corretamente cada cliente, resultado significativo quando os resultados possíveis são dicotômicos.

5.2 Resumo dos resultados

O **Quadro 8** apresenta um sumário dos resultados das hipóteses, apresentados na seção **5.1** deste capítulo:

Quadro 8: Resumo dos resultados das hipóteses

Hipótese	Descrição	Sinal	Significância Estatística	Resultado
H1	A implementação do indicador da taxa de churn apresenta impacto negativo no churn clientes	Positivo	Significante ***	Hipótese não suportada. A implantação do indicador da taxa de churn, resulta em maior churn de clientes
H2	A maior experiência do gerente de atendimento apresenta impacto negativo no churn clientes	Positivo	Não Significante	Hipótese não suportada. A maior experiência do cliente não é relevante para explicar o churn de clientes
H3	O maior tempo de casa do gerente apresenta impacto negativo no churn clientes.	Positivo	Significante ***	Hipótese não suportada. O maior tempo de casa do gerente resulta em um maior churn de clientes
H4	A maior quantidade de clientes por gerente apresenta impacto positivo no churn de clientes.	Positivo	Não Significante	Hipótese não suportada. A quantidade de clientes por gerente, não é relevante para explicar o churn de clientes

Nota: Nível significância *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Fonte: Formulado pelo autor (2020)

6. Conclusão, limitações e sugestões para pesquisas futuras

Neste, que é o último capítulo desta dissertação, apresentar-se-á, na seção **6.1**, as conclusões gerais da investigação e, em seguida, na seção **6.2**, são discutidas as respectivas limitações deste estudo de caso e sugestões para trabalhos futuros.

6.1 Principais Conclusões

Os resultados obtidos a partir deste estudo de caso sugerem algumas conclusões interessantes. O objetivo principal deste trabalho, apresentado no primeiro capítulo dessa dissertação, estruturou-se em identificar se a implementação de uma ferramenta de gestão (indicador da taxa de *churn*), que deveria incentivar os gerentes a reduzirem o *churn* de clientes das suas respectivas carteiras de clientes, ajudou, de fato, o banco a reduzir a evasão total de clientes e, além disso, a presente pesquisa também se esforçou para apurar quais fatores afetam significativamente a probabilidade de reter tais clientes.

Sabe-se que a perda de clientes é a principal barreira para uma empresa elevar sua rentabilidade (Cheney, 2016; Zaharik, 1993; Reichheld; Sasser, 1990). Para cumprir os objetivos, investigando as hipóteses propostas para o desenvolvimento deste trabalho, utilizou-se o modelo probabilístico de Regressão Logística Multinível, descrevendo, para cada indivíduo presente na base analisada de clientes do banco, o risco de abandono.

Prever quais clientes tendem a migrar para instituições financeiras concorrentes, com a intenção de agir para evitar essa migração, é uma questão que pode ser resolvida por meio de metodologias de análise preditiva, permitindo que a empresa realize uma gestão proativa (Khakabi, 2010; Gholamian, 2010; Namvar, 2010; Vandecruys *et al.*, 2008). Consequentemente, após a aplicação do modelo de *churn*, foi possível traçar o perfil dos clientes mais propensos a se tornar *churn*, como também os menos propensos.

É possível afirmar, portanto, diante dos resultados obtidos, que o modelo proposto pode ser eficiente para determinação do risco de abandono de clientes, a partir de variáveis transacionais. Nos 1469 clientes analisados, as variáveis explicativas significativas, no modelo para determinar a propensão a *churn* foram a implementação do indicador da taxa de *churn* (**H1**) e o tempo de casa do gerente (**H3**).

A tabela de classificação mostrou que a taxa de acerto geral do modelo de Regressão Logística é de 77,22% e que as taxas de acerto dos grupos individuais indicam consistência na

previsão. O grupo que cancela apresentou taxa de acerto de 97,85% e a curva ROC apresentou um poder de explicação de 0.82.

A combinação dessas medidas de avaliação do ajuste e da precisão do modelo indica a aceitação deste como um modelo de Regressão Logística Multinível significativa. Assim, de acordo com as variáveis assumidas no modelo em questão, conclui-se que o perfil do cliente com maior risco de encerrar o relacionamento com o banco reúne as seguintes características: este cliente é aquele que é atendido por gerente com maior tempo de experiência e maior tempo de casa. Já a implantação do indicador da taxa de *churn*, para gerenciamento analítico, influenciou negativamente na retenção de clientes.

É importante destacar que a metodologia, estratégia e *design* de pesquisa aplicadas nesta dissertação, compostas por métodos qualitativos, mostraram-se apropriadas para responder às perguntas de pesquisa e apoiar os objetivos. Com isso, este estudo trouxe revelações importantes e respostas às questões de pesquisa propostas, além oferecer diversas contribuições teóricas e práticas.

6.2 Limitações e sugestões para estudos futuros

Este estudo está baseado em uma amostra de dados de uma única instituição financeira, sendo esta a maior limitação do estudo. Os resultados refletem a efetividade da ação estratégia implementada e características da equipe de gerentes comerciais deste banco, especificamente, sendo que a mesma ação realizada em outros bancos ou setores podem apresentar variações.

Os dados retirados do banco de dados da instituição financeira, objeto deste estudo, não contemplavam variáveis importantes de satisfação e situações de atrito, tampouco uma variável com capacidade de indicar a desidratação de clientes, exigindo que a variável dependente tivesse de ser elaborada pelo autor.

O período da base de dados utilizada é relativamente curto e, por se tratar de um banco médio, a quantidade de clientes, relativamente pequena, quando comparada a bancos de varejo ou empresas do setor de telecomunicação, por exemplo, nas quais os estudos do fenômeno do *churn* são mais difundidos.

É importante ressaltar, ainda, que a redução do volume de negócios nem sempre é decorrente de *churn*. O cancelamento ou redução de produtos oferecidos, a quantidade de

operações ou volumes transacionados podem ser motivados pela simples falta de necessidade. No entanto, outra limitação do estudo é que não foi possível realizar essa distinção, configurando uma limitação do método.

Quanto aos esforços para pesquisas futuras, esta dissertação abre diversas frentes a serem exploradas.

Em relação à metodologia, novas variáveis explicativas podem ser testadas, assim como outras proposições para a variável dependente. O modelo testado neste estudo poderia ser replicado em outros estudos de casos para aumentar a validação das descobertas. A partir de um esforço conjunto de pesquisadores, é possível comparar os fatores críticos de sucesso presentes em cada estudo, considerando cada contexto organizacional e tornando, assim, os constructos apresentados nessa dissertação mais robustos.

Por fim, a comparação dos efeitos entre diferentes setores pode apresentar resultados interessantes a serem avaliados. Recomenda-se, portanto, a comparação do modelo logístico com outros instrumentos, como Árvores de Decisão, Redes Neurais, entre outras.

Referências bibliográficas

ACCENTURE. The new customer imperative: retaining and acquiring customers in a changed banking landscape. **The Point**, v. 9, n. 4, p. 1-8, 2009.

AMIN, A., AL-OBEIDAT, F., SHAH, B., ADNAN, A., LOO, J., ANWAR, S. Customer churn prediction in telecommunication industry using data certainty. **Journal of Business Research**, v.94, p. 1-12, 2018.

ASCARZA, E., NESLIN, S., NETZER, O., ANDERSON, Z., FADER, P., GUPTA, S., SCHRIFT, R. In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions. **10th Triennial Invitational Choice Symposium**, v.5, p. 65-81, 2017.

AXELSSON, R., NOTSTAM, A. **Identify Churn: A study in how transaction data can be used to identify churn for merchants**, 2017.

BARRETO, I. F., CRESCITELLI, E., FIGUEIREDO, J.B. Resultados de marketing de relacionamento: proposições de modelo por meio de mapeamento cognitivo. **Revista Brasileira de Gestão de Negócios**, v. 17, n. 58, p. 1371-1389, 2015.

BERRY, L. L. Relationship Marketing of Services – growing interest, emerging perspectives. **Journal of the Academy of Marketing Science**, v. 23, n. 4, p.236-245, 1995.

BOTELHO, D., TOSTES, F. D. 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 behaviorally loyal clients in a non-contractual FMCG retail setting. **European Journal of Operational Research**, 164(1), p. 252-268, 2015.

BUTTLE, F. **Customer Relationship Management**. Routledge, 2004.

BUREZ, J., 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, 2008.

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

CLARO, D. P. **Marketing de relacionamento: conceitos e desafios para o sucesso do negócio**. Série Working paper Insper (WPE – 076), 2006

CAPRARO, A., BRONIARCZYK, S., SRIVASTAVA, R. Factors Influencing the Likelihood of Customer Defection: The Role of Customer Knowledge. **Journal of the Academy of Marketing Science**, 31(2), p. 164-175, 2003.

DATTA, P., MASSAND, B., MANI, D. R. LI, B. Automated Cellular Modeling and Prediction on a Large Scale. **Issues on the application of data mining**, v. 14, p. 485-502, 2001.

DE CAIGNY, A., COUSSEMENT K., DE BLOCK, K. W. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision tree. **European Journal of Operation Research**, n. 269, pp. 760-772, 2018.

EIBEN, A. E., KOUDIJS, A. E., SLISSER, F. Genetic modelling of customer retention. In: **European Conference on Genetic Programming**. Springer, Berlin, Heidelberg. p. 178-186, 1998.

CHENEY, C. Understanding *Churn*: Why Do Customers Unsubscribe? **Promotional Marketing**, [s. l.], p.1, 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.

DAVID W. H., STANLEY, L. **Applied Logistic Regression**. Wiley Publishing, 2000.

DAWES, J., SWAILES, S. Retention sans frontieres: issues for financial service retailers. **International Journal of Bank Marketing**, v.17, n. 1, p. 36-43, 1999.

DE CAIGNY, A., COUSSEMENT, K., DE BOCK, K. W. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. **European Journal of Operational Research**, v. 269, n. 2, p. 760-772, 2018.

DELEN, D. A comparative analysis of machine learning techniques for student retention management. **Decision Support Systems**, v. 49, n. 4, p. 498-506, 2010.

DOMAZET, I. Development of Long-term Relationships with Clients in Financial Sector Companies as a Source of Competitive Advantage. **Economic Sciences Series**, v. 62, n. 2, p. 1-11. 2010.

Dwyer, F., SCHURR, P., OH, S. Developing Buyer-Seller Relationships. **Journal of Financial Services Marketing**, 7(2), p. 186-198, 1987.

GAURAV, K. Impact os Relationship Marketing Strategy on Customer Loyalty, **ICFAI Journal of Management Research**, v.7, n.11, p.7-21. 2008.

GANESH, J., ARNOLD, M. J., REYNOLDS, K. E. Understanding the customer base of service providers: an examination of the differences between switchers and stayers. **Journal of Marketing**, 64(3), p. 65-87, 2000.

GUENZI, P., PELLONI, O. The impact of interpersonal relationships on customer satisfaction and loyalty to the service provider. **International Journal of service industry management**, v. 15, n. 4, 365-384, 2004.

GUJARATO, D. N.m Monteiro, M. J. C. (Trad.). **Econometria básica**, v.4. ed. Rio de Janeiro: Elsevier, p. 812, 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, 2014.

HANDLEY, L. Customer retention: brave new world of consumer dynamics. **Marketing Week Online Edition**, v. 21, 2003.

HALINEN, A., Tahiten, J. A process theory of relationship ending. **International Journal of Service Industry Management**, 13(2), p. 163-180, 2002.

HAIR, J. F. *et al.* **Análise multivariada de dados**. Bookman Editora, 2009.

HOLTROP, N., WIERINGA, J., GIJSENBURG, M., VERHOEF, P. No future without the past? Predicting churn in the face of customer privacy. **International Journal of Research in Marketing**, 34, p. 154–172, 2017.

KATELARIS, L., THEMISTOCLEOUS, M. Predicting Customer Churn: Customer Behavior Forecasting for Subscription-Based Organizations. **European, Mediterranean, and Middle Eastern Conference on Information Systems**, p. 128-135, 2017.

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.

KRASNIKOV, A., JAYACHANDRAN, S., KUMAR, V. The impact of customer relationship management implementation on cost and profit efficiencies: evidence from the I.S. commercial banking industry. **Journal of Marketing**, v. 73, p.61-76. 2009.

KEAVENEY, S. Customer Switching Behavior in Service Industries: An Exploratory Study. **Journal of Marketing**, 59(2), p. 71-82, 1995.

KERAMATI, A., GHANEEI, H., MIRMOHAMMADI, S. M. Developing a prediction model for customer churn from electronic banking services using data mining. **Financial Innovation**, 2(1), p. 1-13. 2016.

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, 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.

JAHANZEB, S., JABEENM 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.

LIOU, J. J. H. A novel decision rules approach for customer relationship management of the airline market. **Expert systems with Applications**. 36(3), p. 4374-4381, 2009.

LINOFF, G. S., BERRY, M. JA. **Data mining techniques: for marketing, sales, and customer relationship management**. John Wiley & Sons, 2011.

LOPES, L., BRITO, C. H., & ALVES, H. Lost and win-back customers: towards a theoretical framework of Customer Relationship Reactivation. **41st EMAC Conference**, 2012.

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

MENARD, S. Coefficients of determination for multiple logistic regression analysis. **The American Statistician**, v. 54, n. 1, p. 17-24, 2000.

NIE, G., ROWE, W., ZHANG, L., TIAN, Y., SHI, Y. Credit card churn forecasting by logistic regression and decision tree. **Expert Syst Appl**. 38, p. 15273–15285, 2011.

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

OLIVER, R. Whence Consumer Loyalty? **Journal of Marketing**, 63, p. 33-44, 1999.

OWCZARCZUK, M. *Churn* models for prepaid customers in the cellular telecommunication industry using large data marts. **Expert Systems with Applications**, v. 37, n. 6, p. 4710-4712, 2010.

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.

Reichheld, F. F., Sasser, W. E. Jr.. Zero defections: Quality comes to service. **Harvard Business Review**, 68(5), p. 105–111, 1990.

RIBEIRO, Á. H.P. Database marketing: A ferramenta básica para o marketing individualizado e interativo. **Jornal O Povo**, Fortaleza, 1997.

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

RYALS, L., PAYNE, A. Customer Relationship Management in Financial Services: Towards Information Enabled Relationship Marketing. **Journal of Strategic Marketing**, v.9, p.3-27. 2001.

SHARMA, A., PANIGRAHI, P.K. A neural network based approach for predicting customer churn in cellular network services. **Int J Comput Appl**, 27(11), p. 26–31, 2011.

SHAH, D., KUMAR, V., QU, Y., CHEN, S. Unprofitable Cross-Buying: evidence from consumer and business markets. **Journal of Marketing**, v. 76, p. 78-95. 2012.

SHERER, L. J., TOALDO, A. M. M. Um estudo da atitude do consumidor frente às estratégias de retenção das quatro maiores operadoras de telefonia móvel do Brasil. **Revista Brasileira de Gestão de Negócios**, v.13, n.40, p. 243-263. 2011.

SINGH, R., KHAN, I. A. An approach to increase customer retention and loyalty in B2C world. **International Journal of Scientific and Research Publications**, 2(6), p. 606-611, 2012.

TSAI, C. F., LU, Y. H. Customer churn prediction by hybrid neural networks. **Expert Systems with Applications**, 36(10), p. 12547-12553, 2009.

VAN DEN POEL, D., LARIVIERE, B. Customer attrition analysis for financial services using proportional hazard models. **European Journal of Operational Research**, 157(1), p. 196-217, 2004.

VAN DOLEN, W. *et al.* Customer-sales employee encounters: a dyadic perspective. **Journal of Retailing**, v. 78, n. 4, p. 265-279, 2002.

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

VASILUI, D. M. Approach to customer relationship management (CRM) – The New Key Sales Success. **The USV Annals of Economics and Public Administration** v.12, n.15, p. 144. 2012.

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

Verbeke, W., Martens, D., Mues, C., & Baesens, B. Building comprehensible customer churn prediction models with advanced rule induction techniques. **Expert systems with applications**, 38(3), p. 2354-2364, 2011.

WEST, B. T., Welch, K. B., Galecki, A. T. **Linear mixed models: a practical guide using statistical software**. 2. Ed. Boca Raton: Chapman & Hall/CRC Press, 2015.

WOLDRIDGE, J. M. **Introdução à econometria: uma abordagem moderna**. São Paulo: Cengage Learning, p.5. 2010.

YU, X., GUO, S., GUO, J., & HUANG, X. An extended support vector machine forecasting framework for customer churn in e-commerce. **Expert Systems with Applications**, 38(3), p. 1425-1430, 2011.

YENGIN, I., KARAHOCA, A., KARAHOCA, D. E-learning success model for instructors' satisfactions in perspective of interaction and usability outcomes. **Procedia Computer Science**, v. 3, p. 1396-1403, 2011.

ZHAO, L., GAO, Q., DONG, X., DONG, A., DONG, X. K-local maximum margin feature extraction algorithm for churn prediction in telecom. **Cluster Computing**, 20(2), p. 1401–1409, 2017.